APPAM Pre-Conference Workshop:

Understanding and Using the Entrepreneurship Survey Data

Presented: November 16, 2022

Quentin Brummet | EPOP Principal Investigator
Michael Yang | EPOP Sampling and Weighting Statistician
Katie Johnson | EPOP Research Methodologist
Karen Grigorian | EPOP Project Director

The Entrepreneurship in the Population (EPOP) Survey Project

The EPOP Survey…

- Is a new publicly available resource

- Supports understanding about the scope of entrepreneurial activities across the U.S.

- Provides various measures of entrepreneurial behavior at national, state, and local levels

- Includes 5 years of survey collection from 2022 to 2026 to support trend analysis

- The questionnaire and data from EPOP:2022 are available to analysts today

Acknowledgment

The EPOP Survey Project is both guided and funded by the Ewing Marion Kauffman Foundation

We are grateful for this support from the Foundation

This Workshop covers…

Discussion of the specifics of the EPOP Survey

- Measurement of entrepreneurial activities

- Sample and study design

Using the data

- Overview of items in the questionnaire

- Examples of data analysis

Data access, support, and sharing

→ **We want to hear from you!**

Workshop Housekeeping

Data querying and code examples will be conducted in Stata

- Additional program language code is available upon request

Ask questions at any point!

Agenda

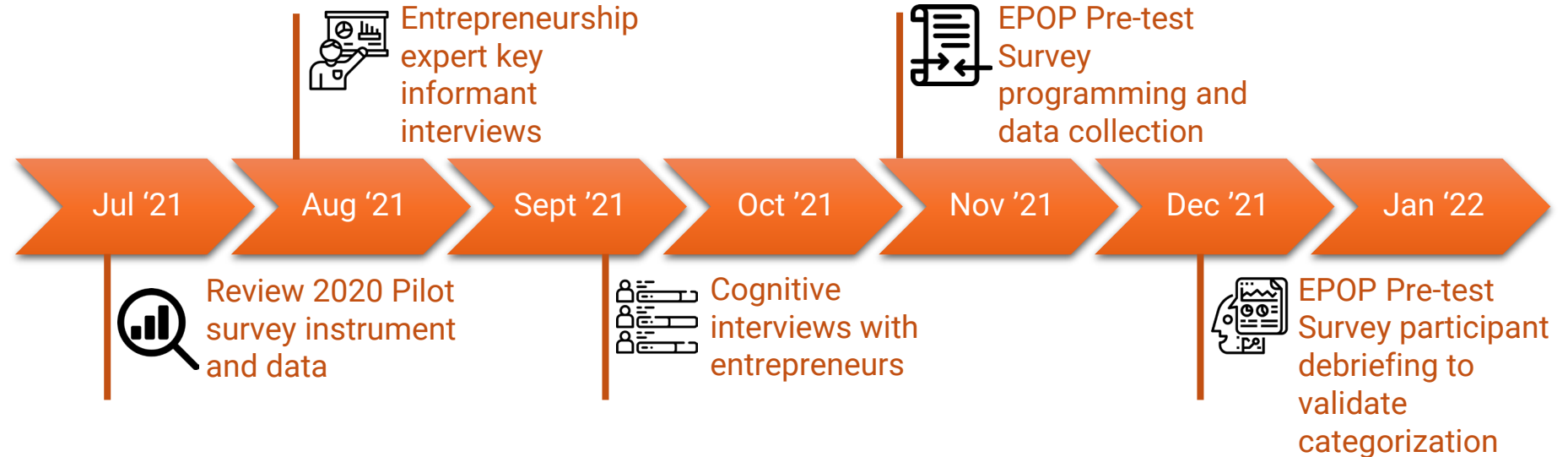|		
01	Key Concepts and Measurement	11:00 – 11:30 am
02	Sample and Study Design	11:30 am – noon
	~~LUNCH BREAK~~	noon – 12:15 pm
03	Using the Data	12:15 – 1:45 pm
	~~BREAK~~	1:45 – 2:00 pm
04	Data Access, Support, and Sharing	2:00 – 2:30 pm
05	General Q & A	2:30 – 3:00 pm

Key Concepts and Measurement

Taking a comprehensive perspective on the entrepreneurial pathway

Overview of the Survey Development Process

EPOP SURVEY INSTRUMENT VALIDATION

Entrepreneurship expert key informant interviews

EPOP Pre-test Survey programming and data collection

| Jul '21 | Aug '21 | Sept '21 | Oct '21 | Nov '21 | Dec '21 | Jan '22 |

Review 2020 Pilot survey instrument and data

Cognitive interviews with entrepreneurs

EPOP Pre-test Survey participant debriefing to validate categorization

Measuring Entrepreneurial Activities in the U.S.

The EPOP Survey measures 7 different entrepreneurial activities:

1. Current Business Ownership

2. Current Freelancing

3. Entrepreneurship Planning

4. Former Business Ownership

5. Former Freelancing

6. Withdrawn Entrepreneurship Planning

7. Gig Working

Employment Status and Entrepreneurial Measurement

Because entrepreneurial activities are often conducted while in addition to primary employment, the EPOP Survey asks detailed questions about employment status.

- By asking this follow-up question of respondents who report they are not working, the survey ensures more informal work activities are captured and asked about which is important for determining an accurate measurement of gig work and the full suite of entrepreneurial activities.

- The survey also collects information on characteristics of the job, such as whether it is with a for-profit or non-profit employer, industry, etc.

"In the last week, did you work for pay at a job or business?"

"In the last week, did you do ANY work for pay, even for as little as one hour?"

Measuring Current and Former Business Ownership and Freelancing Activities

Once job information has been captured, the EPOP Survey asks questions to gauge entrepreneurial activities directly.

- Respondents are asked if they have ever owned a business or worked as a freelancer -- some respondents report that they currently own a business even though it is not reported as a current job.

- If an individual reports they are a current business owner, they are not asked about former business ownership; similarly, current freelancers are not asked about former freelancing.

    – While there are reasons for this logic, it will be revised in future surveys.

Consider respondents whose first or second job is <u>NOT</u> a business owner, but who report EVER owning a business

The EPOP Survey directly asks these respondents: "Do you still own this business?"

- Those who say yes are classified as a "current business owner"

- Those who say no are classified as a "former business owner"

  - A **similar process** is followed for asking about freelance work

Three components of "current business ownership" measurement
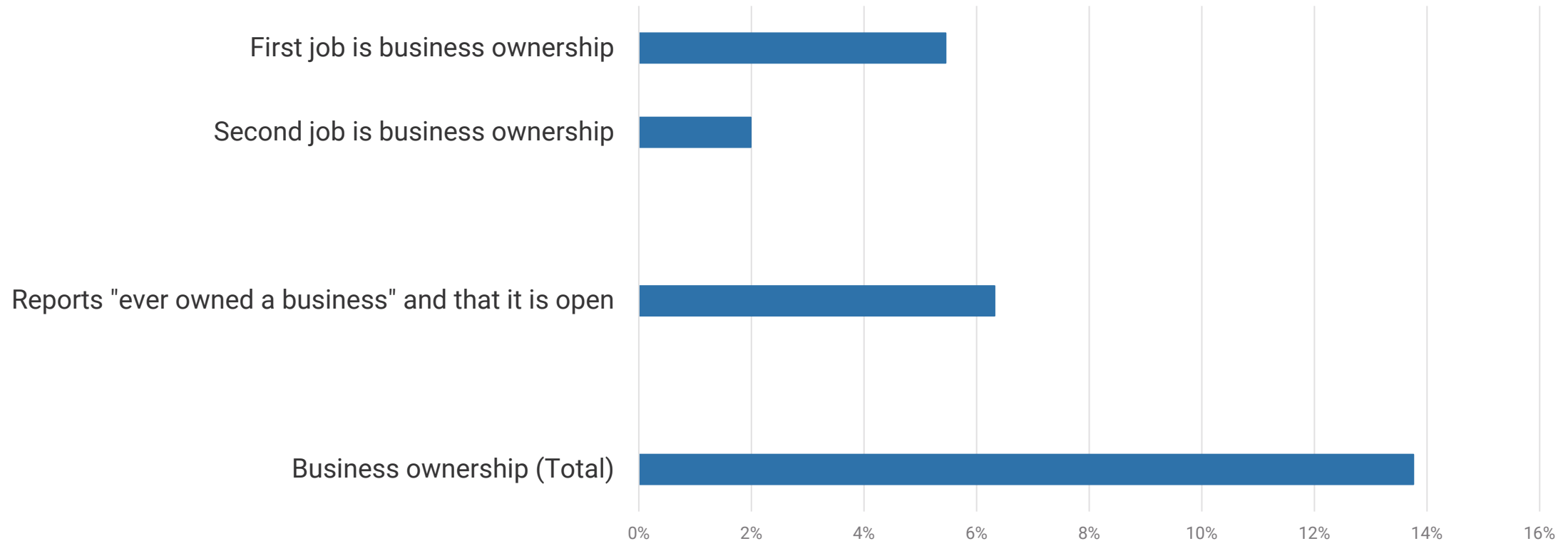
Source: NORC, Entrepreneurship in the Population Survey: 2022.

Entrepreneurship Planning

"Are you, alone or with others, currently trying to start a new business, including any form of self-employment, freelancing, consulting, or independent contracting, or selling any goods or services to others?"

- Note this does not require that individuals have taken a certain number of steps towards opening a business or any specific step, but rather it allows nascent entrepreneurs to classify themselves.

- Future questions ask respondents for the steps they have taken towards entrepreneurship.

Withdrawn Entrepreneurship Planning

"Have you, alone or with others, ever considered starting a new business, including any form of self-employment, freelancing, consulting, or independent contracting, or selling any goods or services to others but decided to wait or change your mind?"

• Only asked if individuals report that they have never engaged in any other entrepreneurial activities.

Gig Working

- Along with the question, respondents are provided with examples of common gig work platforms (e.g., Uber, Etsy, Fiverr).

- This is asked for each job.

- If individuals report two jobs that are not gig work, they are asked if they are engaged in any other gig work in addition to their two primary jobs.

- Because of the potential for confusion on what counts as gig work, the survey contains future items on whether the gig work activity uses an online app and asks for the name of the gig work platform.

"Some people earn money through short, paid tasks or jobs online or in-person that are conducted through companies that coordinate payment for the service. This is sometimes referred to as 'gig work.'"

Additional Measures of Gig Work

"Is your gig work conducted through a company that coordinates payments or relationships with clients?"

"Is the company that coordinates payments or relationships with clients for your gig work an online app?"

"What is the name(s) of the company that coordinates payments or relationships with clients for your gig work?"

