

U.S. Financial Health Pulse - 2020 Trends Report Transactional Data Methodology Supplement

Updated: October 2020

This document provides an in-depth explanation of the methodology used to collect and analyze transactional data in the U.S. Financial Health Pulse 2020 Trends Report. If you have any questions or would like to discuss further, please contact us at: pulse@finhealthnetwork.org.

I. Research Overview

The [U.S. Financial Health Pulse](#) is a rigorous, widely publicized, and regularly refreshed snapshot of financial health. Leveraging the [FinHealth Score®](#) measurement methodology, longitudinal consumer surveys, and transactional records, the Pulse provides a regularly refreshed snapshot of the country's financial health and how it is changing over time.

Survey data are collected through the University of Southern California's "[Understanding America Study](#)," (UAS) a nationally representative probability-based internet panel. The surveys employ a cross-sectional and longitudinal design, allowing researchers to explore how financial health changes for the country and for individual respondents over time.

Transactional data are collected from study participants who consented to link their financial accounts to a secure online platform that leverages [Plaid's](#) API. These data allow researchers to track key elements of consumers' financial lives in near-real-time and disaggregate trends by demographics, socioeconomic characteristics, and survey data responses.

II. Data Collection

Pilot Phase

In April 2019, the Financial Health Network and USC's Center for Economic and Social Research sent an initial consent survey to a random sample of 888 UAS panelists (via surveys [UAS 181](#) and [UAS 192](#)) inviting them to share their transactional and account data with researchers. The survey included an initial screening question to determine if respondents banked online or using a mobile device.¹ Since the Pulse data-sharing platform only allows users to connect accounts that are online or mobile-accessible, individuals who do not bank online or via a mobile device were not eligible to participate in the study. If respondents said that they did bank online or using a mobile device, they were invited to link their financial accounts (such as checking accounts, savings accounts, credit cards, mortgages, 401(k)s, student loans, etc.) to a secure online platform that leverages Plaid's API.

Participants logged onto the platform using a unique URL and followed interactive instructions to provide their credentials for online financial accounts they owned. Respondents were provided information about the secure storage of their financial information and instructions about how they

¹ The question was asked: "Q178. In the past 12 months, have you used internet banking? [Using the internet to access your bank or credit union account. This can be done by accessing your bank's web page through the browser on your phone, tablet, or computer.]" and "Q179. In the past 12 months, have you used mobile banking? [Using a mobile phone to access your bank or credit union account. This can be done either by accessing your bank's web page through the web browser on your mobile phone, via text messaging, or by using an application downloaded to your mobile phone.]"

could unlink their accounts. One hundred panelists ultimately linked at least one financial account during the pilot.

Participants received an initial incentive for every institution they linked to the platform and an additional smaller incentive for each month the institution remained actively linked to the platform. We tested two different incentive schemes: one where participants received \$5 for every financial institution they linked and one where participants received \$10 for every financial institution they linked. Both groups of respondents received \$1 for each month that a financial institution remained actively linked. We found that participants who received \$10/account linked an average of one more account than those who received \$5/account (participants receiving \$10 per account most commonly linked another credit card). Based on these results, we decided to offer all participants \$10/account when we expanded the research beyond the pilot.

Scale-Up Phase

In the first half of 2020, we revised the data sharing platform based on lessons learned from the pilot. We reviewed participant feedback from the pilot and responses from eligible participants who elected not to participate. The most commonly cited reasons for not participating in the pilot were not wanting to share financial data and concerns about security. Based on these comments, as well as feedback from external reviewers, we simplified the participant instructions and emphasized the security of participants' data in the study invitation and throughout the data sharing platform. We also streamlined the account linking flow to make the linking process more step-by-step and provided examples of common accounts that participants could link, including checking accounts, savings accounts, credit cards, prepaid cards, brokerage accounts, and employer-sponsored retirement accounts.

