"NATIONAL SURVEY OF MORTGAGE ORIGINATIONS"

OMB NUMBER 2590-0012

SUPPORTING STATEMENT

The Federal Housing Finance Agency (FHFA or the Agency) is seeking approval for a three-year extension of the Paperwork Reduction Act (PRA) clearance for the National Survey of Mortgage Originations (NSMO). OMB has assigned the NSMO control number 2590-0012, which is due to expire on April 30, 2020. The NSMO is a recurring quarterly survey of individuals who have recently obtained a loan secured by a first mortgage on single-family residential property. The survey questionnaire is sent to a representative sample of approximately 6,000 recent mortgage borrowers each calendar quarter and consists of 96 multiple choice and short answer questions designed to obtain information about borrowers' experiences in choosing and in taking out a mortgage. The NSMO is sponsored by FHFA and is one component of the National Mortgage Database Program, an ongoing joint effort of FHFA and the Consumer Financial Protection Bureau (CFPB) (for PRA purposes, the NSMO is sponsored by FHFA). A copy of the survey questionnaire sent out in the first quarter of 2020 is included as Attachment 1.[1] FHFA is also seeking clearance to pretest future iterations of the survey questionnaire and related materials from time to time using focus groups.

A. JUSTIFICATION

1. Circumstances necessitating the collection of information

The NSMO is a component of the "National Mortgage Database" (NMDB) Program which is a joint effort of FHFA and the Consumer Financial Protection Bureau (CFPB). The NMDB Program is designed to satisfy the Congressionally-mandated requirements of section 1324(c) of the Federal Housing Enterprises Financial Safety and Soundness Act.[2] Section 1324(c) requires that FHFA conduct a monthly survey to collect data on the characteristics of individual prime and subprime mortgages, and on the borrowers and properties associated with those mortgages, in order to enable it to prepare a detailed annual report on the mortgage market activities of the Federal National Mortgage Association (Fannie Mae) and the Federal Home Loan Mortgage Corporation (Freddie Mac) for review by the appropriate Congressional oversight committees.

---

[1] In addition, copies of the questionnaire in both English and Spanish can be accessed at http://www.fhfa.gov/nsmo.

[2] 12 U.S.C. 4544(c).

Section 1324(c) also authorizes and requires FHFA to compile a database of otherwise unavailable residential mortgage market information and to make that information available to the public in a timely fashion.

As a means of fulfilling those and other statutory requirements, as well as to support policymaking and research regarding the residential mortgage markets, FHFA and CFPB jointly established the NMDB Program in 2012. The Program is designed to provide comprehensive information about the U.S. mortgage market and has three primary components: (1) the NMDB; (2) the NSMO; and (3) the American Survey of Mortgage Borrowers (ASMB).

The NMDB is a de-identified loan-level database of closed-end first-lien residential mortgage loans that is representative of the market as a whole, contains detailed loan-level information on the terms and performance of the mortgages and the characteristics of the associated borrowers and properties, is continually updated, has an historical component dating back to 1998, and provides a sampling frame for surveys to collect additional information. The core data in the NMDB are drawn from a random 1-in-20 sample of all closed-end first-lien mortgage files outstanding at any time between January 1998 and the present in the files of Experian, one of the three national credit repositories. A random 1-in-20 sample of mortgages newly reported to Experian is added each quarter.

The NMDB draws additional information on mortgages in the NMDB datasets from other existing sources, including the Home Mortgage Disclosure Act (HMDA) data that are maintained by the Federal Financial Institutions Examination Council (FFIEC), property valuation models, and administrative data files maintained by Fannie Mae and Freddie Mac and by federal agencies.

The NSMO was developed to complement the NMDB by providing critical and timely information on newly originated mortgages directly from the borrower. Such information—not available from other sources— provides information on the borrowers' experiences with the mortgage origination process. In particular, the survey questionnaire is designed to elicit directly from mortgage borrowers information on the characteristics of the borrowers and on their experiences in finding and obtaining a mortgage loan, including: their mortgage shopping behavior; their mortgage closing experiences; their expectations regarding house price appreciation; and critical financial and other life events affecting their households, such as unemployment, large medical expenses, or divorce. The survey questions do not focus on the terms of the borrowers' mortgage loans because these fields are available in the Experian data. However, the NSMO collects a limited amount of information on each respondent's mortgage to verify that the Experian records and survey responses pertain to the same mortgage.

The NSMO has been conducted quarterly since 2014, with the most recent wave—wave 25 (Quarter 1 of 2020)—having been mailed on February 3, 2020. Each wave of the NSMO is sent to the primary borrowers on about 6,000 mortgage loans, which are drawn from a simple random sample of the 80,000 to 100,000 newly originated mortgage loans that are added to the National Mortgage Database from the Experian files each quarter (at present, this represents an approximately 1-in-15 sample of loans added to the National Mortgage Database and an

approximately 1-in-300 sample of all mortgage loan originations). By contract with FHFA, the conduct of the NSMO is administered through Experian, which has subcontracted the survey administration through a competitive process to Westat, a nationally recognized survey vendor.[3] Westat also carries out the pre-testing of the survey materials.

FHFA also obtains data from a separate survey, the ASMB, which is also based on the NMDB. While the NSMO solicits information on newly originated mortgages the ASMB focuses on borrowers' experience with maintaining existing mortgages. This includes their experience maintaining mortgages under financial stress, their experience in soliciting financial assistance, their success in accessing federally sponsored programs designed to assist them, and, where applicable, any challenges they may have had in terminating a mortgage loan.[4]

2. Use of data

FHFA views the NMDB Program as a whole, including the NSMO, as the monthly "survey" that is required by section 1324 of the Safety and Soundness Act. Core inputs to the NMDB, such as a regular refresh of the Experian data, occur monthly, though NSMO itself does not. In combination with the other information in the NMDB, the information obtained through the NSMO is used to prepare the report to Congress on the mortgage market activities of Fannie Mae and Freddie Mac that FHFA is required to submit under section 1324, as well as for research and analysis by FHFA and CFPB in support of their regulatory and supervisory responsibilities related to the residential mortgage markets. The NSMO is especially critical in ensuring that the NMDB contains uniquely comprehensive information on the range of nontraditional and subprime mortgage products being offered, the methods by which these mortgages are being marketed and the characteristics—and particularly the creditworthiness—of borrowers for these types of loans.

