

**READ ME FILE**  
**Pew 2007 Muslim American Study**  
**Data release May 22, 2008**

**Section 1. Introduction**

In 2007, the Pew Research Center conducted what is believed to be the first-ever national telephone survey of a probability sample of Muslim Americans, a rare, dispersed, and highly diverse population. The study examined the political and social values, religious beliefs and practices, and life experiences of Muslims living in the U.S. today. The poll was conducted by telephone using three sampling sources; a geographically stratified random digit dial sample of the general public, which entailed screening 57,549 households to yield 354 completed interviews with a Muslim respondent; a commercial database of 110 million households, of which more than 450,000 included people with likely Muslim first names and surnames; and a re-contact sample of English-speaking Muslim households from previous nationwide surveys conducted since 2000. Full details about the survey methodology are available in the report “Muslim Americans: Middle Class and Mostly Mainstream,” which is available for download at <http://pewsocialtrends.org/assets/pdf/muslim-americans.pdf>.

This document provides important information for use of the datasets. Additional information about question wording and response options is provided in the accompanying codebook.

**Section 2. Important Note Regarding Calculation of Standard Errors and Test Statistics**

Due to the complex design of the Muslim American study, formulas commonly used in RDD surveys to estimate margins of error (standard errors) are inappropriate. Such formulas would understate the true variability in the estimates. Accordingly, we used a repeated replication technique, specifically jackknife repeated replication (JRR), to produce replicate weights which allowed for the calculation of standard errors for this study. Repeated replication techniques estimate the variance of a survey statistic based on the variance between sub-sample estimates of that statistic. The sub-samples (replicates) were created using the same sample design, but deleting a portion of the sample, and then weighting each subsample up to the population total. The units to be deleted were defined separately for each of the three samples (RDD geographic strata, list, recontacts), and within each frame by the strata used in the sampling. A total of 100 replicates were created by combining telephone numbers to reduce the computational effort.

A statistical software package designed for complex survey data is required to correctly compute standard errors and test statistics. Popular packages include WesVar (available free of charge from Westat <http://www.westat.com/wesvar/demo/index.html>), Stata, SAS, SUDAAN, and SPSS Complex Samples. Point estimates may be computed using either dataset (*Pew 2007 Muslim American Study.sav* or *Pew 2007 Muslim American Study with Replicate Weights.sav*). Standard errors and test statistics should only be computed using *Pew 2007 Muslim American Study with Replicate Weights.sav*.

### Section 3. Description of the Datasets

#### *“Pew 2007 Muslim American Studyc.sav”*

This SPSS dataset contains the survey response variables, the full sample weight (WEIGHT), and several auxiliary variables, such as respondent region and the language in which the interview was conducted. We recommend using the full sample weight in analysis so as to properly reflect the sample design. **This dataset should NOT be used to compute standard errors or test statistics because it does not contain the replicate weights.**

#### *“Pew 2007 Muslim American Study with Replicate Weights – for computing standard errors.sc.sav”*

This SPSS dataset contains all the variables in *Pew 2007 Muslim American Studyc.sav* along with the replicate weights. There are 100 replicate weights in total, and they are labeled finalr1, finalr2,...,finalr100. The replicate weights were created using the jackknife repeated replication (JK2) method. The full sample weight is labeled WEIGHT. The respondent identification variable is labeled RESPID.

### Section 4. Protection of Respondent Confidentiality

Several variables in this study were recoded or excluded from the public dataset in order to protect respondent confidentiality. As a matter of policy, the Pew Research Center does not release respondent names, contact numbers, or any other uniquely identifying information, and for this study all of this personally identifying information has been deleted from the Center’s records as well as those of the survey field house so that it no longer exists in any form. The table on the following page summarizes the measures taken to protect confidentiality during creation of the public datasets.

<u>VARIABLE</u>	<u>CHANGE TO ORIGINAL DATASET</u>	<u>EXPLANATION</u>
fips_name	VALUES: if population > 200,000 county is reported, otherwise state is reported	Excluded county and state for respondents in counties with fewer than 200,000 respondents; for these cases only the state is coded.
BIRTH1	DELETED	Suppress low frequency birth countries to avoid disclosing unique family histories
birthplace_resp	Recoded version of BIRTH1 into regions – see codebook for details	Suppress low frequency birth countries to avoid disclosing unique family histories
FATHER1	DELETED	Suppress low frequency birth countries to avoid disclosing unique family histories
birthplace_father	Recoded version of FATHER1 – see codebook for details	Suppress low frequency birth countries to avoid disclosing unique family histories
MOTHER1	DELETED	Suppress low frequency birth countries to avoid disclosing unique family histories
birthplace_mother	Recoded version of MOTHER1 – see codebook for details	Suppress low frequency birth countries to avoid disclosing unique family histories
Femalemuslimadults (Q.Z6)	VALUES = 0, 1, 2, 3+	Suppress counts greater than 3 to avoid disclosing unusual HH compositions
Malemuslimadults (Q.Z6a)	VALUES = 0, 1, 2, 3+	Suppress counts greater than 3 to avoid disclosing unusual HH compositions
Hhchildren (Q.Z6b)	VALUES = 0,1,2,3+	Suppress counts greater than 3 to avoid disclosing unusual HH compositions
childmuslim2 (Q.Z6d)	VALUES = 0,1,2,3+	Suppress counts greater than 3 to avoid disclosing unusual HH compositions
qd1@ (verbatim responses to Q.D1)	DELETED	Exclude verbatim responses in order to avoid disclosing a unique way of speaking.
qe14@ (other-specify verbatim responses to Q.E14)	DELETED	Exclude to avoid disclosing a unique religious background, which in combination with other responses could violate confidentiality
qe15	VALUES (1) Age 0 to 9 (2) Age 10 to 19 (3) Age 20 to 29 (4) Age 30 to 39 (5) Age 40 to 49 (6) Age 50 to 59 (7) Age 60+ (99) "Don't know/Refused".	Collapse responses into categories to avoid disclosing a unique religious background, which in combination with other responses could violate confidentiality
qe16@ (verbatim responses to Q.E16)	DELETED	Exclude verbatim responses to avoid disclosing a unique way of speaking.
qh3a@ (verbatim response to Q.H3a)	DELETED	Exclude verbatim responses to avoid disclosing a unique way of speaking.
qz1@ (other-specify verbatim responses to Q.Z1)	DELETED	Exclude verbatim responses to avoid disclosing a unique way of speaking.
qz2@ (other-specify verbatim responses to Q.Z2)	DELETED	Exclude to avoid disclosing a unique religious background, which in combination with other responses could violate confidentiality
qz4	VALUES (1) 1923-1979 (2) 1980-1989 (3) 1990-2000 (4) 2001-2002 (5) 2003-2007 (9) "Don't know/Refused"	Collapse years into categories to avoid disclosing unique personal background information