In June 2020, we invited the rest of the UAS panel (8,056 panelists) to participate in the study using the revised consent survey and updated data sharing platform. By the end of July 2020, 738 new individuals had linked at least one financial account to the platform (see the "sample composition" section for more information on demographic composition of sample). In all, 835 participants (out of the total of 8,797 participants invited) linked at least one financial account, yielding a participation rate of 9.5%. As of July 31, participants had linked 5,219 accounts across 2,312 financial institutions; the mean number of accounts linked by participants was 6.3 and the median was 4.²

² An additional 26 individuals linked accounts that never provided information via the Plaid linkage. We exclude those people from our analysis and reporting.

Image 1. Landing page for participation in transactional data collection

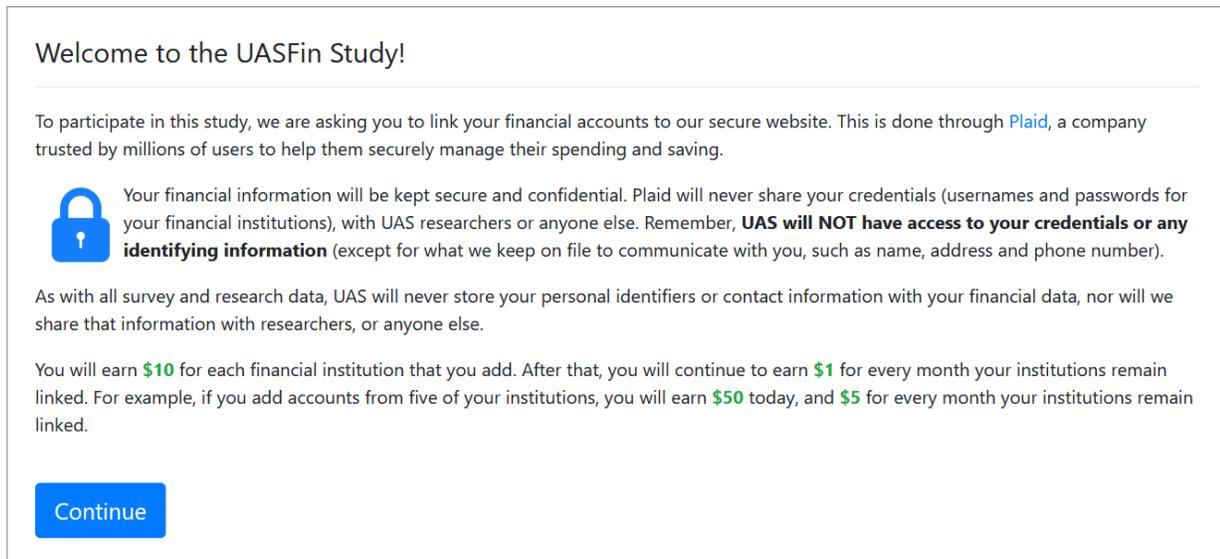


Image 2. Initial account linkage page for a checking account

Please add the financial institution where you have a **checking** account.
(This will also connect other accounts you have at this institution.)

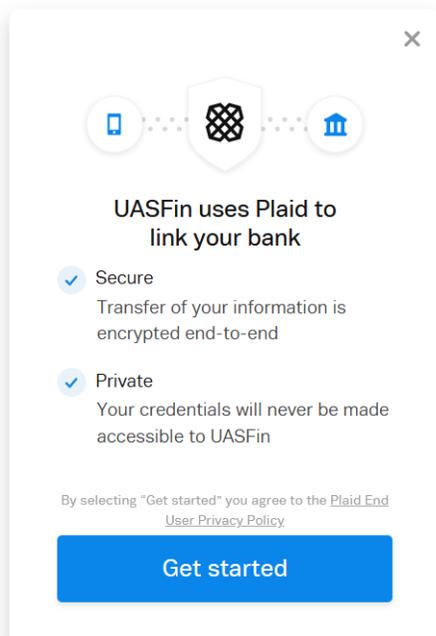


Image 3. Prompt for linking less common accounts

Do you have any other **financial accounts** at other institutions that you would like to add?