In November 2018, FHFA and CFPB released the first loan-level dataset collected through the NSMO for public use. The first release contained data for mortgages originated in 2013 through 2017. An updated version of the dataset that added 2017 mortgage originations was released on February 20, 2020. This dataset provides a resource for research and analysis by federal agencies, by Fannie Mae and Freddie Mac, and by academics and other interested parties outside of the government.[5]

---

[3] The Fair Credit Reporting Act, 15 U.S.C. § 1681 *et seq.*, requires that the survey process, because it utilizes borrower names and addresses drawn from credit reporting agency records, must be administered through the credit repository (in this case Experian) in order to maintain consumer privacy.

[4] OMB has assigned the ASMB control no. 2590-0015, which expired on July 31, 2019. The ASMB was conducted annually from 2016 through 2018, but was not conducted in 2019. FHFA expects to conduct the survey again in late 2020.

[5] The February 2020 NSMO public use dataset can be accessed here: *https://www.fhfa.gov/nsmodata*.

FHFA is also seeking OMB approval to continue to conduct cognitive pre-testing of the survey materials. The Agency uses information collected through that process to assist in drafting and modifying the survey questions and instructions, as well as the related communications, to read in the way that will be most readily understood by the survey respondents and that will be most likely to elicit usable responses. Such information is also used to help the Agency decide on how best to organize and format the survey questionnaires.

3. Use of information technology

The NSMO uses machine-readable paper questionnaires in English and also gives recipients the option of completing the survey online in either English or Spanish. At first, approximately one third of the overall surveys were completed online, but the share of online surveys has approached 50 percent in recent waves. Completed paper questionnaires are scanned and the responses are automatically uploaded into the electronic National Mortgage Database.

4. Efforts to identify duplication

As explained above, the majority of data included in the National Mortgage Database is drawn from existing sources—primarily, the consumer credit database maintained by Experian; the HMDA data released by FFIEC; and administrative data in the possession of FHFA, its regulated entities, and other federal agencies. As described under Item #1, the NSMO questionnaire is designed to obtain critical and timely information that is not available from existing sources. The survey obtains this information directly from borrowers, who are likely to be the most reliable and accessible—and, in some cases, the only—source for this information.

5. Impact on small entities

The information collection will not have a significant economic impact on a substantial number of small entities. The survey recipients are individuals only.

6. Consequences of less frequent collection and obstacles to burden reduction

Section 1324 of the Safety and Soundness Act requires that FHFA undertake a survey of mortgage markets on a monthly basis.[6] While the performance data on existing mortgages in the National Mortgage Database is pulled from the Experian database on a monthly basis, newly originated mortgages are added to the National Mortgage Database on a quarterly basis. The NSMO questionnaires are sent to a random sample of borrowers that originated their mortgages in the year and quarter that corresponds to the quarterly draws of newly originated mortgages from the Experian database. One important purpose of the survey is to monitor loan origination trends. While monthly housing surveys would provide the optimal feedback regarding these trends, FHFA believes that quarterly surveys are sufficient.

---

[6] *See* 12 U.S.C. § 4544(c).

7. Circumstances requiring special information collection

There are no special circumstances that require FHFA to conduct the information collection in a manner inconsistent with the guidelines provided in this Item 7.

8. Solicitation of comments on information collection

In accordance with the requirements of 5 CFR 1320.8(d), FHFA published a request for public comments regarding this information collection in the *Federal Register* on December 10, 2019.[7] The 60-day comment period closed on February 10, 2020. FHFA received no comment letters.

9. Provision of payments or gifts to respondents

Until recently, survey recipients received a cash payment of five dollars as an inducement to complete and return the NSMO questionnaire. Recipients who failed to respond to the first two survey solicitations also received an additional cash inducement of five dollars. Recent waves of the survey, however, have shown a slow but steady decline in the response rate, a problem facing many other surveys like NSMO. In response, FHFA has been experimenting with the cash incentives.

In wave 22, one half of the usual 6,000 borrowers were randomly selected to receive a ten-dollar cash incentive with the first mailing instead of the typical five-dollar incentive. In wave 25, all borrowers were sent an initial incentive of ten dollars. In terms of the second incentive, one random half of the non-respondents were sent the normal five-dollar cash incentive as in previous waves and the other random half were sent a letter informing them that they will be sent a twenty-dollar incentive upon completion of the survey. In wave 26 (to be sent in Quarter 2 of 2020), one random half of the borrowers will receive an initial incentive of five dollars and the other random half will receive an initial incentive of ten dollars, but all non-respondents will receive a letter informing them that they will be sent a twenty-dollar incentive upon completion of the survey. Data collected during these experiments will be used to make decisions about the NSMO methodology in 2020.

Each cognitive pre-testing participant may receive approximately fifty dollars as an incentive payment.

10. Assurance of confidentiality

With respect to the confidentiality of survey responses, the cover letter that accompanies each NSMO questionnaire contains the following statement:

> This survey is voluntary, and we ask that you not identify yourself in any way
> when you return your questionnaire in the enclosed postage-paid return

---

[7] *See* 84 FR 67447 (Dec. 10, 2019). A copy of the 60-day Notice is included as Attachment 2.

envelope. The code numbers on the survey are there to aid in processing and keep track of returned surveys. No names or other identifying information is ever included in the data.

The questionnaire itself contains a statement, required by the Privacy Act,[8] informing recipients that "[s]ubmission of the survey authorizes FHFA to collect the information provided and to disclose it as set forth" in the current System of Records Notice (SORN) for the National Mortgage Database.[9] The questionnaire also instructs recipients not to include their names or addresses when completing the questionnaire.

