



# Housing Discrimination against Same-sex Couples, and Transgender Individuals, Data & Documentation

February 2016

*Prepared for:*

U.S. Department of Housing and Urban Development

## **I. Introduction**

The data from the Pilot for Estimate of Housing Discrimination against Lesbian, Gay and Transgender People (hereafter referred to as HDS-LGT) are divided by raw data (collected directly from testing and extracted from a database maintained internally at the Urban Institute) and analysis data (cleaned data used for analysis, including universe flags).

The lesbian, gay and transgender fieldwork was conducted independently as three separate tracks but data were all recorded into one combined database. Therefore, this submission includes raw data files containing data from all three fieldwork tracks. The analysis files, however, have been separated into three datasets: for remote tests, in-person tests, and multivariate regression analysis.

All data files are provided as SAS datasets, except for the narratives file, which is provided in Microsoft Excel format. We have included a codebook for each dataset that indicates variable names, data source, labels, type, in-person/remote (raw data only) and form order (raw data only). Please note that sensitive identifying information about both the testers and the agents has been redacted from the SAS datasets (not including the narratives file), and the codebooks clearly reflect these redactions.

Additional detail about the structure of the data files, the level of observation for each file, and how the files fit together is provided below. Section II provides a description of the data collection process. Section III describes the raw and analysis files and provides guidance on their use. Section IV describes each data file. Section V provides the unique identifier(s) for each of the raw data files.

## **II. A Guide to the Testing Process**

The raw data from HDS-LGT were collected during field testing. Different protocols were followed for tests conducted in-person and tests conducted remotely (by phone and email).

For in-person testing, each test began with the selection of an advertisement, advance contact (by a specialized caller), followed by appointment contact (by testers), continuing through to an in-person visit. During field testing HDS-LGT testers collected data by filling out report forms online. Each form used during field testing corresponds to a raw data file that has been included with this deliverable.

For remote testing, each test began with the selection of an advertisement, advance contact (by a specialized caller), followed by a remote test contact (by testers). Because these tests did not include an in-person component, testers attempted to acquire information about available rental units over the phone. They also attempted to make an appointment to meet with the agent; if successful, they later cancelled the appointment.

### III. Data Files

#### Raw Data Files

The raw data files include the following datasets:

- lgt\_tester (tester application)
- lgt\_taf (test authorization form, preliminary assessment form and final assessment form)
- lgt\_advancecall (advanced contact form)
- lgt\_assignment (test assignment form)
- lgt\_apptcall (appointment contact form)
- lgt\_rental (site visit form)
- lgt\_availunit (available unit form for in-person tests)
- lgt\_availunitt (available unit form for remote tests)
- lgt\_rentnarr (site visit narrative)
- lgt\_followup (agent follow-up form)

The lgt\_taf data file contains information from three HDS-LGT forms: the test authorization form, the preliminary assessment form and the final assessment form. Data found in the lgt\_taf files can be thought of as “database administrative data” because they are not entered by testers. Instead, these files include information about the test including the housing advertisement, the outcome of the Urban Institute quality review, and final disposition date of the test (whether it passed quality review or not).

Lgt\_rentnarr is provided in Excel format because SAS truncates character variables at 1,024 characters. These narratives have not been redacted by the Urban Institute and may contain sensitive information about testers and agents.

*Linking data across forms.* Data from the various forms can be linked using the test identifier, CONTROL, and the IDs of the testers involved in the test, TESTERID. CONTROL takes the form SS-RT-XXXX-A, where SS indicates the site code (a list of the three site codes can be found in the codebook); R indicates that all tests are rental tests; T indicates the orientation and track of the testers as follows:

- A = Lesbian, white, in-person
- B = Lesbian, minority, in-person
- C = Gay, white, in-person
- D = Gay, minority, in-person
- E = Lesbian, remote
- F = Gay, remote
- G = Transgender, convey
- H = Transgender, non-convey;

XXXX is the test number; A is always set to 1. For example, DC-RG-1516-1 can be read as a transgender test in Washington, DC where the protected tester conveyed their orientation to the housing agent. The test was based on the 16th sampled ad in the testing track during the 15th week of testing.