Here are common types of other financial accounts that you might have:

Savings and Investments

- Certificate of Deposit (CD)
- Money Market Account
- Health Saving Account (HSA)
- Flexible Spending Account (FSA)
- College Savings Accounts (529 Plan, etc.)

Loans and Lines of Credit

- Student Loan
- Auto Loan
- Line of credit
- Other Personal Loans

Data Security

Given the sensitive nature of transactional data, the USC research team cleaned the dataset and removed data fields that could potentially contain personally identifiable information (including vendor id, transaction description, and institution name) prior to sharing the data set with the Financial Health Network. Thus, the dataset used for analysis included Plaid's transaction categorizations, information on the account used, the transaction dollar amount, the running balance in accounts, and the date of transaction. The Financial Health Network does not have access to participants' credentials (e.g. username or password). We also took additional precautions to ensure that the data were stored securely and with minimal risk of re-identification through use of Amazon Web Services S3 buckets and structured credentials for the analytical team at the Financial Health Network. All members of the analytical team completed IRB Human Subjects research training.

III. Dataset Construction

Inclusion Criteria

In order to construct the transactional data set used for analysis in the 2020 Pulse Trends Report, we developed and applied filtering criteria to the collected data. We also applied additional definitions and inclusion thresholds for certain metrics, such as minimum account activity thresholds (see the metric definitions table for complete information).

To be included in the Pulse transactional data set, a financial account must have been:

1. Actively linked at beginning of the study period (January 1, 2020)
2. Actively linked at end of the study period (July 31, 2020)
3. Active linked for 80% throughout the study period

Ensuring that accounts were actively linked via Plaid meant that we received a regular flow of data for that account and could therefore be confident in the completeness of the data. When account linkages go inactive because of changed account credentials or participants unlinking their accounts, for example, the data flow stops and financial activity that occurs during the period of inactivity may be missed.

We chose these inclusion criteria to ensure that we had sufficient data on all individuals in the sample during the study period and to ensure that there were no systematic time-based skews within the data (e.g., many individuals missing data at the beginning or end of the study period). Applying such inclusion criteria reduced the overall size of the sample, but was necessary to ensure the data quality of the final sample.

The sample sizes for all accounts dropped significantly after applying our inclusion criteria, reflecting the amount of missing account data in the dataset. The largest drop occurred in loan and brokerage accounts: while Plaid’s [“Assets” product](#) could retrieve historic transaction and balance data from before a person linked a particular checking account, savings account, or credit card, it could not do the same for loan or brokerage accounts. Therefore, very few of these accounts met the inclusion criterion of having an active link at the beginning of the study period.

Sample Sizes by Linked Account Type Before and After Applying Inclusion Criteria

Account Type	No inclusion criteria # of individuals with linked accounts	Applied inclusion criteria # of individuals with linked accounts
Any account	835	629
Checking Account	769	499
Savings Account	547	321
Credit Card	495	360
Any Loan	265	14
Brokerage / Investment Account (401ks, IRAs, and other retirement or investment accounts)	192	4
Paypal	108	77
CD	30	16
Money Market Account	26	15
Cash Management	11	2
Prepaid	6	1
Other	23	3

Comparisons to account ownership as indicated in survey data

We compared account ownership reported by transactional data collection participants on a Pulse survey (fielded between April 20 and May 7, 2020) to the actual accounts linked in order to assess the consistency of the account data in the two sources (see table below). We found fairly high consistency for checking accounts, savings accounts, and credit cards, indicating that the financial accounts linked by participants were generally the same as those that they reported owning in the Pulse survey. The high degrees of consistency for these accounts, and lower amounts of consistency for other types of financial accounts, helped us make the decision to focus our analysis solely on credit cards and liquid accounts in the 2020 Pulse Trends report.