Categories of Gig Platforms:

1. Confirmed Gig Platform: Services

2. Confirmed Gig Platform: Selling/Renting of Goods

3. Confirmed Gig Platform: Online Surveys

4. Payment Provider

5. Unconfirmed Gig Work

6. Unlikely Gig Work

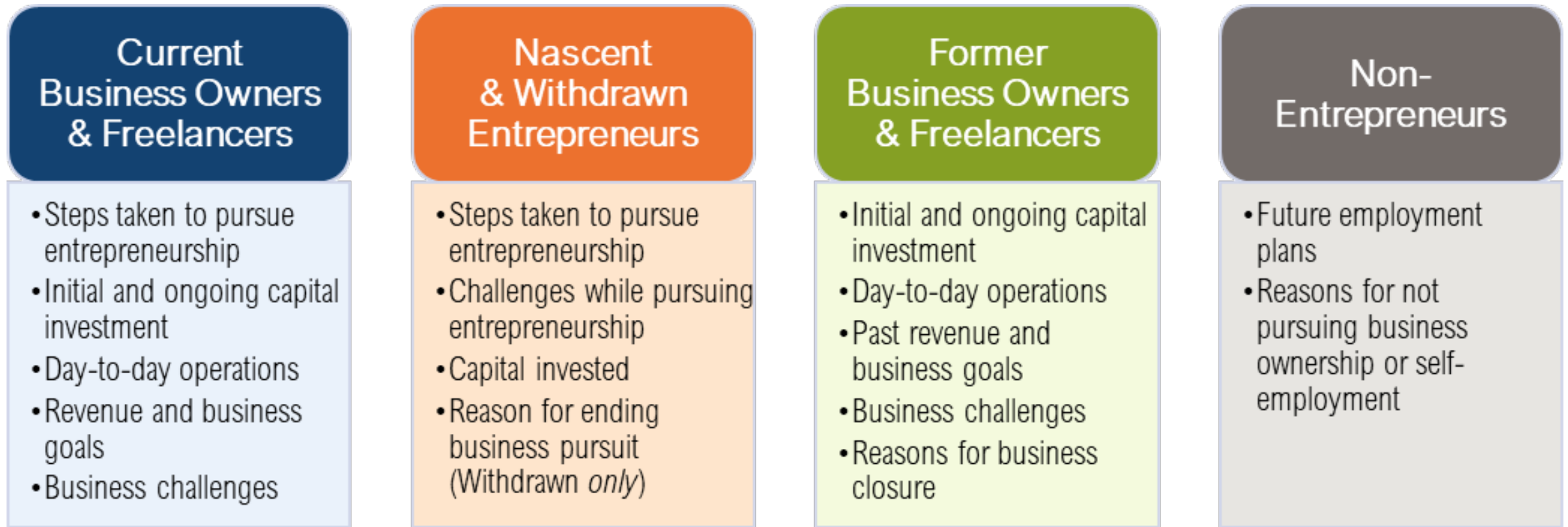# Measuring Entrepreneurial Activities – the EPOP Survey Approach

The EPOP Survey separately measures multiple, non-exclusive entrepreneurial activities to measure the full spectrum of business formation activity in the U.S.

- The EPOP Survey screener is intentionally designed to capture the full range of entrepreneurial activities.

- Individuals can report the variety of entrepreneurial activities in which they are engaged.

- Taking a modular approach to capturing entrepreneurial activity is designed to support myriad research needs.

Survey Pathways and Prioritization

- By design, some respondents report engagement in multiple entrepreneurial activities.

- To limit the survey burden, respondents are assigned to a specific survey pathway focused on one entrepreneurship activity based on analytic priority.

- The survey pathway logic prioritizes current business owners/freelancers, followed by nascent entrepreneurs, former business owners/freelancers, and withdrawn entrepreneurs.

- Individuals engaged in no entrepreneurial activities receive questions asking about their perceptions of entrepreneurship and future employment plans.

- There is one exception – gig work! Any respondent engaged in gig work is asked about it regardless of pathway or prioritization.

- Further information on the pathways is available in the EPOP Survey release brief:

    https://epop.norc.org/content/dam/epop/media/publications/pdf/epop2022-brief_american_entrepreneurship.pdf

# Summarized Survey Pathways and Follow-up Survey Items Collected

Current Business Owners & Freelancers

- Steps taken to pursue entrepreneurship
- Initial and ongoing capital investment
- Day-to-day operations
- Revenue and business goals
- Business challenges

Nascent & Withdrawn Entrepreneurs

- Steps taken to pursue entrepreneurship
- Challenges while pursuing entrepreneurship
- Capital invested
- Reason for ending business pursuit (Withdrawn *only*)

Former Business Owners & Freelancers

- Initial and ongoing capital investment
- Day-to-day operations
- Past revenue and business goals
- Business challenges
- Reasons for business closure

Non-Entrepreneurs

- Future employment plans
- Reasons for not pursuing business ownership or self-employment

The EPOP Survey is quite literally a "new" resource!

The Year 1 EPOP Survey data or EPOP:2022 is "hot off the presses"!

• Year 1 EPOP Survey collection closed in June 2022.

• The fully weighted data set was finalized in October.

• Initial analysis of the data has focused on:

  – National estimates of entrepreneurial activity by major demographic groups

  – Understanding the incidence of gig work

  – Top reasons for and steps taken towards entrepreneurship and business formation

  – State and MSA estimates for the interactive dashboard on EPOP@norc.org

Unweighted Estimates of Entrepreneurship Activity by Sample Type
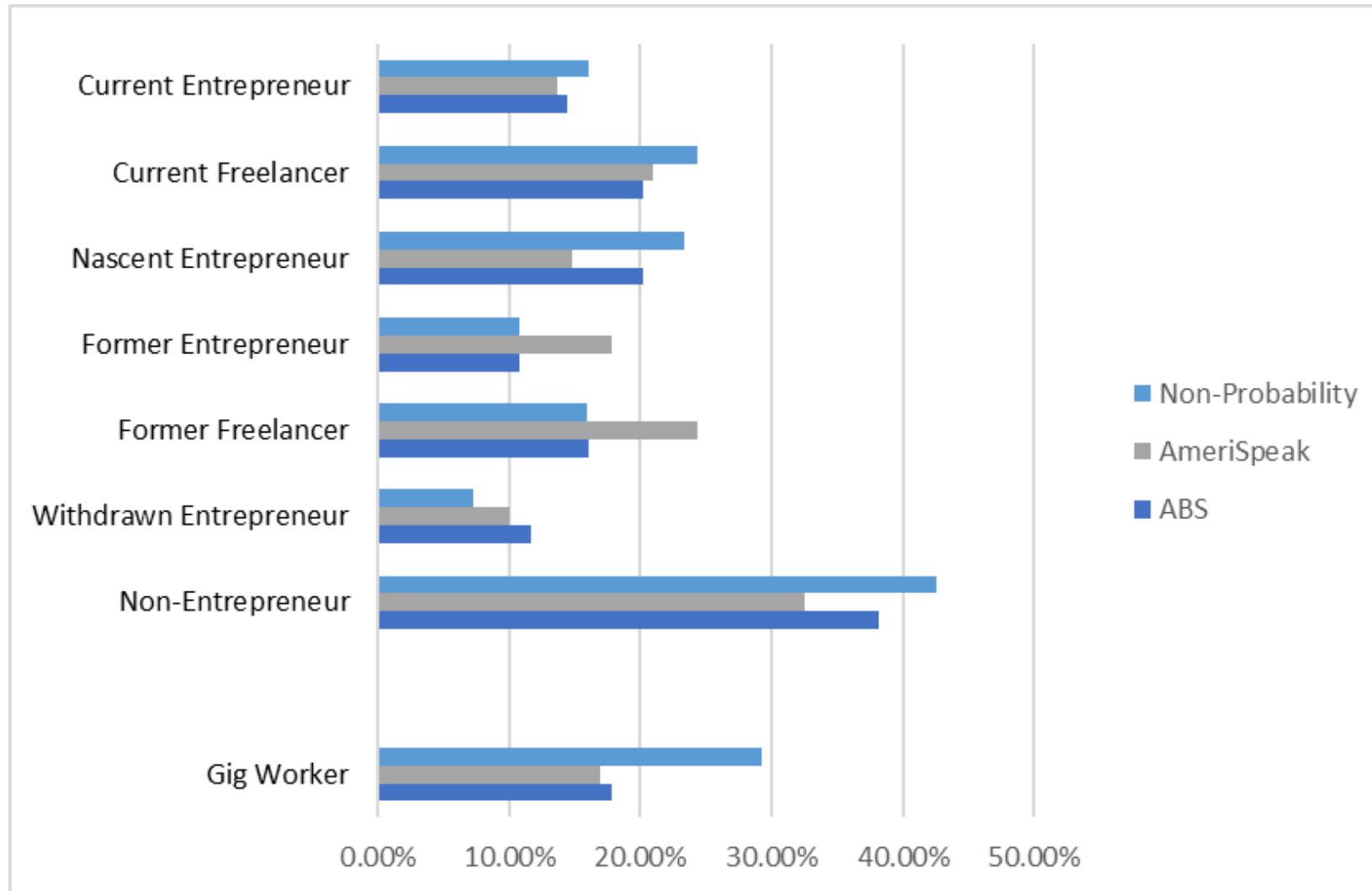
Source: NORC, Entrepreneurship in the Population Survey: 2022.

# Weighted Estimates of Entrepreneurship Activity for Total Sample and Probability Sample Only
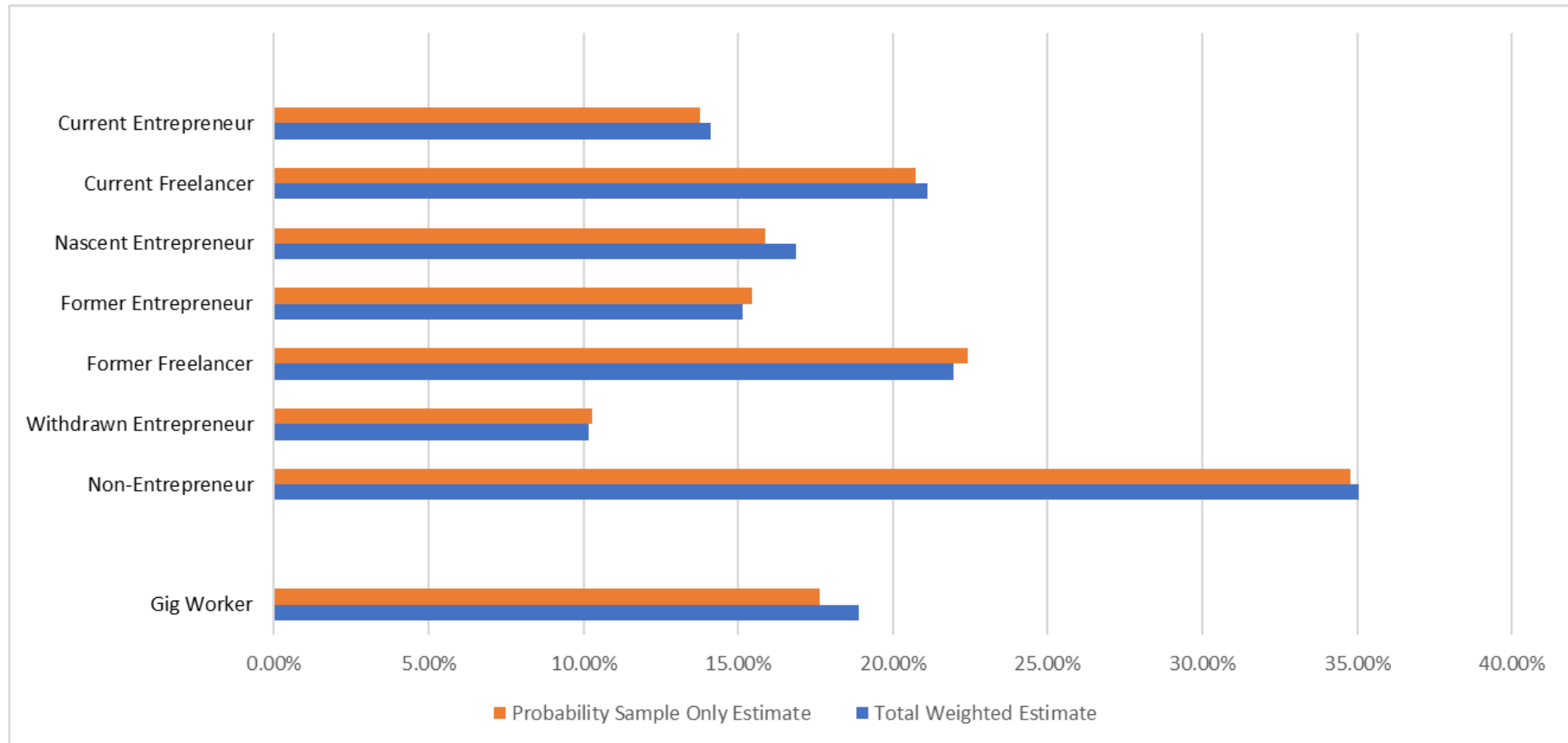
Source: NORC, Entrepreneurship in the Population Survey: 2022.

