

TECHNICAL APPENDIX

This document serves as the technical appendix for the U visa report which can be found on the USCIS website at uscis.gov/data.

1.0 Research Design

Current USCIS electronic data systems do not capture, nor store, many highly pertinent data points on U visa petitioners and derivatives. Complete and reliable information on how many and which types of crimes have been certified by law enforcement officials on the Form I-918, Supplement B, U Nonimmigrant Status Certification (Form I-918B), or which agencies are certifying Form I-918 is not available electronically. Information from the Form I-192, Application for Advanced Permission to Enter as a Nonimmigrant, such as which inadmissibility grounds were waived, is also not stored electronically. Further, USCIS electronic data systems do not capture qualitative information such as the helpfulness of the victim to the prosecution or types of evidence submitted by the principal petitioner. Given the inherent limitations in USCIS electronic data systems, USCIS designed a mixed-methods research study to provide information on principals, derivatives, and filing trends in the U program. This study is comprised of analysis of the electronic data captured by USCIS data systems, data collected through a manual review of select hard copy petitions and associated Alien files, and information gleaned from interviews with subject matter experts in the U program.

1.1 Electronic data

Although USCIS electronic systems do not house all variables of interest, USCIS extracted and analyzed data housed in CLAIMS 3 for all U visa petitioners (principals and derivatives) who filed petitions from 2012 through 2018. Using electronic data provided general trends in the number of petitions received over time, approval and denial rates and some demographic information on principal and derivative petitioners. For example, information on gender, country of birth, and age can be derived from USCIS electronic data.

1.2 Manual File Review and Sampling

For the manual file review, the study team comprised of subject matter experts in U visas, benefit integrity, and/or research design. The study team could not review all U visa principal and derivative petitions submitted to USCIS in order to answer questions regarding this population. USCIS selected a stratified random sample of 1,108 U visa principal petitions from Fiscal Year 2012 through 2018. By using a representative sample, USCIS is able to estimate the occurrence of specific characteristics of principals and derivatives. The research team oversampled to ensure the minimum number of petitions would be available for review during the field work in Vermont.

The research team developed a structured data collection instrument for principals and one for derivatives designed to capture information from petitions to answer the research questions. The data collection instruments were pre-tested with U visa petitions located at the AAO in September 2018 and updates were made to the instruments.

The research team traveled to Vermont in November 2018 and manually reviewed 483 of the 1108 U visa principal petitions in the sample using the data collection instruments. Given the short amount of time in Vermont and the planned interviews of subject matter experts, the study team was cognizant that not all 1,108 petitions would be reviewed. The research team spent approximately two and a half work weeks reviewing petitions. In addition, the team collected data from 108 derivative A-files that were bundled with the principal petition. Following the review of the files, the research team entered information from the data collection instruments into a Microsoft Access database to facilitate analysis.

During the design of the study, the team decided to focus on gathering data from principal petitioners for a number of reasons. First, most of the pieces of information deemed higher priority are contained in the A-file of the principal petitioner (such as documentation of helpfulness, qualifying criminal activity, etc.). Second, the research team recognized that gathering a large number of files for this review would be extremely burdensome for a number of operational reasons. Given this operational burden, the research team decided to not identify a separate random sample of derivative petitions. Instead, researchers decided to review derivative petitions that were physically bundled with the principal petitions only.

Generally, derivative and principal petitions are kept together while pending an adjudicative action and separated once USCIS approves or denies the principal petition. In practice, this meant that, generally, researchers only reviewed derivative petitions when the principal petition was still pending an adjudicative outcome. Due to this, analysis of the derivative petitions was limited. At the time of the review, USCIS was adjudicating petitions received in mid-2014. Although this necessarily limited the analysis for derivative petitions from the random sample of principal petitions, the research team determined the trade-off in terms of reduced burden to the agency and the Vermont Service Center, in particular, was worth it.

