# Alien numbers of migrant children are often sequential: Implications for data security and possible estimation

*Camille Fassett*[*]

*29 January 2020*

In a dataset recording thousands of undocumented migrant children who were detained in federally contracted centers in the United States between 2013 and 2018, the alien numbers of the children — unique identifiers assigned by the government — are generally sequential and positively correlated with the date the child entered the detention facility.

My hypothesis for this dataset after initial indicatory observations was that alien numbers were perfectly correlated with the entry date of a child into detention. Although the correlation is not perfect, it is strong, and this has implications for data security and estimation. This report will detail the dataset used for this analysis, the investigative steps taken to understand the relationship between alien numbers and dates of entry into detention facilities, and the potential leakage of sensitive data caused by sequential alien number assignments. Finally, I will evaluate explanations for outlier data, and explore possibilities for further work on this topic with additional data, including the potential to estimate the total population size of detained immigrant children at a given time.

## Introduction

The data used for this report was obtained by the research and data collective Lucy Parsons Labs in 2018. The dataset contains the names, alien numbers, countries of origin, birthdates, dates of entry in the detention facility, and other personal information for thousands of children held in shelters in the midwest of the United States between 2013 and late 2018. While Lucy Parsons Labs and the Human Rights Data Analysis Group are unable to release the dataset to the public due to the sensitivity of the personal identifying information it contains, it may be possible to release some cleaned version of it in the future.

The shelters at which the children in the dataset were detained are privately run facilities contracted by the federal government to provide services until a suitable guardian is identified. These shelters are overseen by the Department of Health and Human Services' Office of

---

[*]Human Rights Data Analysis Group

Refugee Resettlement (ORR). Some of the children were ultimately reunited with a relative living in the United States, while others were reunited with a parent in ICE custody. Others still were deported from the country, or turned over to ICE custody once the child turned 18. The records in the dataset detail the schedules, daily lives, family histories, and ultimate releases from detention.

# Outside information

United States Citizenship and Immigration Services (USCIS) is a government agency, nested within the Department of Homeland Security, that manages the United States immigration system. USCIS notes on its website that: "USCIS Number, or Alien Registration Number, Alien Number, or A-Number, is a unique, 9-digit number assigned to a noncitizen by the Department of Homeland Security that is listed on the front of Permanent Resident Card (Green Card)."

A version of the West Virginia Income Maintenance Manual chapter on Aliens, Refugees and Citizenship, posted online by the West Virginia Department of Health and Human Services, clarifies that alien numbers are assigned by local Field Control Offices that process the earliest action related to the individual. This seems consistent with our hypothesis that alien numbers are generally assigned sequentially; however, it is unclear whether they would be assigned sequentially within a given pre-numbered folder jacket.

A report by the Government Accountability Office (GAO) in December 2019 wrote that its internal analysis of ICE data showed that the number of arrests of juveniles varied, but increased overall, from calendar years 2015 through 2018.

"The Number of Arrests of Juveniles Increased Overall. The number of ERO (Enforcement and Removal Operations) arrests of juveniles increased overall by 53 percent from calendar years 2015 through 2018, as shown in figure 29.2 During the two years the Priority Enforcement Program was in effect, ERO arrests of juveniles increased 47 percent (from 887 arrests in 2015 to 1,307 arrests in 2016). Following issuance of the 2017 DHS memo, ERO arrests of juveniles increased 76 percent in calendar year 2017 (2,294 arrests), and decreased 41 percent in calendar year 2018 (1,361 arrests)."

Since the overall arrest count of undocumented juveniles has increased in recent years, it follows that the number of children that are held in private detention facilities has likely also increased in the same time frame.

# Data and Graphs

While looking closely at the dataset during the process of cleaning and standardization for potential analysis, we realized that for many sets of siblings, or for parents and children, the alien numbers were only one number apart from one another. When we saw the alien numbers clustered together, we wondered if they were assigned sequentially. This discovery

led to our hypothesis that alien numbers are highly correlated with entry dates into detention, which if true could make estimating the total population of detained children possible.

Alien numbers—generally comprised of 9 digits—are assigned not only to detained migrants, but to any non-citizen remaining in the United States for some length of time, such as greencard holders. Thus, if given an alien number with no additional information, it would be difficult to determine if the individual were a documented visa holder, or an undocumented immigrant detained by a border agency.

The total number of unique alien numbers extracted from the documents is 6,947. However, because the dataset also contains limited information about relatives of the children in custody, a small amount of these alien numbers are associated with parents rather than children. For the purposes of our analysis, only alien numbers with an associated entry date into detention were included. This reduced the frequency of non-child alien numbers in the dataset for analysis, but did not completely eliminate it. The number of unique alien numbers with associated entry dates is 5,076.

Our assumption upon discovering the close alien numbers among relatives was that alien numbers would rank-order correlate perfectly with entry dates. They do not, and in parts III and IV of this report, I will detail outlier data points and explore possibilities for breaks from the trend.