Men are more likely than women to report working as current business owners, yet both men and women are equally likely to report currently planning to open a business.
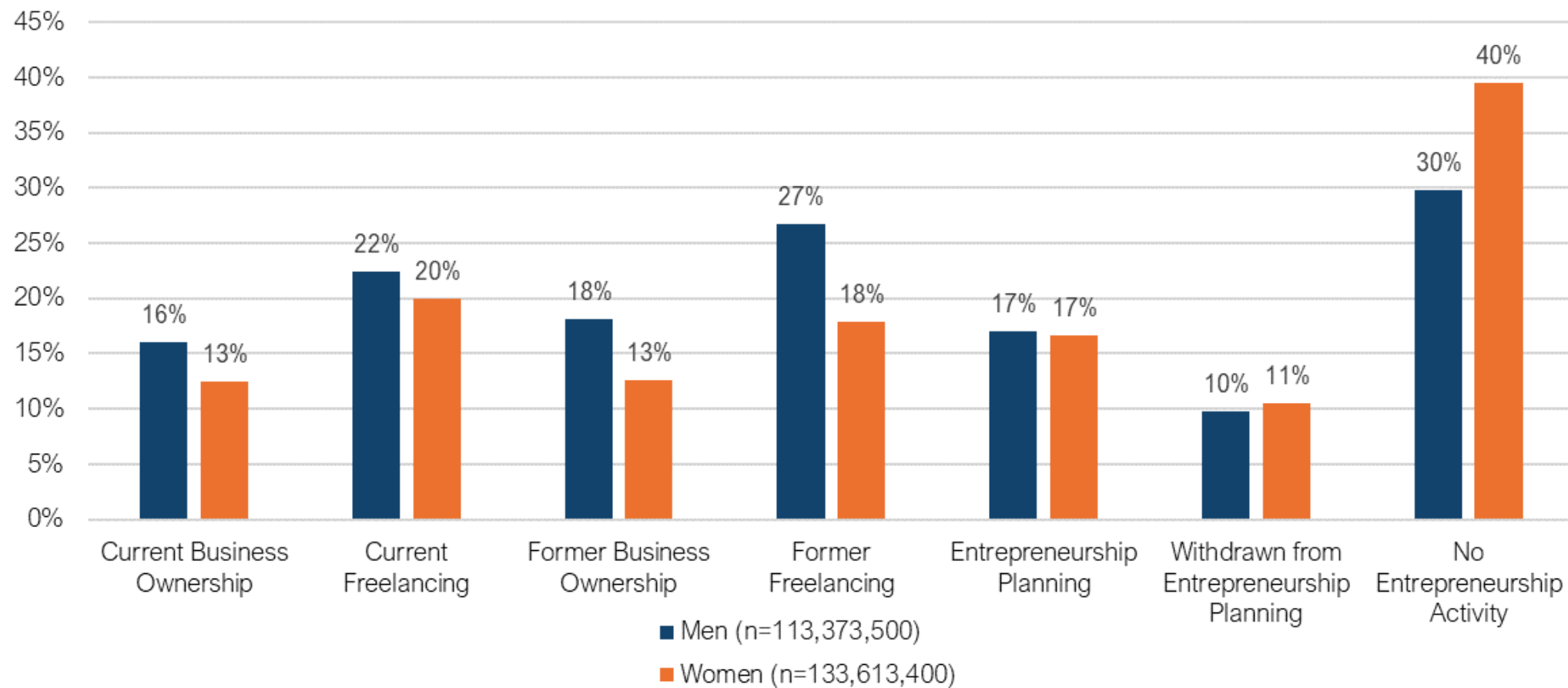
Source: NORC, Entrepreneurship in the Population Survey: 2022.

Entrepreneurial activities vary by race and ethnicity of respondents.
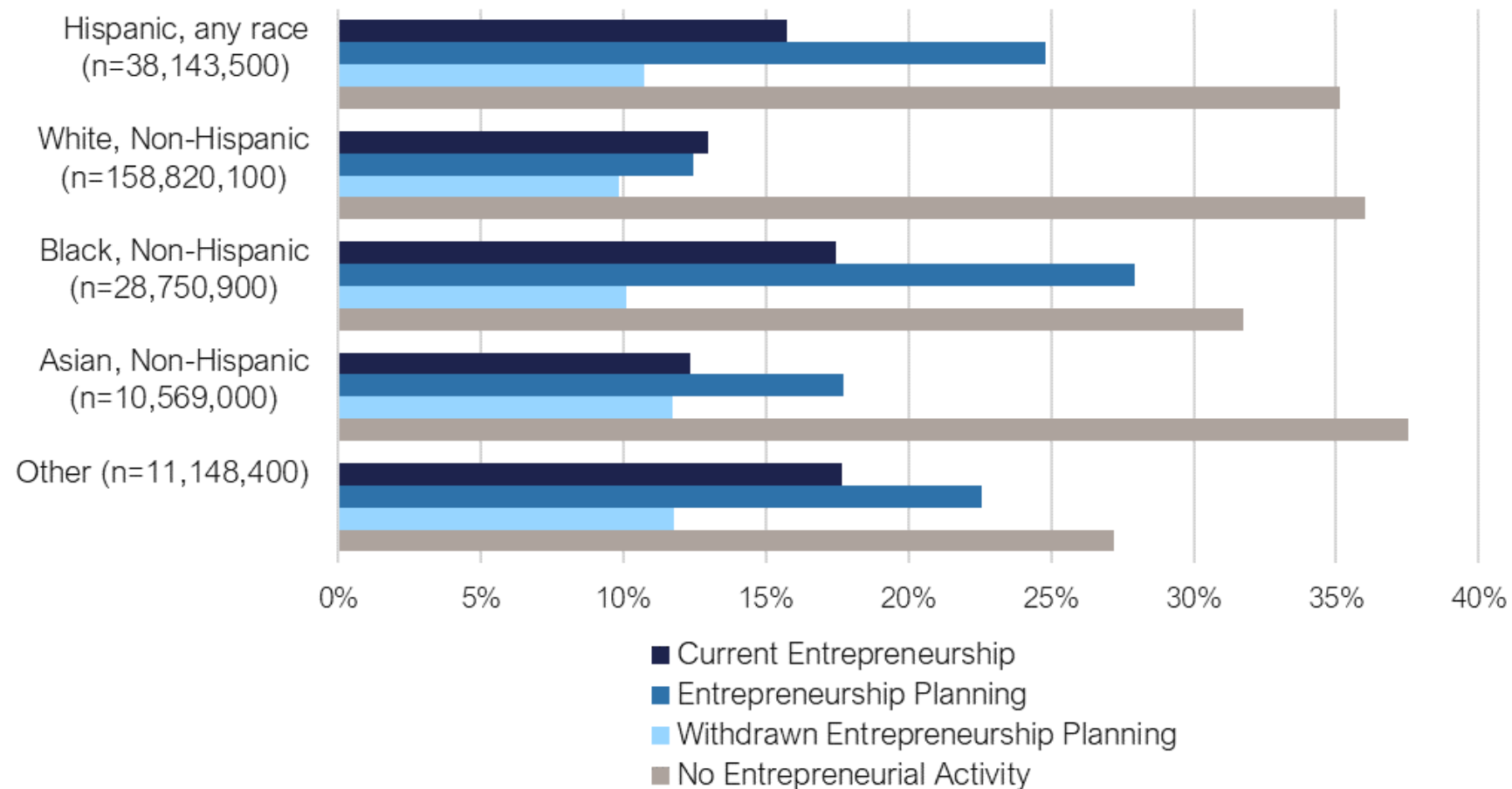
Note: "Other" includes non-Hispanic multi-race, non-Hispanic other reported races, and not reported.
Source: NORC, Entrepreneurship in the Population Survey: 2022.

Many individuals in the U.S. report being involved in the gig economy, which is strongly related to other entrepreneurial activities.

Entrepreneurship Activity Status	Reported Gig Worker Status	
	Not a Gig Worker (n=200,651,300)	Gig Worker (n=46,780,500)
No Entrepreneurial Activity	40.8	10.6
Any Entrepreneurial Activity	59.2	89.4
Current Business Ownership	10.9	27.7
Current Freelancing	12.6	57.7
Entrepreneurship Planning	11.0	42.0
Former Business Ownership	15.6	12.9
Former Freelancing	22.7	18.8
Withdrawn Entrepreneurship Planning	11.3	5.5

Source: NORC, Entrepreneurship in the Population Survey: 2022.