Each HDS-LGT tester is assigned a unique TESTERID. TESTERID takes the form SSNNN. SS indicates the site code; NNN is a numeric digit unique to each tester. For example, DC137 is a tester in Washington, DC and was the 137<sup>th</sup> tester in that site to complete an application. Every form that a tester fills out in HDS-LGT has the tester's ID number appended. Data from their tester application are reported in the file LGT\_TESTER. TESTERID cannot be linked across sites, and while a tester may have conducted both in-person and remote tests in the same site, separate tester ID numbers would have been assigned for their in-person testing vs. remote testing activities.

Suppose, for example, an analyst wants to create a measure of the number of rental units inspected on an in-person test. To link the rental site visit form for a given tester to the available units that they were told about on that test, the analyst would do a one-to-many merge of the LGT\_RENTAL and LGT\_AVAILUNIT files using the CONTROL and TESTERID variables. The resulting matched file could then be used to generate a count of the number of available units inspected on each tester's portion of the test.

*Sequence.* Several of the HDS-LGT forms can be filled out more than once per test. In these cases, the raw data in a given file are organized by sequence number. In general, the sequence indicates the order in which the tester completed the forms. For example, if a tester was told about three rental units on a test, the variable SEQUAU on the Available Unit Form reflects the order in which the tester completed the forms. Additional detail about sequence can be found in Section V of this memo.

A similar concept, test sequence, indicates the order in which testers were assigned to conduct the test. For HDS-LGT, either the control or protected tester is randomly selected to begin an assigned test first. The variable SEQUENCE on the Assignment Form indicates which tester was sent first.

*Deleted Records.* During the course of testing, forms may be deleted by Test Coordinators or Regional Coordinators. This would occur when a mistake was made or duplicate forms were saved to the database. Deleted records have been removed from this data submission.

*Quality Review.* After each test was completed, a quality review was conducted by the Urban Institute. The quality review checked that protocol violations and data entry errors were not made. If the Urban Institute reviewer indicated that a test had passed quality review, they made a note on the Final Assessment Form. This data submission includes only tests that passed quality review.

#### Analysis Data Files

For analysis, the Urban Institute created three data files – one for in-person tests, one for remote tests and one for multivariate analysis. These files were created by combining information across different forms and multiple reports of the same forms for a given test. Analysis files include the universe flags necessary to reconstruct outcomes from the report.

For example, to create the measure for whether one tester was told about more units on an in-person test, we combined records from the site visit form and available unit form, and applied the flag for tests where both testers met with an agent.

The analysis data files include the following datasets:

- lgt\_inperson\_analysis (analysis file for in-person tests)

- lgt\_remote\_analysis (analysis file for remote tests)
- lgt\_mv (analysis file for multivariate regression analysis)

*Flags.* For each outcome variable, we created a flag indicating whether a particular test should be used in a given analysis. The flags are used to restrict the analysis to those tests that passed quality review and where both testers of a matched pair had documented information available for the calculation. The latter includes restrictions such as “both testers met with an agent” or “both testers were told about an available unit.” For example, rents are compared only when both testers were told about at least one unit and reported a valid rent.

Note: replication of our analysis cannot be completed using the raw data unless these restrictions are accounted for. An analysis of the data with different restrictions is possible using the raw data files, but will almost always yield estimates different from those reported in the study.

#### IV. Data Files

The lists below provide an overview of the contents of the data found in each of the files included with the data deliverable. Lastly, Record-level describes the level of observation, or what each row in the dataset represents.

##### Raw Data Files

File: lgt\_tester

- Corresponding HDS-LGT Form: Tester Application Form
- Description: Contains information that each tester is required to submit prior to participating as an HDS-LGT tester.
- Record-level: Each record represents one HDS-LGT tester.

File: lgt\_taf

- Corresponding HDS-LGT Form: Test Authorization Form, Preliminary Assessment Form & Final Assessment Form
- Description: Contains unique information about each test, including advertisement information and whether the test passed quality review.

- Record-level: Each record represents one test.