Section 1324 of the Safety and Soundness Act authorizes FHFA to modify the mortgage data released to the public as necessary to ensure that it contains no "representation of information that permits the identity of a borrower to which the information relates to be reasonably inferred by either direct or indirect means."[10] For each sampled loan and its associated borrower(s), Experian provides its survey subcontractor, Westat, with the identifying information it needs to administer the survey. However, the data on borrowers and loans that is accessible to FHFA, CFPB, and any other authorized user of the National Mortgage Database, including data obtained through the NSMO, does not include any direct identifying information such as borrowers' names, addresses, or Social Security numbers or the name of any financial institution.

Westat mails a survey questionnaire to the borrower(s) on each sampled mortgage loan at the property address associated with that mortgage. It then uses an encrypted key to track the surveys so that it can compile and maintain the survey opt-out list and identify non-responders to whom it must send follow-up correspondence. All returned questionnaires and any non-delivered mail are sent directly to Westat, not to FHFA, CFPB, or Experian. To maintain the de-identified nature of the data and the confidentiality of the survey responses, Westat purges all responses of any identifying information before providing the collected information to FHFA's National Mortgage Database Program staff for further processing (which is described in Part B of this Supporting Statement).

Similarly, while Westat knows the identity of the cognitive pre-testing participants, that information is not conveyed to FHFA and is not included in the National Mortgage Database in any form.

---

[8] 5 U.S.C. § 552a.

[9] *See* 80 FR 52275 (Aug. 28, 2015); 81 FR 95595 (Dec. 28, 2016). Copies of the current SORN for the National Mortgage Database and a subsequent revision are included as Attachment 3.

[10] *See* 12 U.S.C. §§ 4544(c)(3), (4).

11. Questions of a sensitive nature

Of the 96 questions on the most recent iteration of the survey questionnaire, approximately 20-25 might be considered to be of a sensitive nature by particular borrowers. Questions that FHFA has identified as potentially sensitive include those requesting information on loan terms, purchase price, household income and other sources of funds, employment status, level of education, age, sex, race, ethnicity, and marital status.

Each of those questions is designed to elicit information that FHFA is required by statute to collect or that is otherwise essential to fulfilling the purposes of the NSMO and the National Mortgage Database Program as a whole. While FHFA understands that some survey recipients will be reluctant to answer questions about these potentially sensitive topics, the Agency believes that others will look upon doing so as an opportunity to express themselves about issues of concern to them.

12. Estimates of the hour burden of the information collection

This information collection comprises two components: (1) conducting the survey; and (2) pre-testing survey questionnaires and related materials through the use of cognitive testing. FHFA estimates that the total annualized hour burden imposed upon members of the public by this information collection will be 12,030 hours: 12,000 hours associated with conducting the survey and 30 hours associated with pre-testing the survey materials. Because the survey recipients and cognitive testing participants are individuals only, there are no hourly costs associated with the burden estimates. The overall burden estimates are based on the following calculations:

1) Conducting the Survey

The estimated annualized hour burden associated with conducting the NSMO is 12,000 hours. The NSMO questionnaire will be sent to 6,000 recipients quarterly. Although, based on historical experience, the Agency expects that only 20 to 30 percent of those surveys will be returned, it has assumed that all of the surveys will be returned for purposes of this burden calculation. Based on the reported experience of respondents to prior NSMO questionnaires, FHFA estimates that it will take each respondent 30 minutes to complete the survey, including the gathering of necessary materials to respond to the questions.

Recipients read and complete survey questionnaire and return the completed form to the survey subcontractor:

- Completion time per recipient: 0.5 hours
- Survey mail-outs annually: 4
- Recipients per survey: 6,000
- Total recipients annually: 24,000
- Total hours annually: 12,000 hours

2) Pre-Testing of Survey Materials

The estimated annualized hour burden associated with the pre-testing of the survey materials is 30 hours.

Selected individuals participate in cognitive testing to pre-test the survey questionnaire and related materials:

- Time per participant: 1 hour
- Total participants annually: 30
- Total hours annually: 30 hours

13. Estimated total annualized cost burden to respondents

FHFA has not identified any costs to respondents.

14. Estimated cost to the federal government

The estimated annual burden to the federal government is $732,000 and 400 hours, calculated as follows:

FHFA analyst embeds NSMO data into a query-based electronic database and carries out data cleaning, imputation, and non-response bias weighting:

- Processing time per survey: 100 hours
- Total surveys annually: 4
- Total hours: 400
- Hourly rate: $80 (includes salary, benefits, and overhead)
- Total cost: $32,000

In addition, approximately $700,000 will be paid annually to the contractor hired to conduct the surveys. Of this, approximately $250,000 will be attributable to the cash incentive payments to survey recipients; approximately $250,000 will be for printing and assembly costs; approximately $100,000 will be for postage costs; and approximately $100,000 will be for other fixed costs.

$32,000 (hourly cost) + $700,000 (paid to subcontractor) = $732,000.

15. Reasons for change in burden

The estimated burden has not changed.

16. Plans for tabulation, statistical analysis and publication

On April 18, 2018, FHFA and CFPB released aggregate data from the 6,285 responses pertaining to borrowers who obtained a mortgage in 2016 and whose loans were reported to Experian.[11] The data provide an overview of the mortgage market and borrowers' experiences in 2016. These unweighted responses were about one-third of the sample drawn from mortgages originated in 2016. On February 20, 2020, FHFA and CFPB released additional loan-level data for public use collected through the NSMO.[12] The analytical techniques used in preparing the data are discussed in detail in Part B of this Supporting Statement.

Going forward, FHFA and CFPB plan to publish similar data annually, with each presenting NSMO data pertaining to borrowers who obtained a mortgage during a particular calendar year. The agencies expect to release reports on NSMO data for loans originated in 2018 and subsequent years approximately 24 months after the end of the calendar year to which the data pertains.