Optional ways to use survey items to measure incidence of gig work
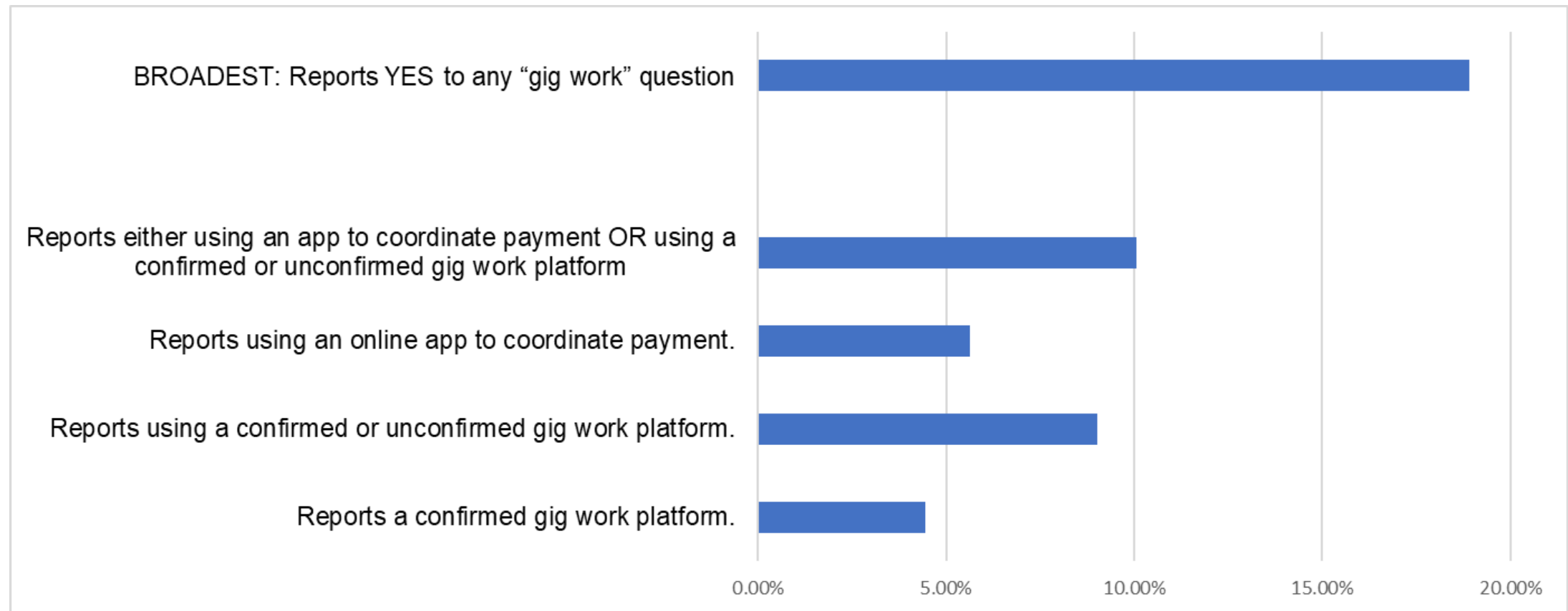
Source: NORC, Entrepreneurship in the Population Survey: 2022.
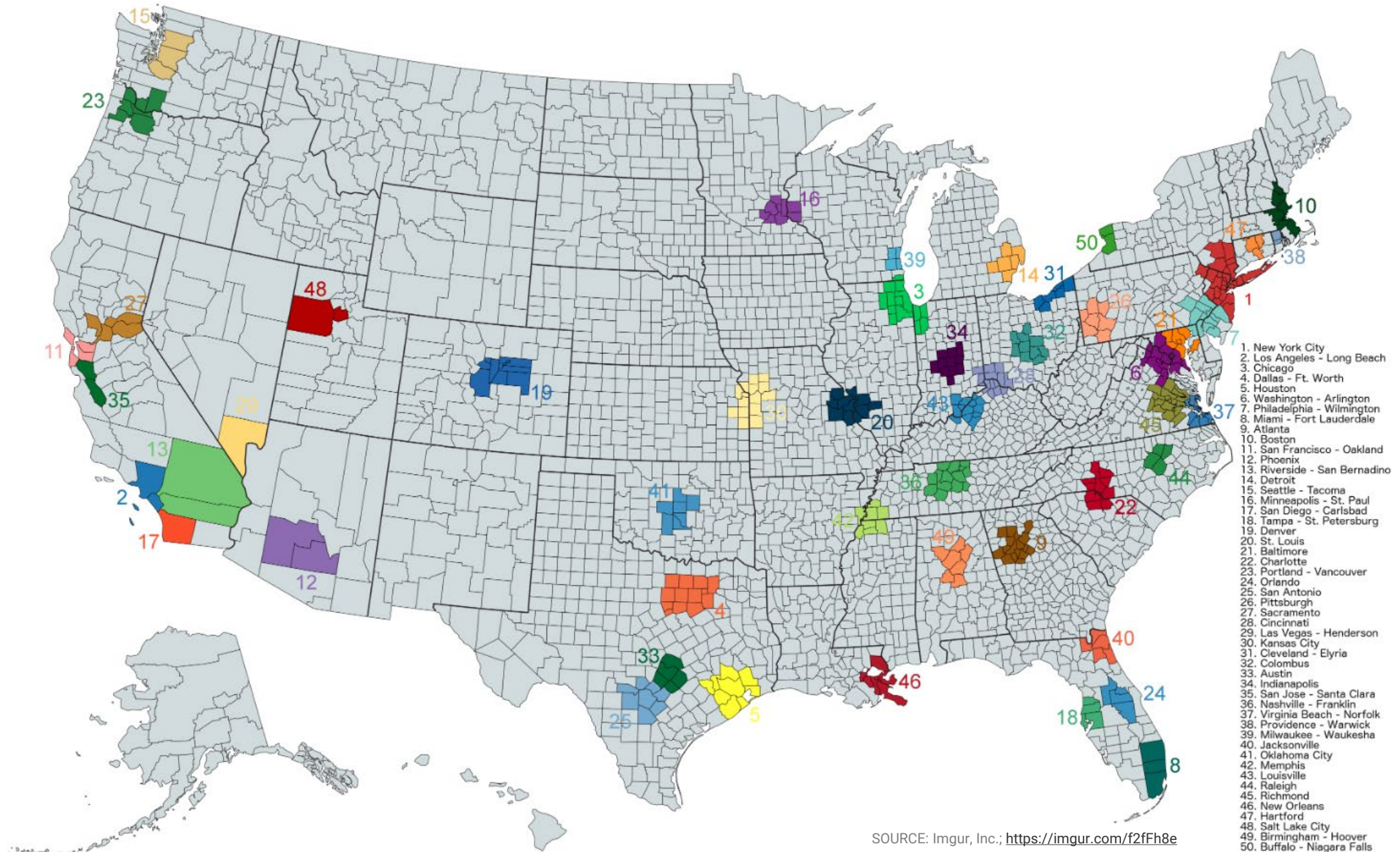
Sample and Study Design

The EPOP Study Design Overview

- The EPOP Survey is designed to understand the scope of entrepreneurial activities from the target population
    - Noninstitutionalized adults 18 years and older
    - Living in the United States

- Additionally, the survey captures
    - A robust set of demographic characteristics
    - Information on the behaviors, challenges, and resources available to individuals during the entrepreneurial process
    - The reasons for leaving or never engaging in entrepreneurship or self-employment
    - A brief topical question module updated annually; EPOP:2022 focused on the pandemic impact

- The design includes ~30,000 annual, cross-sectional surveys
    - 32,021 complete surveys in the EPOP:2022 data set

- The survey is be conducted for 5 annual cycles from 2022 through 2026

EPOP sample is designed to support various estimation objectives

The EPOP sample design supports the following estimation objectives:

- National estimates by demographics such as race/ethnicity, gender, age, and education, but not necessarily by the cross of these demographic variables

- State level estimates by race/ethnicity and gender, but not necessarily by the cross of these variables

- Metropolitan statistical area (MSA)-level estimates by race/ethnicity and gender for the fifty most populous MSAs per the 2020 Census, but not necessarily by the cross of these variables
    - The 50 most populous MSAs per the 2020 decennial census

1. New York City
2. Los Angeles – Long Beach
3. Chicago
4. Dallas – Ft. Worth
5. Houston
6. Washington – Arlington
7. Philadelphia – Wilmington
8. Miami – Fort Lauderdale
9. Atlanta
10. Boston
11. San Francisco – Oakland
12. Phoenix
13. Riverside – San Bernadino
14. Detroit
15. Seattle – Tacoma
16. Minneapolis – St. Paul
17. San Diego – Carlsbad
18. Tampa – St. Petersburg
19. Denver
20. St. Louis
21. Baltimore
22. Charlotte
23. Portland – Vancouver
24. Orlando
25. San Antonio
26. Pittsburgh
27. Sacramento
28. Cincinnati
29. Las Vegas – Henderson
30. Kansas City
31. Cleveland – Elyria
32. Colombus
33. Austin
34. Indianapolis
35. San Jose – Santa Clara
36. Nashville – Franklin
37. Virginia Beach – Norfolk
38. Providence – Warwick
39. Milwaukee – Waukesha
40. Jacksonville
41. Oklahoma City
42. Memphis
43. Louisville
44. Raleigh
45. Richmond
46. New Orleans
47. Hartford
48. Salt Lake City
49. Birmingham – Hoover
50. Buffalo – Niagara Falls

SOURCE: Imgur, Inc.; https://imgur.com/f2fFh8e

Three sample sources for statistical efficiency and cost effectiveness

AmeriSpeak® Panel Sample (n=25,190)

- A large probability-based panel designed to represent the U.S. household population, constructed and maintained by NORC.

  – Sample selection probabilities are known.

Addressed Based Sample (ABS) (n=25,883)

- The ABS frame is constructed from the Delivery Sequence File (DSF) used by the USPS for mail delivery across the country; the DSF contains all delivery point addresses serviced by the USPS.

  – Sample selection probabilities are known.

Nonprobability Panel Samples

- The nonprobability or opt-in panels are composed of volunteers who were recruited online and who receive some compensation for completing online surveys.

  – Sample selection probabilities are unknown and need to be estimated.

A stratified sampling design is used to select the probability sample (AmeriSpeak and ABS)

Primary sampling strata (geographies):

- Each state that does not contain any MSAs constitutes a primary sampling stratum

- For states that contain one or more MSAs, each MSA and the rest of state outside the MSA(s) make a primary sampling stratum

- MSAs that include counties from multiple states are divided into multiple primary sampling strata, one for each state

Secondary sampling strata:

- Each primary stratum is further divided into three secondary sampling strata: Hispanic, non-Hispanic Black, and non-Hispanic Other

Probability sample allocation ensures a minimum sample size for key analysis domains

- Each state (including DC) receives a total of 175 complete surveys from the probability samples to support state-level estimation;

- Each of 50 MSAs receives a total of 100 complete surveys from the probability samples to support MSA-level estimation;

- Within each state, the Hispanic and non-Hispanic Black substrata each receives 20 percent of the total state allocation, or twice the respective state population benchmark percent for each group, whichever is lower;

- Within each MSA, the Hispanic and non-Hispanic Black substrata each receives 25 percent of the total MSA allocation, or twice the respective MSA population benchmark percent for each group, whichever is lower.

Nonprobability samples are quota samples with unknown population coverage and sample selection mechanism

- The nonprobability samples are selected by the sample vendors based on the target number of complete surveys by geography and race/ethnicity provided by NORC

- Nonprobability panels may not have complete coverage of the target population

- The sample selection probabilities are unknown for the nonprobability samples

- Model-based methods are needed to develop survey weights for the nonprobability samples

Data Collection and Processing Timeline

Data collection ran from February 15 to June 6, 2022

– Collection for the AmeriSpeak sample began February 15 and ended June 6

– Collection for the ABS began February 28 and ended June 6

– Collection for the nonprobability sample began May 3 and ended June 3

EPOP Survey data collection

Weighting and non-disclosure review

Data release and dissemination

EPOP:2022 TIMELINE

| Jan '22 | Feb to May '22 | Jun '22 | Jul '22 | Aug '22 | Sept '22 | Oct '22 |

Finalize EPOP Survey instrument and update programming

Data review and data processing

Data analysis and documentation

SOURCE: All icons sourced from https://www.flaticon.com/.
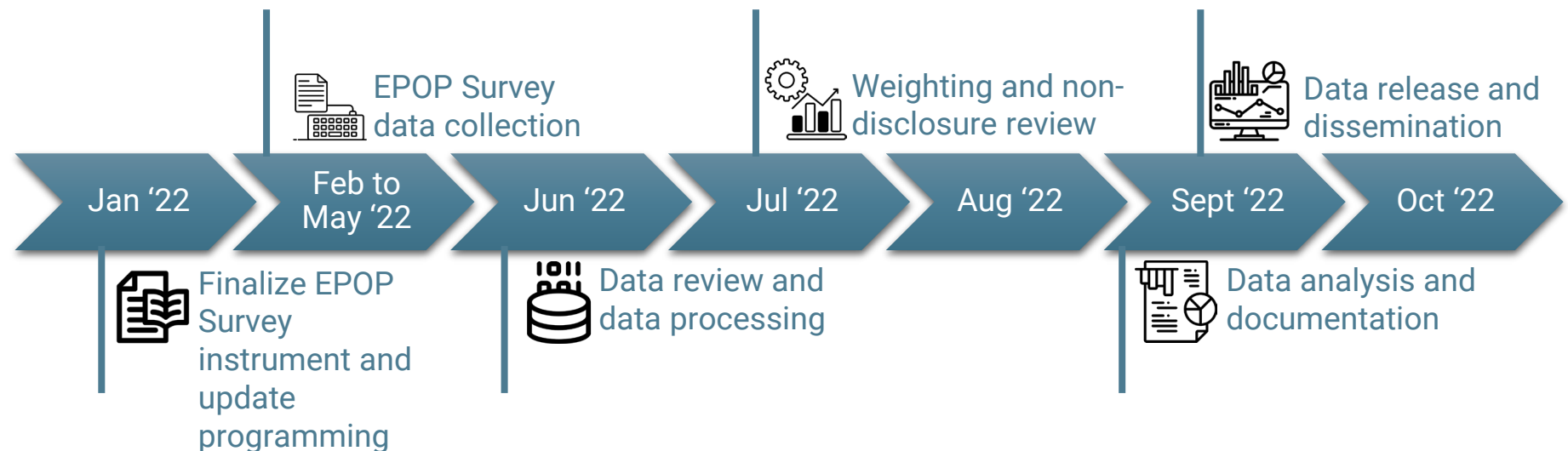
Data Collection Methods

- There were differential data collection protocols for each sample type:
    - The AmeriSpeak Panel sample was invited to participate by email and telephone prompting;
    - The ABS sample was invited to participate by USPS mail, postcard, email, and telephone prompting; and
    - The nonprobability panel participants were recruited via a variety of methods including email and digital network and social media ads.

- Partially completed surveys in the AmeriSpeak and ABS samples were followed-up with emails and prompting calls.

- Data was primarily collected via an online survey. Computer-assisted telephone interviewing (CATI) was a secondary mode available upon request (n=474).