Alien numbers and entry dates into detention centers are positively, but not perfectly, correlated with a rank order correlation of 0.38. This is the number without removing any outliers, even known outliers — such as data points confirmed to be associated with a parent or an incorrect entry date. The alien numbers and their associated entry dates are shown in Figure 1.
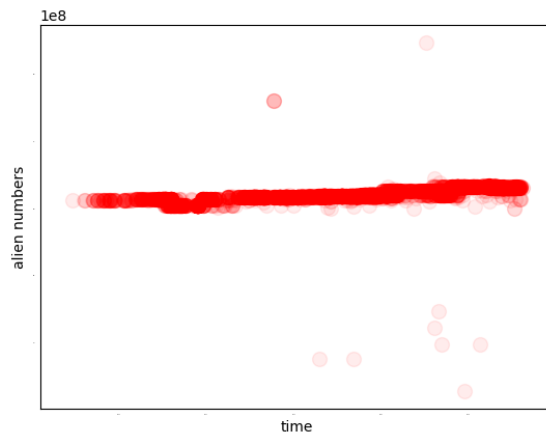


Figure 1: Distribution of alien numbers from 2013–2018

If my hypothesis that alien numbers increase sequentially were correct, we would expect to see a linear increase in alien numbers over time. Figure 1 shows that alien numbers are generally higher over time — as time goes by, alien numbers tend to increase.
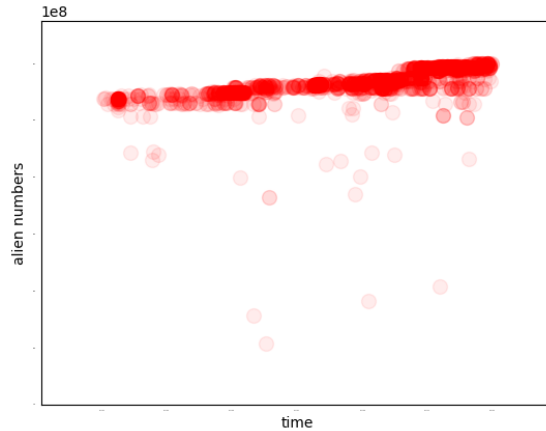
Figure 2: Distribution of alien numbers with a detention entry date in 2016

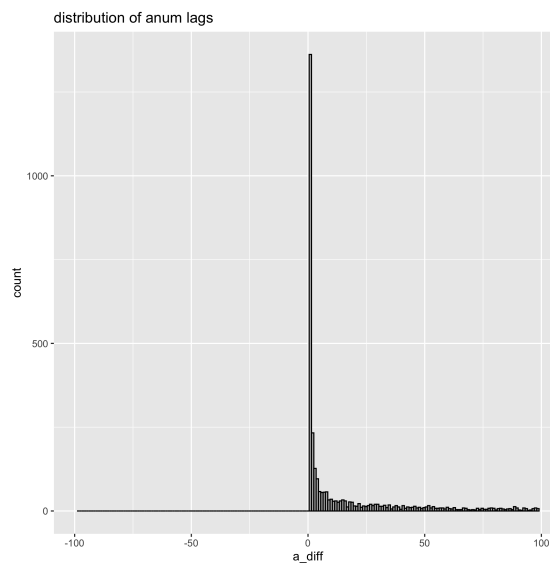Zeroing in on 2016 shown in Figure 2, we see observe several outliers — low alien numbers.



Figure 3: Histogram of alien number differences

To generate the histogram in Figure 3, the differences were taken from a list of sorted alien numbers. In other words, for each alien number in that list, the difference would be the value of the alien number in the previous index position minus this alien number. The distribution of differences is heavily centered on low positive numbers — meaning that generally, the alien numbers increment consecutively and gradually.

26 percent of these alien numbers are perfectly sequential, or have a difference (or lag) of one. Alien numbers cannot have a difference of zero, because alien numbers are unique.
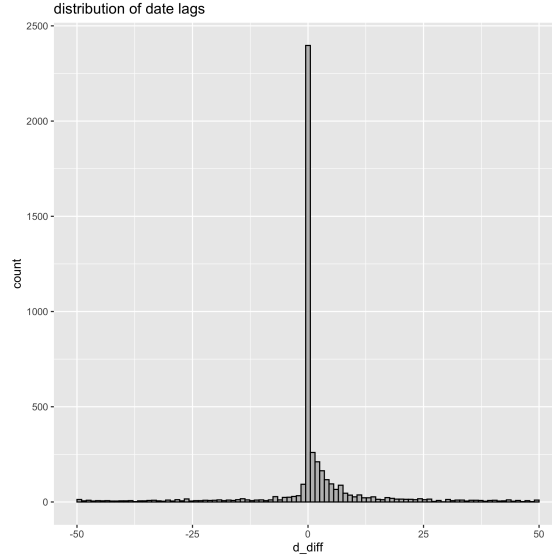
distribution of date lags

Figure 4: Histogram of entry date differences

The distribution of the frequency of differences in entry dates seen in Figure 4 shows a high concentration of differences in entry dates of zero. This is the case when two people with sequential alien numbers have the same entry date.