Account Type	% of study participants who reported owning account on survey who linked the account to the data sharing platform
Checking	93%
Credit card	71%
Saving	80%
Employer-provided retirement account (such as a 401k, 403(b) or Thrift Savings Plan (TSP), etc.)	16%
Prepaid card	1%
Individual retirement account not provided by an employer (such as an IRA, Keogh, SEP, or any other retirement fund)	29%
Cash management, money market, health savings account (HAS), and/or certificate of deposit (CD)	24%
Brokerage account, annuity, profit sharing/stock plan, and/or 529 plan	39%
Employer-provided traditional pension or cash balance plan	1%
Other financial assets or accounts	19%

There are several potential explanations for the observed differences between reported account ownership on the Pulse survey and the rate at which these accounts were actually linked:

- Participants did not have online accounts for transactional accounts they reported owning on the survey
- Participants did not have ready access to online credentials for transactional accounts they reported owning on the survey
- Less-frequently used transactional accounts did not come to mind for respondents during the Pulse transactional data linkage process
- There were technical issues in the account-linking platform

Sample Composition

The overall demographic composition of the final Pulse transactional data set broadly aligns with the demographic composition of individuals who said they banked online or via a mobile device in the U.S. Financial Health Pulse survey fielded in May, 2020. We use this population as our benchmark because it most closely aligns with the group of individuals who were eligible to participate in the transactional data portion of the study, since the Pulse data-sharing platform only captures information from online or mobile-accessible financial accounts. As a result, we would not expect to see individuals who do not have a bank account or who solely use cash represented in the transactional data sample.

However, there are some differences between the two data sets. Compared with the sampling frame, the Pulse sample skews slightly higher income, younger, and less financially healthy. There are also skews in the Pulse sample along gender (with women overrepresented), race and ethnicity (with Black respondents underrepresented), and education (with those with less education underrepresented). A multiple regression analysis confirmed that these demographic differences remained even after controlling for other demographic characteristics.³

Sub-Sample Composition

Because we break the overall sample into smaller subsamples based on account ownership for analysis in the 2020 Trends Report, we also examined the demographics of individuals who had linked a specific financial account. The demographic composition of these subsamples differed slightly from that of Pulse survey respondents who banked online or via a mobile device, but differences remained fairly small and intuitive (for example, individuals who linked a credit card had higher income and were more financially healthy). We also checked the demographic compositions of these samples after applying the inclusion criteria to ensure that doing so was not systematically biasing the remaining sample. Once again, the differences that arose as a result of the criteria were small compared to the raw data and the demographic composition of survey respondents who banked online or with a mobile device.

The table below summarizes the demographic compositions of individuals in different transactional samples before and after applying our inclusion criteria and relative to that of survey respondents who indicated they banked online or with mobile devices. The “raw” columns indicate the samples of *any* individuals who own a given account while the “criteria applied” columns indicate the samples of any individuals who link a given account after we apply our inclusion criteria for that account.

A small number of participants in our transactional data collection did not respond to the Q2 Pulse survey, so we do not have demographic data on a small proportion of the sample (indicated by the “NA” categories in the tables below).

³ Given our interest in sharing nationally representative insights, we considered weighting the dataset of transactions based on the demographic characteristics of the respondents. Ultimately this approach was abandoned for lack of a comparable nationally representative benchmark (though the FDIC Un/underbanked CPS supplement was considered) and because of the relatively small participant sample size of the transactional dataset. Additionally, because our inclusion criteria for each metric created different subsamples of participants that changed over time, weighting these subsamples for analysis could have been impractical (see “metric construction” for more information).