- Respondents were able to take the online survey or phone interview in English or Spanish.

- All respondents were compensated for their participation.

Data Processing

- Complete surveys from all sample sources were combined, reviewed, and processed

  - Raw data was compared to the programmed survey specifications to identify data irregularities

  - Data anomalies were investigated and recoded or eliminated

  - Complete and partial survey records were all reviewed for data integrity and cases in violation were discarded

    - Respondents who completed the survey too quickly to have read question text were discarded

    - Respondents who straight-lined responses were discarded

  - Cases with high item nonresponse or missing critical pathway information were dropped from the final data set

- 32,021 surveys are included in the final EPOP:2022 data set

Item response threshold by sample source and entrepreneurial activity group

Entrepreneurial Activity Group	Sample Type		
	AmeriSpeak	**ABS**	**Nonprobability**
Current Entrepreneur	75%	75%	80%
Current Freelancer	75%	75%	80%
Nascent Entrepreneur	75%	75%	80%
Former Entrepreneur	75%	75%	75%
Former Freelancer	75%	75%	75%
Withdrawn Entrepreneur	70%	70%	70%
General Population/Non-Entrepreneur	70%	70%	70%

Response Rates

Response rate varied by sample type

– For the AmeriSpeak Panel sample

– The survey response rate was 38.2%

– The weighted combined panel recruitment was 23%

– The weighted retention was 82.1%

– For the ABS sample, the response rate was about 6.0%

– For the nonprobability survey panels, the response rate is not reported

Composition of EPOP:2022 complete surveys across the three sample sources

AmeriSpeak (n=9,623) 30%

Nonprobability (n=20,847) 65%

ABS (n=1,551) 5%
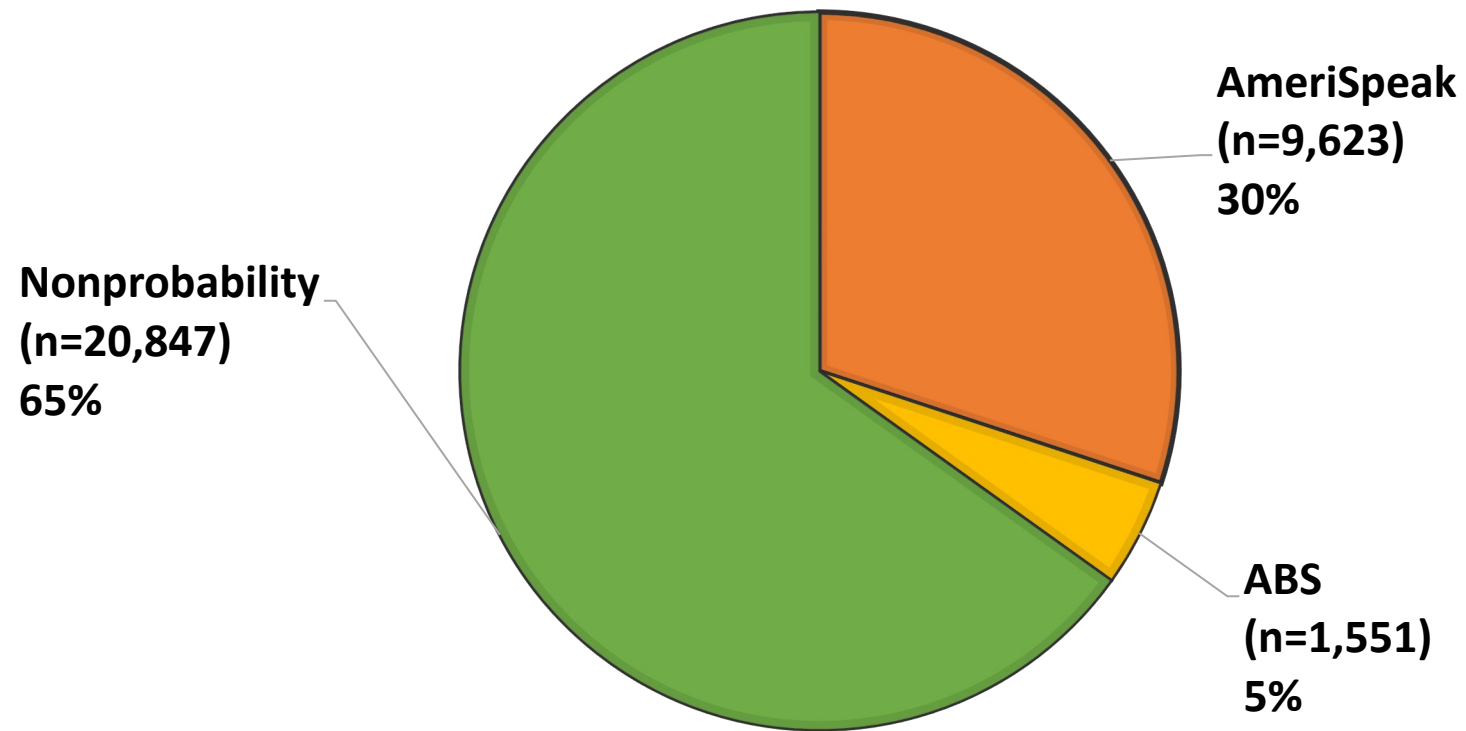
Using the Data

EPOP Survey data are available in both public use and restricted use files

	Public Use File (PUF)	**Restricted Use File (RUF)**
Availability to Users	Readily available on EPOP website	Application process required
Geography	Census region & division (9 categories)	Census region & division, State, Metropolitan Statistical Area, Rural-Urban Area, County
Sample Type Differentiation	No information on sample type is provided	Indicator for address-based (ABS), AmeriSpeak®, nonprobability samples
Weights	Survey weight for combined samples	Survey weight for combined sample AND survey weight for cases in probability-based samples

For more detailed information about PUF and RUF differences:

https://epop.norc.org/content/dam/epop/researchers/pdf/EPOP_DUA_Attachment%20C_RUF_and_PUF_Variable_Comparison.pdf

Aggregation & Data Protection

In some instances, survey responses for a given variable were aggregated, rounded, or recoded to protect respondent confidentiality.

Data privacy changes are indicated in the data files with the variable name suffixes _PUF, _RUF, or _DRV.

To address small cell counts in cross-tabulations, value suppression was also utilized for some variables. Suppressed values in the PUF and RUF are indicated by -7.

Variable	Datafile	Data Protection Type	Name Convention	Example	# of Suppressions
State	RUF only	Original Variable	No change	DEM_STATE	0
Education	RUF only	Aggregated for RUF (7 levels)	_RUF	DEM_EDU_RUF	884
Education	PUF only	Aggregated for PUF (4 levels)	_PUF	DEM_EDU_PUF	884
Census Region	RUF and PUF	Aggregated for RUF and PUF	_DRV	REGION_DRV	0

Accessing public use data

https://epop.norc.org/us/en/epop/researchers/public-use-files.html

https://epop.norc.org

- Researchers

- PUBLIC USE FILES

- PUF request form is at the bottom of this web page

EPOP
Understanding Entrepreneurship in the US

RESEARCHERS ∨ PARTICIPANTS ∨ MEDIA ∨ ABOUT THE STUDY ∨

PUBLIC USE FILES

ABOUT THE DATA

EPOP survey data is available for researchers to explore, analyze, and share findings on entrepreneurship in the US. Public use files (PUFs) are comprehensive data sets available to the general public without restriction. PUFs are fully anonymized, meaning that the risk of re-identifying survey respondents is negligible. This data supports analysis at national, state, and regional levels.

To request EPOP PUFs, please complete the form below and agree to the Conditions of Use. Required fields are marked with an asterisk (*).

When the request form is submitted, the EPOP research team will email you instructions on accessing the PUF data package which includes:

✳ Data files and formatting programs
✳ Questionnaire
✳ PUF Codebook
✳ User Guide
✳ Variable Crosswalk (shows the survey variables and survey pathways for the different categories of entrepreneurship)

Entrepreneurship Category and Activity Variables

Assigned Entrepreneurship Categories by Reported Entrepreneurship Activities

Survey Pathway Priority	Assigned Entrepreneurship Category	Total Surveys	Reported Entrepreneurship Activity					
			Current Business Ownership	Current Freelancing	Entrepreneurship Planning	Former Business Ownership	Former Freelancing	Withdrawn Entrepreneurship Planning
1	Current Business Owner	4,907	4,907	3,189	2,892	0**	728	0
2	Current Freelancer	4,213	0	4,213	1,754	842	0**	0
3	Nascent Entrepreneur	1,467	0	0	1,467	257	274	0
4	Former Business Owner	3,030	0	0	136*	3,030	1,766	0
5	Former Freelancer	3,144	0	0	347*	0	3,144	0
6	Withdrawn Entrepreneur	2,649	0	0	0	0	0	2,649
7	Non-Entrepreneur	12,611	0	0	0	0	0	0
	Total	32,021	4,907	7,402	6,596	4,129	5,912	2,649

*These individuals worked as a freelancer, consultant, or independent contractor within the last 5 years of the survey administration date, although they report no longer working in that capacity of self-employment.

**There is no former business ownership activity reported for current business owners and no former freelancing or consulting activity for those categorized as current freelancers. This is a result of the survey construction as the Year 1 version of the EPOP Survey does not capture former entrepreneurship activity when an individual is currently engaged in the activity. The survey questionnaire will be revised for Year 2 to capture this type of former entrepreneurial activity for current business owners and freelancers.

Source: NORC, Entrepreneurship in the Population Survey: 2022.

Entrepreneurship Category and Activity Variables

Entrepreneurship Category	Variable	Entrepreneurship Activity	Variables (binary)
Current Business Owner	DOV_GROUP == 1	Current Business Ownership	DOV_CUR_ENTR == 1
Current Freelancer	DOV_GROUP == 2	Current Freelancing	DOV_CUR_FREE == 1
Nascent Entrepreneur	DOV_GROUP == 3	Entrepreneurship Planning	DOV_NASCENT == 1
Former Business Owner	DOV_GROUP == 4	Former Business Ownership	DOV_FORM_ENTR == 1
Former Freelancer	DOV_GROUP == 5	Former Freelancing	DOV_FORM_FREE == 1
Withdrawn Entrepreneur	DOV_GROUP == 6	Withdrawn Entrepreneurship Planning	DOV_WITHDRAW == 1
Non-Entrepreneur	DOV_GROUP == 7	No Entrepreneurship Activity	DOV_GENPOP == 1

Taking a closer look at follow-up survey items by assigned entrepreneurship category

Current Business Owners & Freelancers

- Steps taken to pursue entrepreneurship
- Initial and ongoing capital investment
- Day-to-day operations
- Revenue and business goals
- Business challenges

Nascent & Withdrawn Entrepreneurs

- Steps taken to pursue entrepreneurship
- Challenges while pursuing entrepreneurship
- Capital invested
- Reason for ending business pursuit (Withdrawn *only*)

Former Business Owners & Freelancers

- Initial and ongoing capital investment
- Day-to-day operations
- Past revenue and business goals
- Business challenges
- Reasons for business closure

Non-Entrepreneurs

- Future employment plans
- Reasons for not pursuing business ownership or self-employment

Navigating the EPOP Survey Pathways With the Survey Item Crosswalk

- Because the EPOP Survey screens for multiple entrepreneurial activities, this sometimes leads to complexity in the follow-up survey pathways.

- There is a Survey Item Crosswalk to help researchers understand the survey design and facilitate data usage.

- The Crosswalk:
  - Includes all EPOP Survey items
  - Highlights and color codes the survey items asked of each assigned entrepreneurship category
  - Shows the survey items, two ways – survey items by assigned category and assigned category by survey items
  - Is provided to all data users in an Excel format to allow for filtering and isolating items
  - The EPOP: 2022 Survey item Crosswalk may be found here:

    https://epop.norc.org/content/dam/epop/researchers/pdf/epop2022_survey_item_crosswalk.pdf

Steps taken towards entrepreneurship

"The next few questions ask about different steps you may have taken to pursue or develop your business or working for yourself"

Current, former, nascent, and withdrawn entrepreneurs respond to 32-item module.