NSMO data releases and technical documentation are made available on FHFA's public website at: https://www.fhfa.gov/nsmodata.

17. If seeking approval to not display the expiration date for OMB approval of the information collection, explain the reasons why display would be inappropriate

FHFA will display the expiration date for OMB approval.

18. Explain each exception to the topics of the certification statement identified in "certification for paperwork reduction act submission."

There are no exceptions to the topics of the certification statement identified in the "Certification for Paperwork Reduction Act Submission."

---

[11] The April 2018 release is available at:
https://www.fhfa.gov/PolicyProgramsResearch/Programs/Documents/NMDB-technical-report_6pt0_041818.pdf

[12] The February 2020 release is available at: https://www.fhfa.gov/DataTools/Downloads/Documents/NSMO-Public-Use-Files/NSMO-Technical-Documentation-20200220.pdf.

B. COLLECTIONS OF INFORMATION INVOLVING STATISTICAL METHODS

Question 1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

The NSMO represents a universe of 6 to 9 million recently originated closed-end first-lien mortgage loans on single-family residential property that are reported to Experian annually. A 1-in-20 simple random sample of those mortgages are added to the National Mortgage Database each quarter, resulting in approximately 80,000 to 100,000 loans being added quarterly and 320,000 to 400,000 loans being added annually. For each of the twelve quarterly waves of the NSMO that have been mailed to date, FHFA has selected an approximately 1-in-15 random sample of loans from one or more recent quarterly updates of the National Mortgage Database. This represents an approximately 1-in-300 sample of the universe.

For each quarterly survey FHFA randomly selects a sample of approximately 6,000 mortgage loans from those reported to Experian during the preceding quarter and newly added to the National Mortgage Database, with the additional conditions that the loans must have been reported to Experian within a year of origination and that the borrowers must not have been selected for an earlier NSMO survey. As of this writing, the NSMO survey is currently on wave 25 (mailed on February 3, 2020). Table 1 shows the survey field periods to date.

Table 1. Survey Field Periods			
Wave	**Survey Field Dates**	**Calendar Quarter**	**Surveys Mailed**
1	April to June 2014	2014 Quarter 1	15,000
2	June to August 2014	2014 Quarter 2	3,000
3	August to November 2014	2014 Quarter 3	5,992
4	November 2014 to February 2015	2014 Quarter 4	5,795
5	February to May 2015	2015 Quarter 1	5,925
6	May to August 2015	2015 Quarter 2	4,428
7	August to November 2015	2015 Quarter 3	7,352
8	November 2015 to February 2016	2015 Quarter 4	5,913
9	February to May 2016	2016 Quarter 1	5,907

10	May to August 2016	2016 Quarter 2	5,885
11	August to November 2016	2016 Quarter 3	5,904
12	November 2016 to February 2017	2016 Quarter 4	5,919
13	February to May 2017	2017 Quarter 1	5,910
14	May to August 2017	2017 Quarter 2	5,804
15	August to November 2017	2017 Quarter 3	5,809
16	November 2017 to February 2018	2017 Quarter 4	5,707
17	February to May 2018	2018 Quarter 1	5,755
18	April to July 2018	2018 Quarter 2	5,773
19	August to November 2018	2018 Quarter 3	5,759
20	November 2018 to February 2019	2018 Quarter 4	5,770
21	February to May 2019	2019 Quarter 1	5,746
22	May to August 2019	2019 Quarter 2	5,720
23	August to November 2019	2019 Quarter 3	5,737
24	November 2019 to February 2020	2019 Quarter 4	5,676
25	February to May 2020	2020 Quarter 1	5,698
Total			**151,884**

As shown in Table 2, NSMO typically samples about 6,000 new mortgage originations each quarter. Over 23 waves for which FHFA has received data from Experian, nearly 31 percent of the surveys mailed to borrowers of sampled mortgages were completed and 65 percent of completed surveys were received by mail.

Table 2. Survey Samples and Returns								
Wave	Average Sampling Weight	Surveys Mailed	Postal Non-Delivery	Surveys Delivered	Surveys Completed				Opt-Out
					Total	By Mail	Online English	Online Spanish	
1	464.21	15,000	218	14,782	5,793	4,410	1,360	23	169
2	296.14	3,000	37	2,963	1,076	858	214	4	31
3	280.96	5,992	110	5,882	2,073	1,534	524	15	40
4	263.63	5,795	86	5,709	2,020	1,496	514	10	53

5	247.32	5,925	126	5,799	2,089	1,567	520	2	39
6	238.92	4,428	38	4,390	1,489	1,133	353	3	31
7	296.64	7,352	147	7,205	2,494	1,744	744	6	39
8	326.97	5,913	99	5,814	1,899	1,305	587	7	24
9	292.31	5,907	155	5,752	1,824	1,230	584	10	42
10	253.27	5,885	98	5,787	1,765	1,148	607	10	36
11	278.27	5,904	172	5,732	1,733	1,097	627	9	21
12	343.76	5,919	167	5,752	1,778	1,078	687	13	18
13	363.21	5,910	127	5,783	1,885	1,197	675	13	32
14	318.55	5,804	107	5,697	1,681	1,085	588	8	21
15	270.61	5,809	136	5,673	1,537	765	760	12	24
16	305.24	5,707	164	5,543	1,507	757	738	12	26
17	304.31	5,755	112	5,643	1,647	879	762	6	45
18	262.93	5,773	163	5,610	1,536	812	711	13	32
19	266.84	5,759	242	5,517	1,464	760	695	9	29
20	284.5	5,770	206	5,564	1,396	762	627	7	11
21	266.12	5,746	251	5,495	1,511	777	719	15	17
22	213.35	5,720	219	5,501	1,405	757	630	18	29
23	262.92	5,737	235	5,502	1,236	647	579	10	29
Total	303.25	140,510	3415	137,095	42,838	27,798	14,805	235	838
Percent of Mailed Surveys		*100.0%*	*2.4%*	*97.6%*	*30.5%*	*19.8%*	*10.5%*	*0.2%*	*0.6%*
Percent of Completed Surveys					*100.0%*	*64.9%*	*34.6%*	*0.5%*	*NA*

In 2014, the first year of the survey, a modified version was conducted for the first three waves in April, June, and September. Wave 1 (April) included a sample of 15,000 mortgages. This was a catch-up period to cover mortgages originated in 2013 and newly reported to Experian in the archives for June, September, and December 2013. Wave 2 (June) included 3,000 surveys and was for mortgages that were originated in 2013 and newly reported to Experian between January and March 2014. For wave 3 (August), Westat mailed out 6,000 surveys representing mortgages that were originated in 2013 and reported to Experian between March and June 2014 within a year of origination as well as any mortgages originated in 2014 and reported to Experian between January and June 2014.