1.3 Sample Design

A stratified random sample of 483 I-918 petitions adjudicated between 2012 and 2018 was selected. Those adjudicated before 2014 were sampled at half the rate of those adjudicated in more recent years. Furthermore, a sample of petitions that were pending as of the date of the data pull (April 18, 2018) was selected among those received between 2012 and 2018. These had a sampling rate one-third the rate of petitions adjudicated in recent years. Table 1 below shows petition counts from the sampling frame, and Table 2 shows the sampling rates relative to those of petitions adjudicated in recent years. The final sample size was 483, and Table 3 shows the distribution of expected sample counts. The design requirements leading to this distribution of sample is described below, following a discussion of subsampling in the field and measures of precision used in the requirements.

Appendix Table 1: Sampling Frame Counts: Outcome by Fiscal Year of Action*

	2012	2013	2014	2015	2016	2017	2018	Total
Approved	9,955	9,916	10,013	10,022	10,015	9,999	7,810	67,730
Denied	1,517	1,661	3,252	2,228	1,649	1,966	1,194	13,467
Pending	16	48	4,287	28,940	34,258	36,285	19,612	123,446
Total	11,488	11,625	17,552	41,190	45,922	48,250	28,616	204,643

* For pending petitions, the action date is the receipt date; otherwise, action date is the adjudication date

Appendix Table 2: Oversampling Factors: Outcome by Fiscal Year of Action*

	2012	2013	2014	2015	2016	2017	2018
Approved	.50	.50	1	1	1	1	1
Denied	.50	.50	1	1	1	1	1
Pending	.50	.50	.33	.33	.33	.33	.33

* For pending petitions, the action date is the receipt date; otherwise, action date is the adjudication date

Appendix Table 3: Expected Sample Counts: Outcome by Fiscal Year of Action*

	2012	2013	2014	2015	2016	2017	2018	Total
Approved	22	22	44	44	44	44	34	252
Denied	3	4	14	10	7	9	5	52
Pending	0	0	7	42	50	52	28	179
Total	25	25	65	95	101	104	68	483

* For pending petitions, the action date is the receipt date; otherwise, action date is the adjudication date. Stratification was accomplished by sorting and selecting petitions systematically, which can lead to a slight variance between the expected and realized sample sizes within each stratum.

1.3.1 Initial Oversample and Subsampling in the Field

The scope of this study allowed for a sample of approximately 500 to be studied, but 1,108 were initially selected using the same relative sampling rates as shown in Table 2. This oversample was sent to the Vermont Service Center, with the extra petitions selected in the event that some would be unavailable for review, although this did not occur. These petitions had been divided into ten replicate subsamples of about 110 petitions each, each randomly sorted. From these, analysts in the field chose four complete subsamples and a random selection of petitions from a fifth subsample, leading to a total of 483 in the final sample.

1.3.2 Measures of Precision

The sample was subject to design constraints on the precision of a proportion estimated from a stratified random sample, which describes most of the estimates provided in this report. The precision is measured as the half-width of a Wilson Score confidence interval for a binomial proportion that has been adjusted for complex sampling. The Wilson Score interval differs from the more commonly-applied normal approximation interval in that the standard error is not treated as a fixed value, but instead will have implicitly different values for the upper and lower

bounds of the interval. The Wilson Score interval has better coverage properties than the normal approximation interval, which means, for example, that a Wilson Score interval with a reported 90 percent confidence level will have a true confidence level closer to 90 percent than a normal approximation interval. For moderate-to-large sample sizes, the confidence intervals will match closely, and the normal approximation interval may be preferable for its simplicity. The scope of this study, and the constraints on sample size, will in some cases lead to Wilson Score intervals that are markedly different from the normal approximation interval suggesting their use may lead to a meaningful improvement. One characteristic of Wilson Score intervals is that they are often not symmetric around the estimated proportion, while the normal score interval always is.