"Which of the following <u>networking steps</u> did you (or you and your co-developers/collaborators) take to pursue owning your own business?

1. Discussed the idea with a friend, work colleague, classmate, or acquaintance
2. Discussed the idea with a family member
3. Identified and worked with a mentor(s)
4. Networking with experts, colleagues, or acquaintances in the field
5. None of the above

Topic	# of Items
Networking	5
Technical or market research	7
Business development steps with business support organizations	3
Business financing	5
Organizational planning	7
Staffing and growth	5
Total	**32**

Using the "steps" variables

Variable stem is "`PE_STEPS_`"

Each variable is an indicator for each of the 32 items `PE_STEPS_1_1 – PE_STEPS_6_5`

Flexible in for analyses – for example, users can summate each item for a count of steps taken or generate an indicator variable for each of the 6 main planning areas

	PE_STEPS_1_1	PE_STEPS_1_2	PE_STEPS_1_3	PE_STEPS_1_4	PE_STEPS_1_5	PE_STEPS_2_1	PE_STEPS_2_2	PE_STEPS_2_3	PE_STEPS_2_4	PE_STEPS_2_5	PE_STEPS_2_6	PE_STEPS_2_7
1	0	1	1	0	0	0	1	0	1	1	1	0
2	1	0	0	0	0	0	1	0	1	0	0	0
3	1	0	1	0	0	0	1	0	1	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0	1
5	1	0	1	0	0	1	1	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0	1
7	0	0	0	1	0	0	1	1	0	0	0	0
8	0	1	0	0	0	0	1	0	0	0	0	0
9	0	1	0	0	0	0	0	0	1	0	0	0
10	1	0	0	1	0	1	0	1	0	0	0	0

Challenges faced while pursuing entrepreneurship

"Now we want to ask you about some challenges you may or may not have encountered while pursuing owning your own business"

Nascent and withdrawn entrepreneurs respond to 29-item module

"Which of the following <u>customer reach challenges</u> (have you encountered/did you encounter) while pursuing owning your own business?"

1. Finding Customers
2. Setting up the business' digital/online presence
3. Securing a physical location for the business
4. None of the above

Current and former entrepreneurs are given a similar set of questions regarding challenges *while* being an entrepreneur

The coding of both challenges modules is equivalent to the steps module

– Variable stems: `PE_CHALLENGE_` & `BO_CHALLENGE_`

Topic	# of Items
Financial or economic security	6
Business operations	6
Customer reach	4
Resource or support	8
Economy or market	5
Total	**29**

Capital Variables

All respondents who reported current or previous entrepreneurship activities were asked about their business financing related to starting up or pursuing their business

- Financing sources and funding amounts in dollars

- Did respondents receive as much as was requested from each source?

- Did respondents try to obtain funds from other sources that they didn't receive?

Current and former business owners and freelancers were also asked about additional financing after the initial start-up phase

Using the capital variables on initial funding sources

First, respondents indicated which of 14 funding sources they used to start their business. These variables are treated as 14 indicator variables in the data files `(PE_CAPITAL_1_1-PE_CAPITAL_1_14)`

Next, respondents provide funding amounts in dollars for 6 funding areas `(PE_CAPITAL_4-PE_CAPITAL_9)`:

1. Total amount used

2. Personal capital

3. Capital from family, friends, and employees

4. Capital from banks or financial institutions

5. Capital from outside investors

6. Capital from government grants

Capital Visualization Examples

Initial Funding Sources among Current Entrepreneurs

(Bar chart)
- Personal/Friends/Family — ~90%
- Business Credit Card — ~15%
- Guaranteed Loan — ~13%
- Gov Loan — ~7%
- Grant — ~4%
- Bank Loan — ~3%
- Other Sources — ~7%
- Funding Not Needed — ~10%

X-axis: 0% 20% 40% 60% 80% 100%

"What was the total amount of personal capital used to start your business?"

(Pie chart legend)
- Less than $100
- $100 - $999
- $1,000 - 4,999
- $5,000 - 9,999
- $10,000 - 24,999
- $25,000 - 49,999
- At least $50,000

NOTE: This source of capital includes personal savings, personal retirement accounts, home equity loans, and personally borrowed funds.

Administrative Variables and Values

Respondent ID – R_SUID

Sample Type (RUF only) – SAMP_TYPE

PARTIAL_CASE – partial complete cases

Q_LANGUAGE – survey language

Missingness	
-7	suppressed
-5	don't know
-3	refused
.	logical skip

The RUF and PUF are available in Stata, SAS, and .csv file formats

EPOP
Understanding Entrepreneurship in the US

Hello Katie Johnson,

Thank you for visiting the EPOP website and requesting the Public Use Files (PUFs) for the Entrepreneurship in the Population (EPOP) study.

You can now download the PUF data package here.

This package includes the following data and documentation:

- Data files and formatting programs:
    - CSV data file (EPOP_YR1_PUF.csv)
    - STATA data file with formatting applied (EPOP_YR1_PUF.dta)
    - SAS data file (EPOP_YR1_PUF.sas7bdat)
    - SAS program containing labels (EPOP_YR1_PUF_LABELS.sas)
    - SAS program containing formats (EPOP_YR1_PUF_FORMATS.sas)
    - SAS program to apply labels and formats (EPOP_YR1_PUF_APPLY_FORMATS_LABELS.sas)
- Questionnaire
- PUF Codebook (EPOP_YR1_PUF_CODEBOOK.pdf)
- User Guide
- Variable Crosswalk (shows pathways through the survey based on screener assignment)

EPOP provides two sets of analysis weights to support approximately unbiased estimation of population quantities from the complete surveys

Probability sample weights:

- Developed for complete surveys from the probability samples (AmeriSpeak and ABS)
- Available only in the restricted use file (RUF)

Combined sample weights:

- Developed for complete surveys from all three sample sources
- Available in both the public use file (PUF) and the RUF

Probability sample weights are developed in three general stages

Base weights:

- Computed as the inverse of the sample selection probabilities
- Compensate for sampling frame units not included in the study sample

Nonresponse adjusted weights:

- Computed as the product of the base weights and the nonresponse adjustment factor
- Compensate for eligible sample units that fail to complete the survey
- Nonresponse weighting variables: geography, race/ethnicity, gender, age, and education

Final weights:

- Calibrate the nonresponse adjusted weights to population benchmarks
- Compensate for sampling frame coverage errors
- Calibration variables: geography, race/ethnicity, gender, age, and education, including interactions

Combined sample weights are developed using NORC's TrueNorth 2.0 method based on statistical matching and propensity weighting

Statistical matching:

- Statistically match each nonprobability sample unit to a probability sample unit using *StatMatch* in the R package
  - Matching is based on Gower's dissimilarity measure
  - The matching variables are determined using a gradient boosting tree model
  - Final matching variables include basic demographics such race, gender, education, martial status as well as substantive survey response variables
  - Matching divides the probability sample into two sets: the set of units that is matched to a nonprobability sample unit and the set of units that is not matched

Combined sample weights are developed using NORC's TrueNorth 2.0

Propensity weights for the nonprobability sample units:

- Concatenate the matched probability sample and the nonprobability sample
- Use a logistic regression model to estimate the inclusion probabilities of the nonprobability units
- Predictor variables:
    - Demographics like race, education, marital status
    - Substantive survey response variables
    - Ridge regression was used to produce a parsimonious model
- The weights for nonprobability units are computed as the inverse of the estimated probabilities
- Rescale the nonprobability weights such that the sum of the nonprobability sample weights is equal to the sum of the probability sample weights

Combined sample weights are developed using NORC's TrueNorth 2.0

Combined sample weights:

$$
w^{comb}{}_i = \begin{cases} w_i & i \in S_P^U \\ \lambda w_i & i \in S_P^M \\ (1 - \lambda)\, w_i^{NP} & i \in S_{NP} \end{cases}
$$

λ is determined to minimize the mean squared error for some key survey estimates

Final weight and quality assurance checking

The final population estimate for EPOP:2022 is 247,431,811 non-institutionalized U.S. residing adults

- 247,431,811 is the summed survey weight value (WTSURVY) for all 32,021 cases

When producing estimates for various groups or sub-groups be sure to run a quality assurance check:

- The best practice is to calculate the weighted estimates and unweighted case count
- Regardless of your analysis focus, include all sub-groups to verify the sums match the total estimate and case count
- EXAMPLE: When analyzing women, run the totals for the whole sample by gender (demographic variable of interest); verify the weighted estimate and case count sums to the correct totals, and then proceed to use the correct totals for women (the sub-group of interest) to QC all subsequent analysis:

Gender	Case Count	Weighted Estimate
Men	13,642	113,373,522
Women	18,197	133,613,371
Non-binary	158	387,521
Other	5	16,056
Missing	19	41,340
Total	32,021 ✓	247,431,811 ✓

✳NORC

Sample design variables for variance estimation

Design strata (STRATA)

- AmeriSpeak sample
  - Pseudo strata constructed to support variance estimation based on the AmeriSpeak Panel sample design
  - Represent the first stage sampling strata only
- ABS and nonprobability samples
  - Each MSA and the rest of the state constitutes a STRATA

Design PSUs (PSU)

- AmeriSpeak sample
  - Pseudo PSUs constructed to support variance estimation based on the AmeriSpeak Panel sample design
- ABS and nonprobability samples
  - ABS and nonprobability samples are not clustered, so each unit is their own PSU

Weighting and sample design

STATA	
Load data	`use EPOP_YR1_PUF.dta,clear`
Set survey design	`svyset [pweight=WTSURVY], ///` `    strata(STRATA) psu(PSU) singleunit(scaled)`
Weighted mean	`svy: mean AVAR`
Weighted percentage	`svy: proportion AVAR`
Weighted total	`svy: total AVAR`
Weighted one-way table	`svy: tabulate AVAR`
Subset mean	`svy, subpop (if SUBGROUP==1): mean AVAR`
Specifying subgroups	`svy: mean AVAR, over(GROUPVAR)`

R	

Install & load required packages	`install.packages(c("tidyverse", "survey"))`
	`library(haven)`
	`library(survey)`

Load data	`mydata <- read_dta("EPOP_YR1_PUF.dta")`
Set survey design	`mydesign <- svydesign(id = ~PSU, weights = ~WTSURVY, strata = ~STRATA, data = mydata, nest = TRUE)`
Singleton PSU correction	`options(survey.lonely.psu = "adjust")`
Weighted mean	`svymean(~AVAR, mydesign, na.rm = TRUE)`
Weighted total	`svytotal(~AVAR, mydesign, na.rm = TRUE)`
Weighted one-way table	`svytable(~AVAR, mydesign)`
Subset mean	`svymean(~AVAR, subset(mydesign, SUBGROUP == 1), na.rm = TRUE)`
Specifying subgroups	`svyby(~AVAR, by = ~GROUPVAR, mydesign, svymean)`

For more on weighting and variance estimation, see the User Guide:
https://epop.norc.org/content/dam/epop/researchers/pdf/epop-2022-user-guide.pdf