Wave 4, mailed in November 2014, was the first sample that is comparable to subsequent surveys. It comprised sample mortgages newly reported to Experian in the most recent quarter (July to September 2014) that was reported within a year of origination. It is also the first wave for which Experian eliminated potential sample cases deemed to not have legitimate addresses or names prior to mailing. Other than slight changes to two questions, the questionnaire was unchanged from prior waves. This same questionnaire was used for wave 5.

Initial analysis of data from the first four waves of the survey suggested that respondents may have frequently misunderstood or misinterpreted some of the questions, prompting major revisions to the questionnaire for part of wave 6 and all of wave 7 (users should be aware of these interpretation inconsistencies when using data from the earlier waves). For wave 6, surveys for mortgages that were originated in 2014 were mailed on the established schedule and using the original questionnaire; surveys for mortgages originated in 2015 were held back to be mailed with wave 7, using the new questionnaire.

Wave 7 consisted of three samples drawn independently. The first were those 1,236 respondents selected for wave 6 with loans originated in 2015. The second were 4,981 respondents with mortgages newly reported to Experian between April and June 2015 (the normal quarterly sample frame). Finally, a special sample of 1,142 borrowers residing in "remote rural" counties as defined using a USDA criterion with 2014 loan originations reporting to Experian within a year of origination was selected. Each subsample was assigned a different sample weight. All subsequent waves of the survey have included only the regular sample, mailed on-schedule.

Returned questionnaires and online responses were evaluated to determine the set of usable responses. Table 3 summarizes the results of this analysis through the 21 waves which have been completely processed and indicates the four criteria for rejecting a completed questionnaire.

Table 3. Usable Survey Responses							
Survey Wave	Returned	Duplicate or Ineligible	Answered No to Q1	Did Not Finish Survey	Wrong Loan	Usable	Weighted Usable*
1	5,793	84	738	127	216	4,628	6,871,209
2	1,076	15	84	16	38	923	875,467
3	2,073	37	108	36	59	1,833	1,659,752
4	2,020	164	88	46	64	1,658	1,395,466
5	2,089	37	81	46	62	1,863	1,443,963
6	1,489	116	69	29	50	1,225	991,516
7	2,494	65	144	78	98	2,109	2,108,874
8	1,899	42	73	28	59	1,697	1,900,299

9	1,824	38	69	27	37	1,653	1,701,989
10	1,765	59	84	40	58	1,524	1,432,246
11	1,733	41	92	38	38	1,524	1,602,998
12	1,778	58	102	49	50	1,519	1,995,025
13	1,885	48	103	52	54	1,628	2,107,377
14	1,681	50	66	52	44	1,469	1,817,181
15	1,537	30	140	78	33	1,256	1,544,244
16	1,507	27	116	70	26	1,268	1,713,007
17	1,647	25	127	64	34	1,397	1,723,003
18	1,536	18	117	60	34	1,307	1,500,016
19	1,464	23	108	61	31	1,241	1,512,449
20	1,396	22	116	66	25	1,167	1,617,383
21	1,511	29	152	82	23	1,225	1,499,320
Total	*40,197*	*1,028*	*2,777*	*1,145*	*1,133*	*34,114*	*39,012,784*
Percent of Mailed Surveys	*31.1%*	*0.8%*	*2.2%*	*0.9%*	*0.9%*	*26.4%*	*NA*

* The weighted usable total excludes the remote rural sample in wave 7.

The first category of unusable surveys comes from respondents whose sample loans were ultimately removed from the NMDB after the survey had been executed, either because the loans were deemed to have duplicate trade lines and did not meet the criteria for remaining in the NMDB, or because the sample loan was determined to be a second and not a first mortgage. In some instances, the survey response itself led to the removal, as margin notes or comments indicated that the loan was a second lien. This was a particular problem in wave 4.

The second category is a "no" response to the first question (Q1). Q1 is used as a screener question to confirm that the survey respondent took out a mortgage during the reporting period as suggested by Experian's records. In wave 1, a surprisingly high number of respondents (738) said that they had not taken out a mortgage. An analysis of these responses suggests that many people did not consider a refinance a "new" mortgage. Consequently, in wave 2, the wording of Q1 was changed to add the phrase "including any mortgage refinances." With this change, the share of "no" responses to Q1 decreased from 13 percent to 8 percent.

The third category eliminates breakoffs, defined as questionnaires for which the respondent did not answer almost all questions from the middle of the survey through the end, or answered less than 50 percent of the questions overall.

The fourth category is for respondents who provided information on the wrong loan. Although the sampling frame is tied to a particular loan associated with the borrower, the questionnaire does not refer explicitly to that loan. Instead, respondents who have taken out multiple loans during the reference period are asked to report on the "most recent," which, in some instances, has not been the sample loan. This was a particular problem in wave 1 which, as a "catch up" survey, had a relatively long reference period. Also, some respondents who have refinanced their mortgage have reported on the original home purchase mortgage rather than the refinance. Finally, it appears that in a few instances the survey has been sent to the wrong person, with answers bearing no resemblance to the sample loan features as characterized by Experian records. In each of those circumstances the survey response was removed from the data set used for analysis.