Household Income	2020 Q2 Pulse Survey	Link any transactional account		Link a credit card		Link any liquid account	
	Weighted, Online Banked	Raw Data	Criteria applied	Raw Data	Criteria applied	Raw Data	Criteria applied
Less than \$30,000	19%	18%	15%	12%	9%	18%	16%
\$30,000 - \$59,999	28%	24%	23%	20%	18%	24%	25%
\$60,000 - \$99,999	27%	26%	27%	30%	32%	26%	24%
\$100,000 or more	27%	28%	32%	36%	40%	29%	33%
NA	NA	4%	2%	2%	1%	4%	3%

Gender	2020 Q2 Pulse Survey	Link any transactional account		Link a credit card		Link any liquid account	
	Weighted, Online Banked	Raw Data	Criteria applied	Raw Data	Criteria applied	Raw Data	Criteria applied
Women	52%	57%	57%	53%	52%	57%	57%
Men	48%	40%	41%	45%	47%	39%	40%
NA	NA	4%	2%	2%	1%	4%	2%

Race and Ethnicity	2020 Q2 Pulse Survey	Link any transactional account		Link a credit card		Link any liquid account	
	Weighted, Online Banked	Raw Data	Criteria applied	Raw Data	Criteria applied	Raw Data	Criteria applied

Asian American, American Indian or Alaska Native, Hawaiian/Pacific Islander	6%	7%	8%	9%	10%	7%	9%
Black	10%	5%	5%	3%	2%	5%	6%
Latinx	17%	17%	17%	16%	15%	17%	20%
White	64%	63%	63%	64%	68%	62%	59%
Multiple Races	3%	4%	4%	5%	5%	4%	4%
NA	NA	4%	3%	3%	1%	4%	3%

Education	2020 Q2 Pulse Survey	Link any transactional account		Link a credit card		Link any liquid account	
	Weighted, Online Banked	Raw Data	Criteria applied	Raw Data	Criteria applied	Raw Data	Criteria applied
Less than high school	5%	3%	3%	2%	1%	3%	3%
High school	25%	11%	10%	8%	7%	11%	9%
Some college	30%	35%	34%	31%	30%	34%	34%
Bachelor's degree or higher	40%	47%	51%	58%	61%	48%	51%
NA	NA	4%	2%	2%	1%	4%	3%

Age Group	2020 Q2 Pulse Survey	Link any transactional account		Link a credit card		Link any liquid account	
	Weighted, Online Banked	Raw Data	Criteria applied	Raw Data	Criteria applied	Raw Data	Criteria applied
18-25	7%	10%	10%	10%	9%	10%	11%
26-35	21%	23%	24%	26%	25%	23%	25%
36-49	29%	30%	30%	29%	31%	30%	28%
50-64	25%	22%	23%	21%	23%	22%	22%
65+	18%	12%	11%	13%	12%	12%	11%
NA	NA	4%	2%	2%	1%	4%	3%

Currently Working	2020 Q2 Pulse Survey	Link any transactional account		Link a credit card		Link any liquid account	
	Weighted, Online Banked	Raw Data	Criteria applied	Raw Data	Criteria applied	Raw Data	Criteria applied
No	39%	34%	31%	28%	26%	34%	32%
Yes	61%	62%	67%	69%	73%	62%	66%
NA	NA	4%	3%	2%	1%	4%	3%

Financial Health Tier	2020 Q2 Pulse Survey	Link any transactional account		Link a credit card		Link any liquid account	
	Weighted, Online Banked	Raw Data	Criteria applied	Raw Data	Criteria applied	Raw Data	Criteria applied

Healthy	34%	28%	30%	36%	39%	28%	29%
Coping	53%	52%	54%	55%	55%	53%	53%
Vulnerable	13%	15%	13%	6%	5%	14%	14%
NA	NA	5%	3%	3%	2%	5%	4%

IV. Analytical Decisions

In the 2020 Trends Report, we developed metrics to provide additional context to the eight survey data [indicators of financial health](#), focusing on metrics with sufficient account coverage in the transactional sample (i.e. that could be constructed from transactions from checking accounts, savings accounts, and/or credit cards).