Using weights in Stata
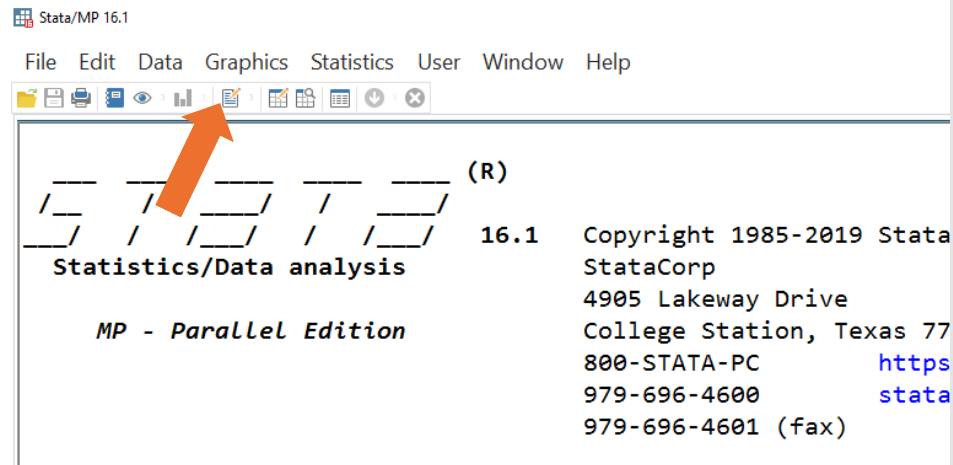

```stata
1   use "YOUR_FILE_PATHWAY_HERE/EPOP_YR1_PUF.dta", clear
2
3   * Specifying the survey design
4
5   svyset [pweight=WTSURVY], strata(STRATA) psu(PSU) singleunit(scaled)
6
7   * Weighted one-way table
8
9   svy: tabulate DEM_GENDER, count format(%9.0f)
10
11  * Weighted proportion
12
13  svy: proportion DOV_GROUP
14  svy: proportion DOV_GROUP, percent
15
16  * Subset Proportion
17
18  svy, subpop (if REGION_DRV==3): proportion DOV_GROUP
19
20  * Subgroups
21
22  tab BO_HOURSWK_1
23  recode BO_HOURSWK_1 (-5 = .) (-3 = .)
24  svy: mean BO_HOURSWK_1, over(DOV_GROUP)
```

Graphing with weights

```
graph hbar [w=WTSURVY],
over(DOV_GROUP)
```
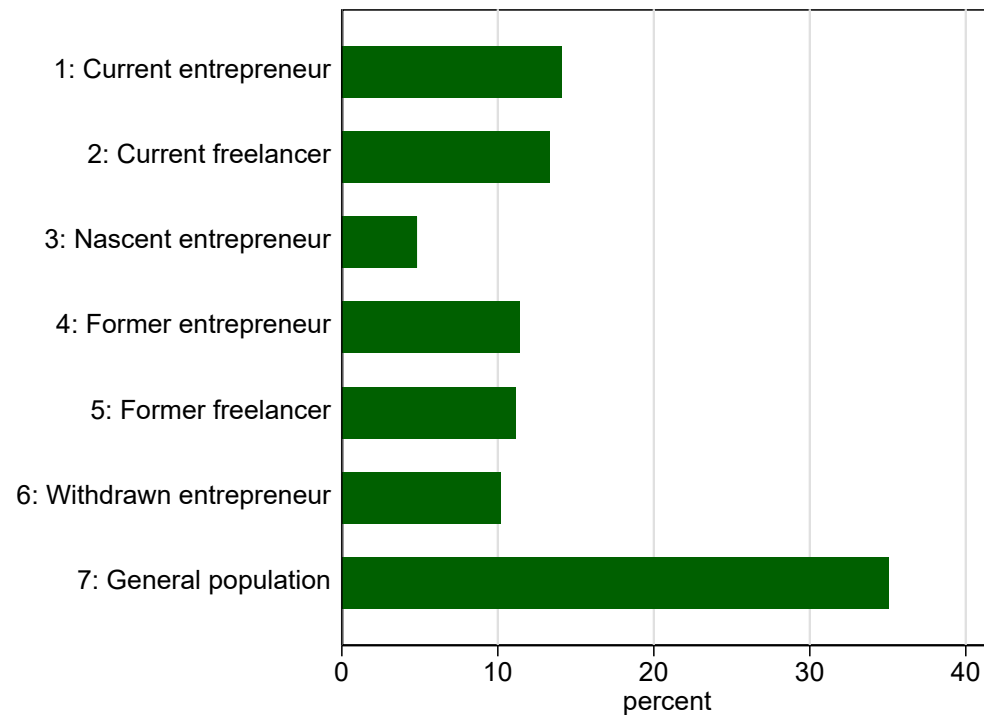
```
gen int_WTSURVY = int(WTSURVY)

histogram BO_HOURSWK_1 [w=int_WTSURVY],
percent bin(30) xtitle("Weekly Hours")
```
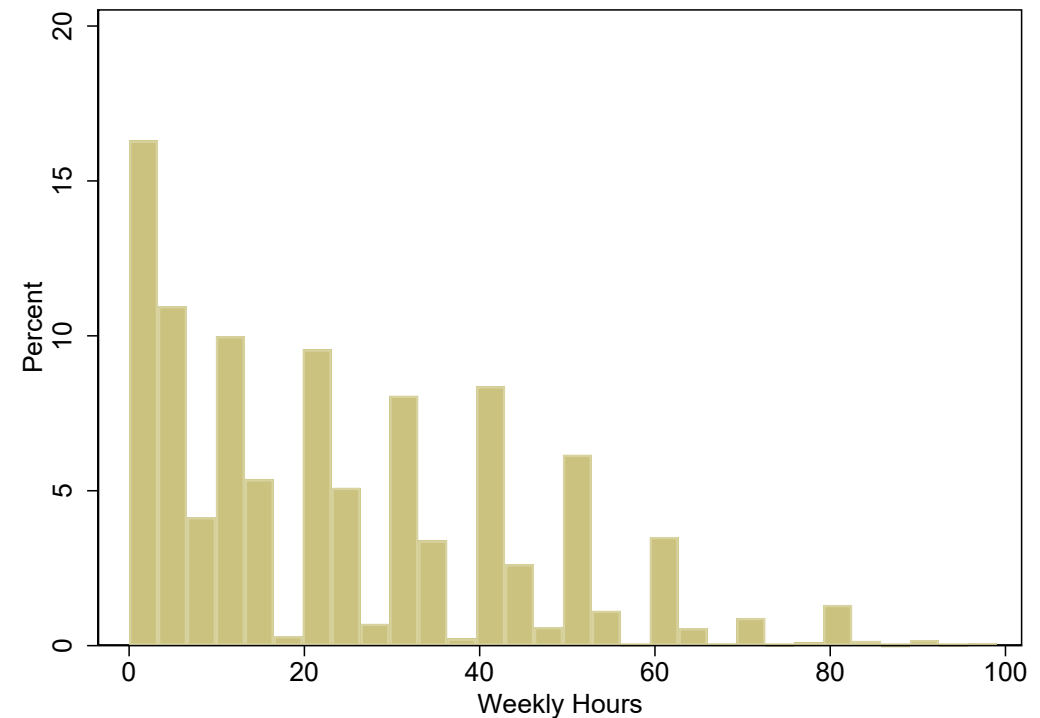
Graphing Entrepreneurial Activities by Gender
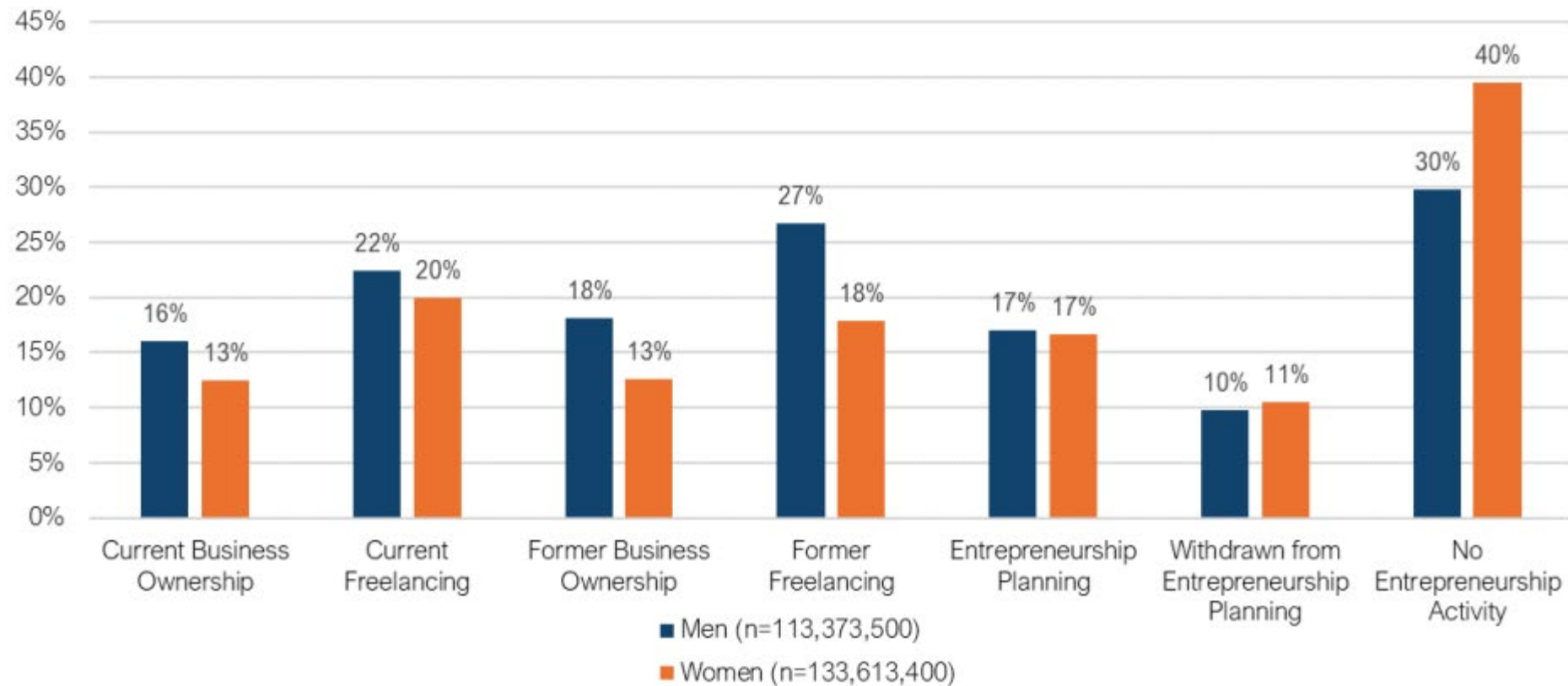
Source: NORC, Entrepreneurship in the Population Survey: 2022.