Overall, for the first 21 waves 34,114 usable responses were obtained from 40,197 returned surveys. This resulted in a usable response rate of 26.4% out of the mailed-out questionnaires.

Question 2. Describe the procedures for the collection of information, including:
- **Statistical methodology for stratification and sample selection,**
- **Estimation procedure,**
- **Degree of accuracy needed for the purpose described in the justification,**
- **Unusual problems requiring specialized sampling procedures, and**
- **Any use of periodic (less frequently than annual) data collection cycles to reduce burden.**

Data for the NSMO is collected through a single-blind mail survey format. The survey's sample selection is discussed earlier in Question 1. The NSMO is a simple, random sample of mortgage originations and is not stratified. Alternatives designed for stratifying, clustering, or cut-off samples were not considered. While subpopulations are of interest, the key purpose of the survey is to gather information on the origination characteristics of the sampled borrowers. No estimation procedures are needed in interpreting the survey responses. Similarly, no hypotheses are being tested and no unusual problems exist that require specialized sampling procedures.

Question 3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

Methods Used to Maximize Response Rates

The methodology for National Mortgage Database, including the NSMO, was initially based upon methodology that was developed and tested during a series of three surveys funded and

carried out by Freddie Mac between 2010 and 2012. After achieving a response rate of just 12 percent on its initial pilot survey in 2010, Freddie Mac retained Dr. Don A. Dillman of Washington State University, a leading expert in mail survey methods, to provide input on ways to maximize response rate. Dr. Dillman provided recommendations in three critical areas: (1) the execution/implementation of the survey; (2) the communications package; and (3) the questionnaire content and format. Freddie Mac adopted Dr. Dillman's recommendations in these areas and, in the second and third pilot surveys, achieved response rates of 60 and 45 percent, respectively. FHFA adopted those recommendations for the NSMO and continues to consult with Dr. Dillman, among others, in developing the survey.

One important recommendation that FHFA adopted for the NSMO was to have four planned mail contacts with the survey recipients. The first contact consists of the questionnaire, an upfront monetary incentive (ten dollars in the current wave), and a cover letter. The second contact, a reminder letter sent to all recipients, occurs in the second week of implementation. The third contact, in the fifth week, is sent only to non-responders and includes a second reminder letter, another copy of the questionnaire, and another incentive. The final contact, a third reminder letter, is sent in the seventh week to non-responders only. A due date for returning the survey questionnaire is included in the last mailing, which closes the communication loop with all survey recipients.[13]

Recent waves of the survey have shown a slow but steady decline in the response rate, a problem facing many other surveys like NSMO. During waves 22 through 25, experiments with survey methodology have been conducted to address the issue of declining response rate. In wave 22, one half of the usual 6,000 borrowers were randomly selected to receive a ten-dollar cash incentive with the first mailing instead of the five-dollar incentive the other half received (and which had been the prevailing rate to that point). In waves 23 and 24, one half of the borrowers received revised cover and reminder letters while the other half received the originals. (Edits were made to the revised letters between waves 23 and 24 to boost response rates.)

In wave 25, all borrowers were sent an initial incentive of ten dollars. In terms of the second incentive, one random half of the non-respondents are being sent the normal five-dollar cash incentive as in previous waves and the other random half are being sent a letter informing them that they will be sent a twenty-dollar incentive upon completion of the survey. In wave 26 (to be sent in Quarter 2 of 2020), one random half of the borrowers will receive an initial incentive of five dollars and the other random half will receive an initial incentive of ten dollars, but all non-respondents will receive a letter informing them that they will be sent a twenty-dollar incentive upon completion of the survey. Data collected during these experiments will be used to make decisions about the NSMO methodology in 2020.

FHFA also adopted two of Dr. Dillman's recommendations regarding the communications package. First, all communications have a friendly tone and reflect a personal and sincere

---

[13] Copies of the four letters (in both English and Spanish) comprising the NSMO communications package used for the February 2020 survey mailing are included as Attachment 4.

request for help. All correspondence is signed by a senior official of both FHFA and CFPB and includes contact information for authenticity. Second, each questionnaire is mailed in a plain white envelope so as not be mistaken for "junk mail."

To further increase response rates, survey recipients are given the option of completing the survey online in either English or Spanish. The first mailing contains an insert, in both English and Spanish, which informs recipients of these options and provides the web addresses to access the appropriate electronic versions of the survey.

Data editing

The survey responses, once delivered to FHFA's National Mortgage Database Program staff, are subjected to thorough editing and review. The initial phase consists of standard data editing— correcting numbers reported in the wrong units, changing answers in responses based on margin notes and comments, assigning responses for questions with open-ended "other" responses, dealing with multiple responses to a question that calls for only one response, and deciding how to handle situations where respondents followed the wrong skip pattern.

One advantage that the NSMO has over other surveys is the availability of credit and administrative data, much of which appears to be quite reliable. These data can be used to assist in the editing and imputation process. Four primary sources of such data are available in processing NSMO:

> (1) credit data from Experian on sample loans;
> (2) data collected by Experian from other data sources on the survey respondents, including loan servicers and data companies;
> (3) information from matches to administrative loan files (Fannie Mae, Freddie Mac, Federal Housing Administration, Department of Veteran Affairs, Rural Housing Services, and Federal Home Loan Banks); and
> (4) information for loans that could be matched to HMDA data (available through calendar year 2018 as of this writing).[14]

The credit and administrative data are used to determine which borrower in the Experian data corresponded to the respondent (and spouse/partner of the respondent) in the survey and to infer

---

[14] Merges with most administrative files are conducted behind a firewall at Experian using borrower name, address, date of birth and Social Security number to ensure the highest quality match accuracy (neither FHFA nor CFPB staff ever receive such information). However, merging of the NMDB data with the HMDA data and the Federal Home Loan Bank loan files have to rely on variables common to both datasets, including the original loan balance, the opening date of the mortgage and the general location of the property (census tract or state/county) but not property address or borrower name. Unfortunately, mortgage servicers report the billing address of the mortgage borrowers to Experian, but this is not necessarily the property address, particularly for mortgages on non-owner-occupied properties. Those, when converted to a census tract for matching the address, may be incorrect. Thus, HMDA merges are less accurate than those employing directly identifying information such as name and Social Security number because the latter are less reliant on address.

the loan the respondent had in mind when answering the survey. These data are also useful in determining if respondents correctly identified their loan as a home purchase loan or a refinance.