As mentioned, the Wilson Score interval was adjusted for the complex design, specifically due to differences in sampling rates across petitions. The estimated proportion was a weighted estimate, with each petition's weight being equal to the inverse of its selection probability. This is referred to as a design weight, and leads to unbiased estimation of population totals and overall proportions. The effect of using weighted estimates was incorporated into the confidence intervals using design effects, which were calculated as the ratio of the sample size (483) and an estimate of the effective sample size from Kish¹. Estimates of the percentage approved, denied, and pending had a design effect equal to 1, while estimates of the percentage of petitions having a particular characteristic, which included both adjudicated and pending petitions in the estimation, had a design effect equal to 1.26.

The upper and lower bounds of an adjusted Wilson Score interval at a 90 percent confidence level are the solutions for p in the following equation:

$$p = \hat{p} \pm 1.645 \sqrt{\hat{p}(1 - \hat{p}) / (deff \cdot n)}, \text{ for an estimated proportion } \hat{p}, \text{ with sample size } n, \text{ design effect } deff, \text{ and } 1.645 \text{ being the z-score for a 90 percent confidence level.}$$

1.3.3 Design Requirements

An estimated approval rate calculated as the number of approved divided by the total number of adjudicated petitions will have a 90 percent Wilson Score confidence interval with a half-width no greater than four percent, assuming an 80 percent approval rate, and a design effect equal to one. Table 4 shows that these constraints are met with a sample size of 270 adjudicated petitions.

¹ The effective sample size of a complex design is the sample size needed from a simple random sample to achieve the same level of precision as the complex design; For a design with differential weighting, Kish's estimate of the effective sample size is $n_{eff} = (\sum_i w_i)^2 / \sum_i w_i^2$, where w_i is the estimation weight for petition i ; See: Kish, L. (1965), *Survey Sampling*, New York: John Wiley & Sons, Inc.

Appendix Table 4: Half-width of a 90 Percent Wilson Score Interval (Precision) by Sample Size, Assuming an Estimate of 80 Percent

Estimate	Sample Size	Lower Bound	Upper Bound	Half-Width
80%	230	75%	84%	4.3%
80%	240	75%	84%	4.2%
80%	250	76%	84%	4.2%
80%	260	76%	84%	4.1%
80%	270	76%	84%	4.0%
80%	280	76%	84%	3.9%
80%	290	76%	84%	3.9%
80%	300	76%	84%	3.8%

The percentage of petitions having a particular characteristic in any Fiscal Year 2014 or later will have a 90 percent Wilson Score confidence interval with a half width no greater than 10 percent, assuming the percentage is 50 percent, and a design effect equal to one. Table 5 shows that these constraints are met with a sample size of 65 petitions for each Fiscal Year 2014 through 2018. The sample size constraints resulting from the two design requirements are satisfied by the oversampling factors described earlier in the overview of the sampling design.

Appendix Table 5: Half-width of a 90 Percent Wilson Score Interval (Precision) by Sample Size, Assuming an Estimate of 50 Percent

Estimate	Sample Size	Lower Bound	Upper Bound	Half-Width
50%	50	39%	61%	11.3%
50%	55	39%	61%	10.8%
50%	60	40%	60%	10.4%
50%	65	40%	60%	10.0%
50%	70	40%	60%	9.6%
50%	75	41%	59%	9.3%
50%	80	41%	59%	9.0%
50%	85	41%	59%	8.8%

1.4 Semi-Structured Interviews

In order to better understand the vulnerabilities and fraud concerns within the U program, the research team also interviewed more than 20 subject matter experts across the Vermont Service Center.

1.5 FBI Record of Arrest and Prosecution (RAP) Sheet Analysis

In order to describe the arrest history of U visa principals and derivatives, USCIS analyzed RAP sheet data to determine if a U visa principal and derivatives received an IDENT response² which indicates that an Alien Number was associated with an arrest or apprehension. USCIS further analyzed the contents of the RAP sheets to determine the number and types of offenses associated with an arrest or apprehension of a U visa petitioner. USCIS analyzed CLAIMS 3 data and RAP sheet data to determine if the person who received an IDENT response was approved for a U visa. Variables like the offense type, charging date, and location of arrest were extracted from these text files, though other relevant variables such as outcome (i.e., whether charges were ultimately dropped, reduced, or led to convictions in court) could not be reliably extracted. In general, these variables were extracted from the text files based on syntax and keyword matching. Errors may result from the mining of these complex text files.