File: lgt\_advancecall

- Corresponding HDS-LGT Form: Advance Contact Form
- Description: Documents each call made by the advance caller/emailer (ACE) to inquire about the availability and accessibility of an advertised housing unit.
- Record-level: Each record represents an advance contact made by the ACE to inquire about housing and accessibility (multiple calls per test are possible).

File: lgt\_assignment

- Corresponding HDS-LGT Form: Assignment Form
- Description: Takes housing advertisements and assigns them to HDS-LGT testers, along with characteristics for testers to use during the test.
- Record-level: Each record represents an assignment made to a tester (each test has up to two records, one per tester).

File: lgt\_apptcall

- Corresponding HDS-LGT Form: Appointment Contact Form
- Description: Documents each contact made by the HDS-LGT tester attempting to make an appointment with the agent.
- Record-level: Each record represents an appointment contact made by the tester (multiple contacts per test are possible).

File: lgt\_rental

- Corresponding HDS-LGT Form: Site Visit Form
- Description: Documents each site visit.
- Record-level: Each record represents a site visit (each test has up to two records, one per tester).

File: lgt\_availunit

- Corresponding HDS-LGT: Available Rental Unit Form

- Description: Documents each available rental unit that agents told testers about during in-person tests only.
- Record-level: Each record represents an available rental unit (multiple available units per tests are possible).

File: lgt\_availunitt

- Corresponding HDS-LGT Form: Available Rental Unit Form
- Description: Documents each available rental unit that agents told testers about during remote tests only.
- Record-level: Each record represents an available rental unit (multiple available units per tests are possible).

File: lgt\_rentnarr

- Corresponding HDS-LGT Form: Site Visit Narrative Form
- Description: Documents narratives written by testers.
- Record-level: Each record represents a rental narrative (each has two records, one per tester).
- Note: Narratives have not been redacted by the Urban Institute and may contain sensitive information about testers and agents.

File: lgt\_followup

- Corresponding HDS-LGT Form: Follow-Up Form
- Description: Documents follow-up contact between tester(s) and agent(s).
- Record-level: Each record represents a contact made between a tester and an agent.

### Analysis Data Files

File: lgt\_inperson\_analysis

- Description: Data file used for in-person analysis.
- Record-level: Each record represents one completed test that passed quality review.
- Additional notes: universe flags included on this file must be incorporated in order to replicate analysis found in the report.



File: lgt\_remote\_analysis

- Description: Data file used for remote analysis.
- Record-level: Each record represents one completed test that passed quality review.
- Additional notes: universe flags included on this file must be incorporated in order to replicate analysis found in the report.

File: lgt\_mv

- Description: Data file used for multivariate regression analysis.
- Record-level: Each record represents one completed test that passed quality review.
- Additional notes: universe flags included on this file must be incorporated in order to replicate analysis found in the report.

## V. Raw Data Files Identifiers

We show the unique identifiers associated with each raw data file in the chart below. Raw data files can be linked using the CONTROL and TESTERID variables. Some files include an additional sequence variable that notes either a form sequence or test sequence.

Raw data files contain information about both in-person and remote tests. However one exception is the available unit forms, which are split into two independent files.

Raw File	Type of Test (In-person or Remote)	Unique Identifiers
LGT_AdvanceCall (Advance Call)	Both	CONTROL
LGT_ApptCall (Appointment Contact)	Both	CONTROL + TESTERID + SEQUENCE
LGT_Assignment (Assignment)	Both	CONTROL + TESTERID
LGT_Availunit (Available Unit in-person)	In-person Only	CONTROL + TESTERID + SEQAU
LGT_AvailunitT (Available Unit remote)	Remote Only	CONTROL + TESTERID + SEQAU
LGT_FollowUp (Follow-up)	Both	CONTROL + TESTERID + SEQFUC
LGT_Narrative (Narrative)	Both	CONTROL + TESTERID
LGT_Rental (Rental Site Visit)	Both	CONTROL + TESTERID
LGT_TAF (Test Assessment Form)	Both	CONTROL
LGT_Tester (Tester Application)	Both	TESTERID