Imputation

After editing and cleaning the survey response data, missing responses are imputed using answers to related questions or statistical models estimated based on credit and administrative data and answers to other questions in the survey. Imputations are designed to replicate the level of inherent inconsistencies between related variables in the actual (non-imputed) responses by the respondents. Actual responses are generally not changed (except in cases where they are edited as described above). In order to preserve the original responses, the raw responses are retained with missing responses coded as such. A parallel set of variables ("X" variables) are constructed where all missing responses are imputed, and necessary responses are edited as described above. Each instance in which an X variable differs from original responses is recorded by a shadow variable ("J" variables) that indicates the method and reason why the change was made. Missing responses typically total about 3 to 5 percent of responses for most questions and only in a few instances were more than 10 percent. The X variables are not created when a directly comparable credit or administrative variable is available for all respondents (*e.g.*, loan amount, loan payment, number of co-signers) as comparable credit or administrative variables can be used in lieu of survey responses in analysis. Instead, Z variables are created in their place to indicate whether the respondents answered the question.

The initial set of imputations are based on inferences drawn from patterns of response. Patterns of missing responses sometimes provides an indication of how the respondent would have answered if they had taken the time to fill out all answers of a group. For example, one question reads, "how important were each of the following…" and provides choices of important or not important. Some respondents only mark "important" for the choices important to them. Other respondents might only mark choices that are not important. When all answers are in a group with only one side answered, the other answers are imputed as the opposite choice. For example, when a respondent only marks choices that are important, the missing questions were imputed as not important.

The survey skips do not always work for every respondent and some respondents miss the lead-in question. The answers to the lead-in question are often imputed based on actual answers to the follow-up questions. For example, one question reads, "how many different lenders/mortgage brokers did you end up applying to" and provides options for one to five. When a respondent chooses one, they skip the next question about reasons they applied to more than one. If the lead-in question was left blank, any yes answer to the follow-up is considered a reason to impute that they applied to more than one lender. All "no" answers to the follow-up questions means that they probably only applied to one lender. When the respondent skips both a lead-in and follow-up question, both are imputed with one of the imputation models.

Once these inferential imputations are taken care of, statistical models are used to impute the remaining missing answers. The most common type of question in NSMO provides a simple yes or no answer. A binomial logistic model provides an estimated probability of a yes answer. For

18

some questions, such as the number of lenders or brokers the respondent seriously considered, the answers are in a logical order. For these types of questions, an ordered logistic model is used to determine the probability of each answer. For other questions the order does not matter, and the answer choices are not related to the previous choice. For these questions, a multinomial logistic model is used, and the reference group is selected to be the most common answer. Again, the model produces a probability of each answer response. A random number is drawn with a different seed for every question and it is then compared to the probability of each response level. When the random number falls below the cumulative probability of an answer, that answer is used as the imputed response. This method injects some randomness to the imputed answers, but the goal is to provide a distribution of imputed answers that mimics the distribution of the answers where no imputation was necessary.

The dependent variable (y_i) in all the models used is a value for the missing answer. The vector of characteristics (x_i) can include information from the credit files or answers to survey questions. Key demographic variables (age, gender, education, ethnicity, and income) are imputed first. For these variables, high quality administrative data are generally available and can be used directly to impute a value for the X variable. For example, lender-reported information provides high quality data on age. Administrative data also provide reliable information on race, income, and interest rate. HMDA data also provide reliable information on race, income, and gender.

The initial statistical imputation models first use all the respondents who provided answers using a standard set of predictors to provide an initial imputation. The models use age, loan amount, credit score, loan type, education and income level. Once the initial imputation values are established, the models are enhanced for any predictor that provides a good fit to the models and these models use actual and imputed values from all respondents. The missing values are imputed statistically using an iterative process where each subsequent run of the model uses the actual responses and the imputed responses from the previous run. Iterating in this way ensures that correlations among the imputed values will better reflect correlations among observations where responses were available.

The regression runs always start with key variables first. As with the initial imputations, the first variables imputed are age, loan amount, credit score, loan type, education and income. The next level covered by the models imputes marital status, race, and ethnicity. The process then moves on to other questions and often follows the order of the survey instrument for less consequential questions. Lead-in questions are always imputed before the follow-up question to keep the follow-up imputations consistent with the lead-in question.

As the recursive models run, the coefficient of each predictor variable in each model is tracked and compared with values from the previous runs. The recursive runs are only stopped when the coefficients have settled down with minimal changes in the last few runs. This ensures that the recursive effect on each model has fed into all the predictions of imputed values and stabilized.

To find the best model for each imputation, the last recursive run is selected, and the actual response is subtracted from the predicted value of the response. The difference represents the

19

error term or the portion of the probability of a response that was not explained by the predictive variables. A large matrix of error terms is constructed, and the values are tested for correlation. Error terms with correlation coefficient of over 0.30 get explored as possible indicators of new predictor variables. Each year, new predictors are placed into the recursive model and the results are tested to see if the model improves. With improved models, the recursive runs are restarted until all the beta coefficients settle down again.

The final imputations rely on a further set of quality control checks. Conditional correlation tables of model residuals are constructed to identify any additional significant explanatory variables which may have been left out of individual equations. Further, imputed values of similar related variables are sometimes adjusted to ensure that the covariances among the imputed answers mirror that of the non-imputed responses.