Look-Back Period

We used a 30-day look-back period when calculating all metrics. Given the day-to-day noisiness of transactional data and the relative infrequency of key financial events (e.g., the monthly cadence of many bills and the relative uncommonness of late fees), we calculated a rolling look-back period over the past 30 days to smooth the metric trends. We considered other look-back periods such as 7- and 60-day intervals, but eventually chose a 30-day period because it strikes the right balance between temporal accuracy and reduction of noise.

Median Values

When looking at dollar value figures in the transactional data set, we typically report the median amount of a metric to avoid outliers affecting the trends. For any given metric, we calculate the median value for each day over the sample for whom the metric is calculated. The table below summarizes the different account groupings we use for analysis, sample sizes, and metric definitions.

Activity Thresholds

We apply basic transaction activity requirements for liquid account inflows and outflows to ensure we are able to actually observe some day-to-day transaction activity. We do not apply activity requirements to other metrics, because transaction activity is less necessary in those instances to calculate the metric (e.g., an individual can hold a liquid account balance without having any transactions in the past 6 months).

Transfers

Transactions categorized as “transfers” are very common in our transactional data; they constitute 19% of all outflow transactions and 30% of total transactions. Including transfers in the analysis of inflows and outflows runs the risk of misidentifying transfers for the purpose of moving funds between personal financial accounts as income or an expenditure. We elected to include transfers in our analysis, though, because we do not need to directly identify income and

expenditures for our metrics and because any potential double counting would be accounted for in our analysis. Furthermore, excluding transfers could run the risk of biasing subsequent analysis against individuals who own investment accounts, brokerage accounts, or any other type of account where transfers are common.

Definitions of Metrics in 2020 Trends Report

Metric	Definition of Metric	Accounts Used for Analysis	# of people	# of accounts	# of institutions
Liquid account inflows	Rolling total of all liquid account inflows, summed over a 30-day look-back period for each individual. We calculate the median value of this metric for the sample on a daily basis.	Account Types: Liquid Accounts = checking, savings, money market, prepaid, and cash management accounts. Inclusion Criteria: All liquid accounts must meet all inclusion criteria (see Box A1). All included individuals must also have at least one linked checking account.	491	1216	623
Liquid account outflows	Rolling total of all liquid account outflows, summed over a 30-day look-back period for each individual. We calculate the median value of this metric for the sample on a daily basis.	Activity Thresholds: At least one transaction per month in a checking account.			
Liquid account balances	Rolling average of total liquid account end-of-day balances, over a 30-day look-back period for each individual. We calculate the median value of this metric for the sample on a daily basis.	Account Types: Liquid Accounts = checking, savings, money market, prepaid, and cash management accounts. Inclusion Criteria: All liquid accounts must meet all inclusion criteria (see Box A1). All included individuals must also	499	1234	632

		<p>have at least one linked checking account.</p> <p>Activity Thresholds: None</p>			
Proportion of sample with late fees	Proportion of individuals with linked credit cards who had at least one expense categorized as "Late Payment" in the past 30 days. We calculate this proportion on a daily basis.	<p>Account Types: Credit Cards</p> <p>Inclusion Criteria: All credit cards must meet all inclusion criteria (see Box A1).</p> <p>Activity Thresholds: None</p>	360	616	774
Credit card charges	Rolling total of all credit card outflows over a 30-day look-back period for each individual. We calculate the median value of this metric for the sample on a daily basis.				
Credit card balances	Rolling average of total credit card end-of-day balances, over a 30-day look-back period for each individual. We calculate the median value of this metric for the sample on a daily basis.				

V. Study Limitations

Sample Size

The total sample size of just over 800 participants in the transactional data set limits our ability to examine smaller demographic subsamples, particularly after we apply our inclusion criteria. We're

considering ways to increase our sample size, but this is a limitation that will be difficult to change in the short-term.

Cash Usage

Since the data-sharing platform only allows users to connect accounts that are online or mobile-accessible, individuals who do not bank online or via a mobile device were not eligible to participate in the study. We are not able to collect data on transactions conducted entirely in cash; the most insight we can gain into cash usage using this dataset is examining ATM withdrawals or cash deposits into a linked account. Therefore, we lack complete coverage of the financial transactions of the unbanked, underbanked, and cash-preferred.