Note: an IDENT response indicates that an individual, in this case a U visa petitioner, was arrested or apprehended for a criminal offense or an immigration-related civil offense only. USCIS analyzed arrests and apprehensions as not all IDENT responses include disposition of the arrest or apprehension such as conviction, acquittal, dismissal or lessening of charges. Therefore, this analysis should be not be considered a comprehensive or complete determination of all arrests and apprehensions of U visa petitioners. For the purposes of this report, we limited the information presented in this report to persons who submitted a U visa principal or derivative petition between October 1, 2012 and June 2018.

2.0 Understanding the Data

In the findings section of the report, there are estimates of characteristics of petitioners and derivatives by individual fiscal year of receipt, where available, as well as for all years in the study (Fiscal Year 2012 through 2018).

Apart from the IDENT data discussed above, the data presented in this report are based on two sources: (1) electronic petition data stored in USCIS electronic data systems (such as CLAIMS 3)

² The Automated Biometric Identification System (IDENT) is the central DHS-wide system for storage and processing of biometric and associated biographic information for national security; law enforcement; immigration and border management; intelligence; background investigations for national security positions and certain positions of public trust; and associated testing, training, management reporting, planning and analysis, or other administrative uses. For the purposes of this analysis, IDENT provided electronic information from the petitioner's Record of Arrest and Prosecutions (RAP) sheet, where one was available. When USCIS submits fingerprints to the Federal Bureau of Investigation (FBI) for criminal history background checks, the FBI response is categorized as "IDENT" or "non-IDENT". When an individual has an IDENT response from the FBI, the results are returned on an Identity History Summary (IdHS; formerly known as a "RAP sheet"). For the purposes of this report, when we use the term "IDENT response", it is defined as an IDENT response from the FBI on an individual's Identity History Summary (or IdHS).

and (2) the manual review of the Alien files and U petitions. Data sources are listed in the source field under each graphic or table in the report.

Where the source lists “CLAIMS 3”, these data are taken from USCIS electronic sources and generally include the universe of U visa principal petitions regardless of outcome, unless stated in the table. For example, findings related to the overall number of petitions received, the outcome of the petitions, the number placed on the waiting list, approved class of admission for derivatives, gender, country of birth, etc. are derived from USCIS electronic data and are based on *all* petitions, or the universe of petitions, not a sample of petitions or subset of petitions. Generally, these data show all years of the program (Fiscal Years 2012 to 2018) except where explicitly noted in the table or graphic title and source notation.

Where the source line indicates “Analysis of Alien File (manual file review),” these findings are based on the random sample of petitions. Therefore, the findings from these data are *estimates* of the characteristic in the larger population of U visa petitioners who submitted a petition between Fiscal Years 2012 and 2018 and are based on the sample of petitions the study team reviewed. For brevity in the report, confidence intervals for all estimates have been excluded from the report. Where possible, USCIS presents estimates for characteristics by year of receipt. Where year over year estimates could not be derived due to small sample sizes, we present estimates for all years in aggregate (Fiscal Year 2012 to 2018).

Some data points in the report are restricted to very specific timeframes or excluded entirely because of limitations in the data collection (specifically, the way USCIS stores principal and derivative petitions and that USCIS was adjudicating petitions from mid-2014 only). For example, the approved waivers for inadmissibility for derivative petitions could not be reliably reported, as there was an insufficient number of approved derivative petitions that were still bundled with approved principal petitions. Likewise, USCIS could only estimate the requested class of admission (i.e., U-2, U-3) for derivatives for pending petitions (i.e., those filed after 2014) because there was an insufficient number of approved derivative petitions bundled with the principal petitioners in the sample.