Non-Response Weighting

There are several ways calculations based on the NSMO raw survey responses may not be representative of the population as a whole. First, as shown earlier in Table 2, the survey waves do not have the same sampling rates. Second, only about one-third of the sampled borrowers completed the survey. Commonly, in survey sampling, some individuals chosen for the sample are unwilling or unable to participate in the survey. Non-response bias is the bias that results when respondents differ systematically from non-respondents. A common method for mitigating possible non-response bias is to use weights to align the characteristics of respondents and the population more closely. This is known as "non-response weighting." Such weights are generally calculated from statistical models. Specifically, the non-response weights in NSMO are designed to "blow up" the usable sample to the total surveys mailed less duplicate and ineligible loans taken out of NMDB.

Often, little is known about survey non-responders, so the statistical models used to construct non-response weights are quite simplistic. Compared with many other surveys, however, NSMO has extensive credit and administrative data on both responding and non-responding borrowers that can be used to estimate non-response weights.

Sample non-response weights are estimated with logistic models separately for each sample wave and within a wave for loans with a single borrower versus those with multiple borrowers. The models estimate the probability of getting a usable response for each wave of the survey. The predictive equations have had pseudo-R-square values ranging from 0.0467 to 0.1654. The models for joint borrowers do better than those for single borrower. The largest pseudo-R-square values were for models estimated on data from wave 20 joint borrowers. Key predictive variables included are loan amount, borrower age, the income relied upon for underwriting, the combined loan-to-value ratio, an indicator of whether it was a home purchase or refinance loan, and the interest rate spread over the prevailing prime interest rate at origination. The models also control for credit score, for geography using Census Divisions, and for demographic characteristics on family composition, race, ethnicity, gender, and educational attainment.

The model's predicted probabilities of response were placed into 5 equal groups of 20 percent each. The average of the response rates from each of these five groups was used to calculate a response weight as the inverse of these five average rates. Once within-wave sample non-response weights are estimated, they are multiplied by the wave sample weight to provide an overall weight.

Table 4 demonstrates the effect of differential sampling weights for the first 21 waves. Column one shows the distribution among various demographic and loan categories of the raw survey responses. Column two provides the distribution using estimated overall weights. Finally, column three shows the average overall weight for each category.

Table 4. Overall Weights, 2013 - 2017 Originations (Waves 1-21)			
	Unweighted Percentage	Weighted Percentage	Average Weight
Loan Category			
Purchase	48.4%	49.6%	1,169
Refinance	51.6%	50.4%	1,112
	100.0%	100.0%	
Loan Size			
Less than $50,000	2.6%	2.6%	1,106
$50,000 to $99,999	14.2%	14.2%	1,139
$100,000 to $149,999	20.2%	20.4%	1,147
$150,000 to $199,999	17.6%	17.7%	1,142
$200,000 to $249,999	13.1%	12.9%	1,123
$250,000 to $299,999	9.8%	9.5%	1,112
$300,000 to $349,999	6.6%	6.6%	1,128
$350,000 to $399,999	4.9%	4.8%	1,115
$400,000 or more	10.9%	11.4%	1,197
	100.0%	100.0%	
Mortgage Term to Maturity			
Less than 15 Years	4.2%	3.7%	1,004
15 Years	16.6%	15.3%	1,050
Between 15 and 30 Years	6.4%	6.4%	1,043
30 Years or More	72.8%	74.6%	1,168
	100.0%	100.0%	
Loan to Value (LTV) Ratio			
Less than 75%	37.9%	35.1%	1,055
75% to 79%	11.4%	11.1%	1,109

80%	9.7%	9.3%	1,095
81% to 89%	10.0%	10.1%	1,155
90% or More	31.0%	34.4%	1,264
	100.0%	100.0%	
Respondent Credit Score			
Lower than 620	4.1%	5.8%	1,592
620 to 639	3.3%	4.3%	1,494
640 to 659	5.0%	6.2%	1,431
660 to 679	5.7%	6.6%	1,323
680 to 699	6.4%	7.2%	1,278
700 to 719	7.5%	8.2%	1,246
720 to 739	9.4%	10.0%	1,209
740 or Higher	58.6%	51.8%	1,006
	100.0%	100.0%	

Question 4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the main collection of information.

FHFA will use information collected from the cognitive testing participants to assist the Agency in drafting and modifying the survey questions and instructions, as well as the related communications, to read in the way that will be most readily understood by the survey respondents and that will be most likely to elicit usable responses. Such information will also be used to help the Agency decide how best to organize and format the survey questionnaire. A copy of the most recent version of FHFA's NSMO cognitive testing guidance document (or "Talk Track"), which was provided to Westat on January 29, 2020, is included as Attachment 5.

Question 5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

The names of and contact information for individual stakeholders from FHFA, CFPB, and Experian, including those who were consulted on statistical aspects of the design and who will

analyze the data, appear in the list included as Attachment 6. FHFA also consulted with the following:

Dr. Mick P. Couper
Survey Research Center and the Institute for Social Research
University of Michigan
426 Thompson Street
Ann Arbor, MI 48104
(734) 647-3577

Dr. Don A. Dillman
Department of Sociology and the Social & Economic Sciences Research Center
Washington State University
Pullman, WA 99164-4014
(509) 335-1511

The subcontractor hired by Experian to carry out the survey and the cognitive testing is:

Westat
1600 Research Blvd,
Rockville, MD 20850

List of Attachments:

1. NSMO questionnaire for 2020 Q1 (mailed February 3, 2020)
2. 60-day PRA Notice published at 84 FR 67447 (Dec. 10, 2019)
3. National Mortgage Database System of Record Act Notices:
    a. SORN published at 80 FR 52275 (Aug. 28, 2015)
    b. Revision to SORN published at 81 FR 95595 (Dec. 28, 2016)
4. NSMO communication package (in English and Spanish) for 2020 Q1
5. Cognitive testing "Talk Track" dated January 29, 2020
6. List of National Mortgage Database stakeholders