To understand the scope of this missing data, we are in the process of linking our transactional dataset to data from the Diary of Consumer Payment Choice, which is fielded to the same survey panel and directly asks respondents about their cash transactions. After merging these two data sets we can examine the extent of volume and dollar value of individuals' cash transactions. Findings from prior nationally representative [Diary of Consumer Payment Choice research](#) indicate that cash transactions make up 6% of transactions by value and 26% of transactions by volume. Given our metrics focus on transaction values and not transaction volume, this point of comparison suggests that we are missing a relatively small amount of cash activity in our Pulse transactional dataset. We may also use the DCPC to examine the differences between reported account ownership and transactional accounts that are actually linked.

Attrition

The total number of linked accounts in our transactional data set has gradually decreased after participants initially linked their accounts. Out of the 835 participating individuals, 65 people no longer have a linked account (8% of the total sample as of September 2020). Furthermore, 20% of all accounts linked had become inactive after being initially linked. This attrition generally occurs when user credentials expire or change, or users voluntarily delink their financial institutions, resulting in inactive account links. Checking accounts, savings accounts, and credit card accounts are most likely to attrite. We hypothesize that this is because frequent use of these accounts prompts financial institutions to force users to reset their password more frequently than for other account types. When the user resets their password with their financial institution, they likely do not also update it in the UAS Plaid interface.

To alleviate user credentials expiring and unlinking, our partners at USC are sending regular reminders to users whose accounts become inactive to encourage them to relink their accounts. Relinkage after an account goes inactive allows us to fill in the gaps when an account was not linked; Plaid fills in the missing data gaps when the account is relinked, provided the period of inactivity is not too long. However, in practice, very few participants have actively relinked accounts. We continue to actively examine how to combat attrition.

Missing Data

We are not able to access all transactional data from a given financial account after a participant links that account. Although the Plaid API allows us to see past transactions that go back a limited time period after an account is linked, accounts are deemed "inactive" in those past dates and balances cannot be observed. This issue also grew as our sample of individuals expanded in June 2020, from 100 to 800+ individuals.

As a result, the sample sizes for metrics are quite sensitive to the analysis period. To address this issue, we used [Plaid's "Assets" product](#) to construct historical account balances. This product allowed us to capture historical account balances at the time an individual linked an account (while we are able to calculate balances ourselves on an ongoing basis after the linkage). However, while some accounts had a year's worth of data going back in time, others only had a few months' worth of data. The "Assets" product allows us to collect more data from participants, but attrition and the limits of historical data collection mean we still face challenges in working with missing data.

Seasonality

The amount of historical data we are able to gather when an individual links an account via Plaid limits our current ability to understand the impact of seasonality on the transactional trends we observe in the Pulse data set. The amount of time that we are able to look back into someone's transactions prior to their linking of an account varies by account type and institution, though the modal length of time is about 6 months. Given the majority of our participants first linked their account(s) in June 2020, we can only retrieve complete data going back to January 2020 for the majority of the sample. As a result, we cannot currently compare trends to previous years of data, and thus may lack some context into which trends are driven by seasonal factors that might be driving these trends. Furthermore, the unique economic context in which we have started our data collection may make this period difficult to use as a baseline for comparison for future data. Over time, though, we will gather sufficient data to allow us to compare to previous years.

VI. Looking Ahead

We will continue to analyze the collected transactional data over time. We aim to improve the data quality and size of the sample over time, making the transactional data and merged survey and transactional dataset even more valuable.

In the coming years we will explore additional topics, including tracking financial health metrics over an extended time period and developing a financial health scoring methodology based on transactional data.

We will continue to release findings from our transactional data analysis through [U.S. Financial Health Pulse](#) reports. Please direct any questions about the transactional data research to pulse@finhealthnetwork.org.