Approximately 83 percent of date differences are greater than or equal to one, meaning that in most cases, alien numbers increase as time goes by. The remaining 17 percent that have a date difference of less than zero, meaning the entry date occurs prior to that of the preceding alien number. These 17 are outliers.

# What we know, and what we don't

Based on our analysis of the dataset, we believe alien numbers are sequential in time, and are potentially assigned in blocks. However, there is some "noise" in our dataset, and the correlation between entry dates into detention and alien numbers is not perfect.

A possible reason for some of this noise is that some alien numbers are those of a relative rather than a child. However, in all likelihood the most significant contributor is that the entry dates in the dataset — which are dates of entry into specific detention facilities — are not the dates that the children were assigned their alien numbers. There are many reasons why this could be the case — children could have been assigned alien numbers at earlier dates, but then be transferred between detention centers as space became available, or held by government agencies for some length of time, or even have entered detention multiple times in their lives.

## Outliers

For the year 2016, I manually reviewed the information available in the dataset for the children with the lowest alien numbers, which, as is visible in the 2016 scatterplot above, are the most obvious outliers. There were previously several additional outliers in this group of low alien numbers, but those were identified to be numbers of older relatives rather than children, and consequently, their entry dates are likely different from the children in this dataset. All but two of the remaining alien numbers that are associated with children have birthdates between the years of 1998 and 2003.

The children with low alien numbers are among the older in the dataset. As mentioned previously, a possible explanation for the unexpectedly low alien numbers in these cases is that these children were assigned their a-numbers at an earlier date — i.e. a child could have been detained and assigned an alien number several years prior, but could have been deported and later entered into detention again in 2016 after coming back to the United States. There is not sufficient information in this dataset to investigate this possibility further.

An alternate explanation for the inconsistencies among alien numbers and entry dates is that if the alien numbers are indeed handed out in blocks by location or field office, the numbers would only be sequential for a given locale. For example, if children detained at facility A were issued alien numbers 1-100, and facility B were issued 101-200, all of the children in facility B would have higher alien numbers regardless of their detention or processing dates. It could be possible to detect this with further work.

## Possibilities for future work

With more information — such as that sufficient to determine the size of the pre-numbered alien number folder jackets, the locations of the Field Control Offices that assigned them, and the dates children were assigned alien numbers (rather than the dates they entered a given private facility)— further data science, analysis, and predictive modeling work could be possible.

Please do not hesitate to get in touch (email: camillefassett [at] gmail [dot] com) if anyone would like to share data, propose ideas for collaboration, or has tips about corroborating datasets.

Estimation possibilities could include:

- The ratio of detained children in ORR custody, compared to all other recipients of alien numbers (detained adults, non-citizen residents of the United States, etc)
- The number of total children in detention for a given time
- How the number of detained children changed with regard to certain policy changes and known events (i.e., Trump's zero tolerance policy)

During World War II, the Allies successfully used the practice of estimating the total size of a set from a much smaller sample within it, in what would become one of the most

famous examples. The Allies preserved German military equipment in the hopes of using it for intelligence gathering purposes, and had access to a small number of serial numbers from German tanks. Allied statisticians correctly guessed that the tanks were numbered sequentially, and were able to estimate from the sample serial numbers the total number of German tanks each month.

Similar estimation methods could be used on this dataset. If alien numbers are assigned in blocks, it could be possible to estimate the number of children within a given detention facility. It could also be possible to determine the detention facility at which a child with a known alien number is being held.

# Conclusion

Alien numbers being sequential could make possible estimations of the population of various categories of non-U.S. citizens, such as detained children. This estimation could serve as a means to corroborate or interrogate reported numbers of people in ORR custody. In light of current and widespread reports of abuse of migrants in border agency custody in the United States, third-party validation of government-reported data would be valuable, and there would be enormous public interest value in knowing the total number of children in detention.

At the same time, patterns in alien number assignments pose risks to the security of the data of non-U.S. Citizens. Knowledge about alien number assignments could make it possible to make conclusions about detained migrants over time, but it also could have significant implications for the security of immigrants' data and identities. If one is aware of this and has sufficient other pieces of information, one could infer sensitive information about an immigrant, or personal details about their history.

When talking about sensitive data, I draw a distinction between potential for identification (a person's name, for example) and authentication (a password). However, problematic or not, some information — including social security and alien numbers, but also sometimes birthdates — are used for both purposes. Social security or alien numbers are unique identifiers used by the government, but are also used to authenticate a person's identity for the purposes of banking or health care access.

Even the knowledge that alien numbers are not randomly generated means that it is possible to guess real alien numbers, which in combination with other kinds of information could be used to impersonate that person and gain access to their accounts.

Conversely, with enough details about a migrant — such as the date they arrived to the United States and the Field Control Office that first processed them — it could be possible to infer their alien number, or narrow it down to a small range of numbers.

# Acknowledgements