Data Access, Support, and Sharing

Accessing EPOP Data
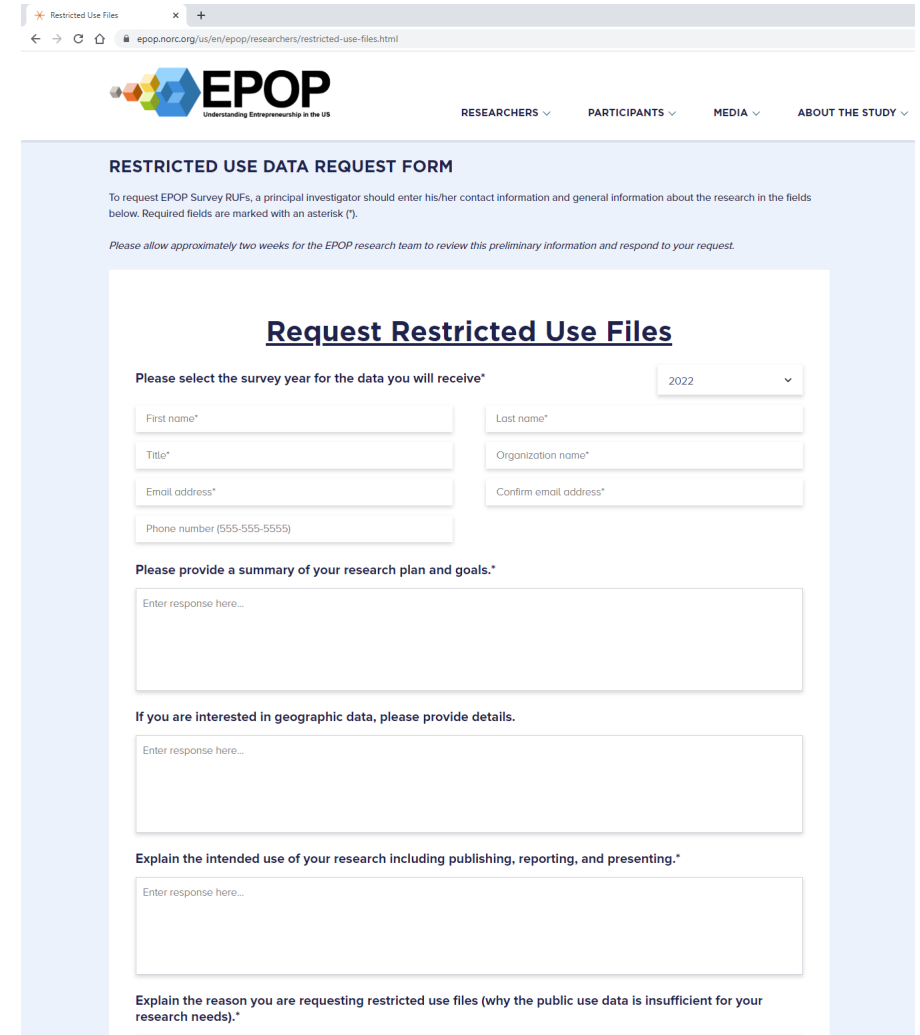
EPOP Survey Project Core Objective: Widely disseminate the data to researchers and practitioners

- Dissemination takes multiple forms to support the whole community
- Annual Data files discussed in today's session
  - Restricted use files available through an application process
  - Public use files immediately available
- Interactive online data
  - Topline statistics at the geographic level
  - Additional interactive tabulations planned
- Curating and sharing publications and findings

Accessing EPOP Data – Restricted Use File Data

Researchers request the EPOP:2022 RUF data via an application process

- The RUF has no direct PII, but does have variables if combined might be potentially identifying

- Initial RUF application form is online: https://epop.norc.org/us/en/epop/researchers/restricted-use-files.html

- The full application requires the following:

  – Data user contact information, including a CV or resume

  – A detailed research plan

  – A detailed data protection plan

  – Online data confidentiality/security training

  – Review and agreement to the terms of the EPOP Data Use Agreement (DUA)

- RUF application review takes 2-3 weeks

- There is no cost for the RUF data

Accessing EPOP Data – Public Use File Data

The EPOP:2022 PUF data is available now

- The PUF has no direct PII and has undergone disclosure review

- The online PUF request form:
  https://epop.norc.org/us/en/epop/researchers/public-use-files.html

- The PUF data:
  - Contains the full EPOP:2022 case set – 32,021 records
  - Includes the survey weight
  - Is appropriate for national and regional estimates
  - Should be used to formulate research plans prior to a RUF application

- PUF request and delivery via email is immediate

- There is no cost for the PUF data

Accessing EPOP Data – Online Interactive Data

The EPOP:2022 data is available in an online interactive application

- The data are shown by state and the 50 most populous MSAs

- The EPOP:2022 online interactive data: https://epop.norc.org/us/en/epop/researchers/interactive-data.html

- There are four (4) specific interactive dashboards:
  - Entrepreneurial Activities by State
  - Entrepreneurial Activities by MSA
  - Entrepreneurial Activities and Demographic Characteristics within a State
  - Entrepreneurial Activities and Demographic Characteristics within an MSA

- Dashboards will be updated annually

- Additional dashboards are planned – stay tuned
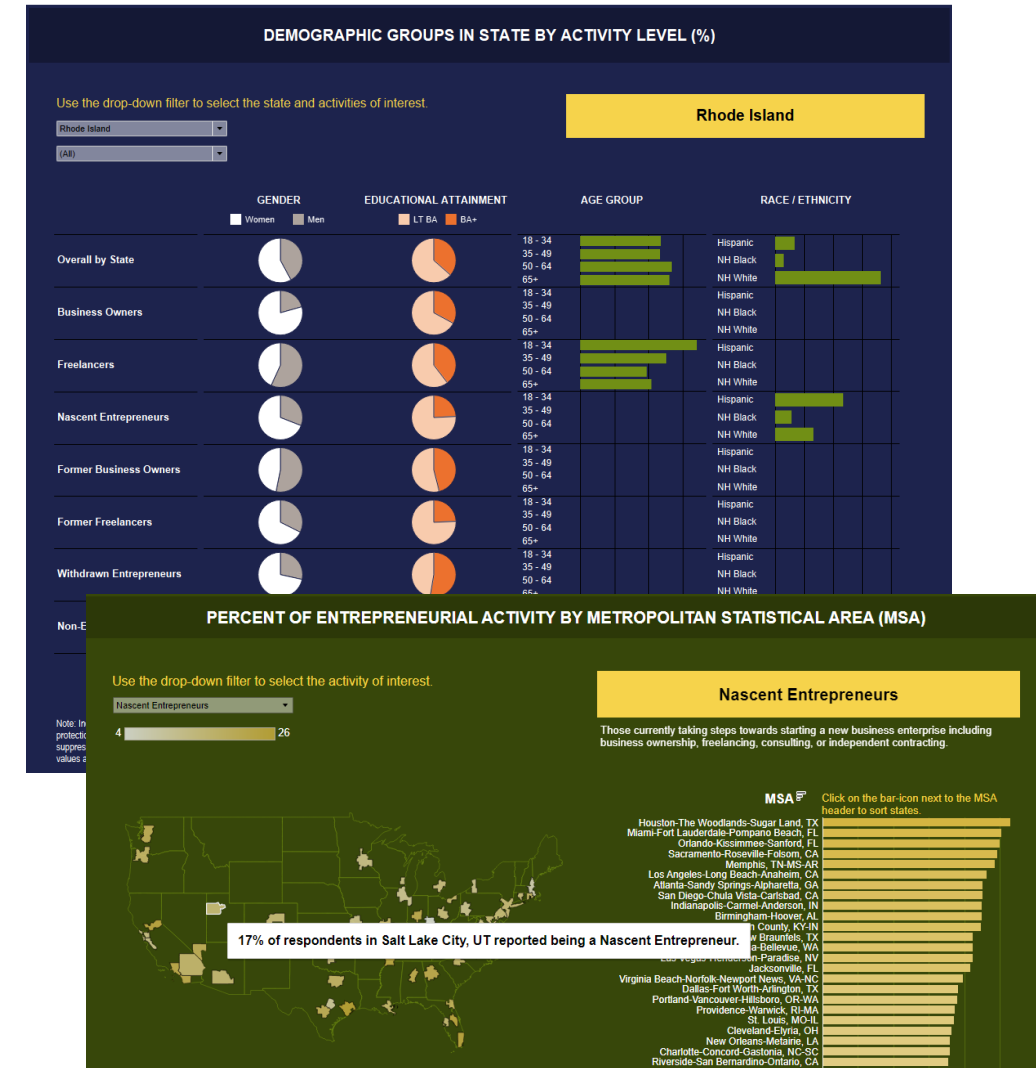
Accessing EPOP Data – Online Interactive Data Custom Comparisons

The view can be customized for comparisons

- Select a location – state or MSA

- Select the entrepreneurial activities of interest for comparison
  - EXAMPLE: Overall compared to Business Owners compared to Nascent Entrepreneurs (or Business Planners)

- Compare across major demographic groups:
  – Gender
  – Age group
  – Race/ethnicity

- Data supporting the dashboard is available in tabular form
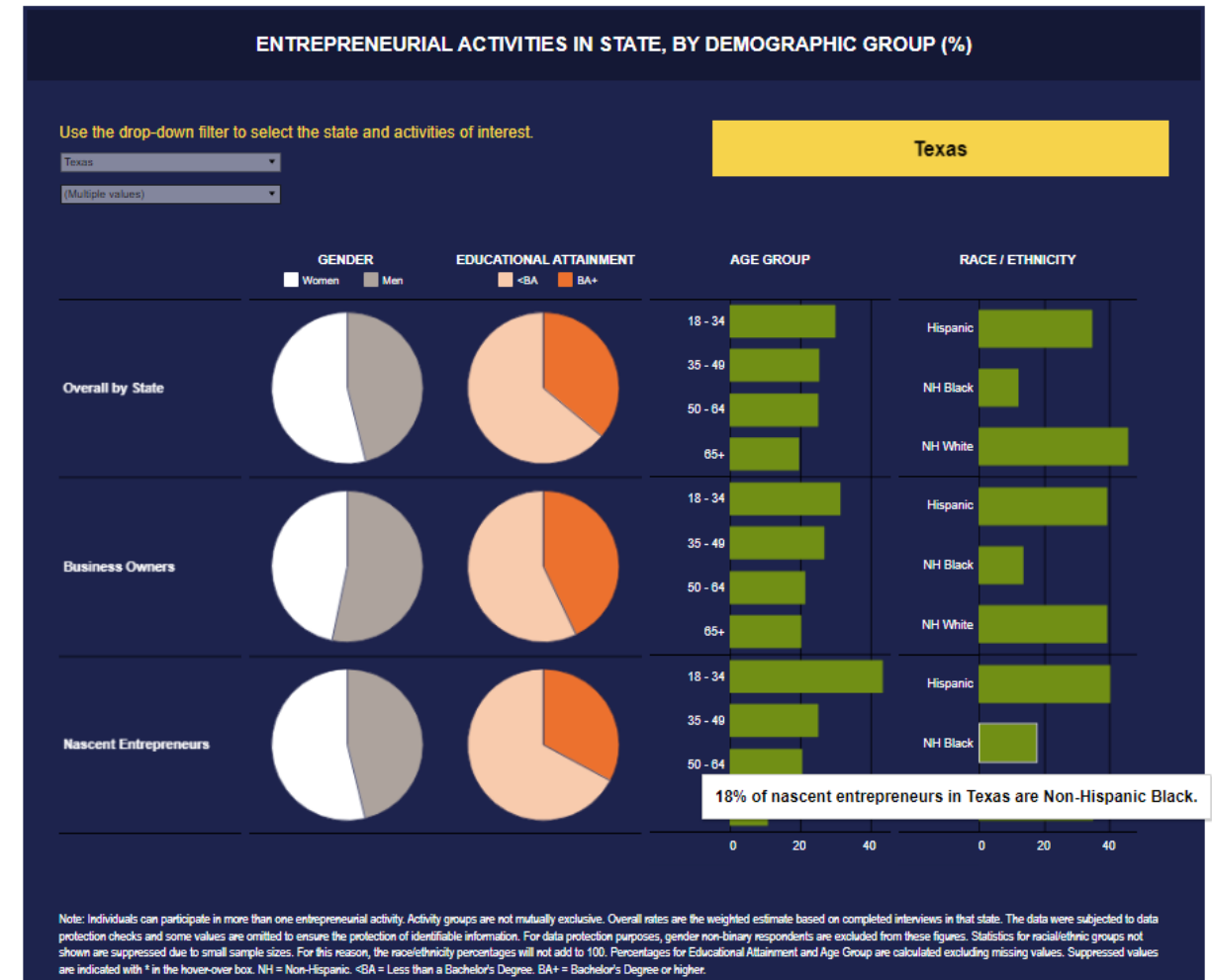
Small Area Estimation

Small Area Estimation is ongoing for Year 1, and estimates will be available for the following:

- All entrepreneurial activity categories and gig work shown in the online interactive data
  - By State
  - By MSA

- Gender and Racial/Ethnic breakdowns of each group of respondents above

Small Area Estimation

- Small Area Estimates "borrow strength" from higher levels of geography

- Models also include information from external sources

  – Examples: American Business Survey, American Community Survey, Business Dynamic Statistics, Non-employer Statistics, and Quarterly Workforce Indicators (QWI)

- As the sample size of an EPOP estimate increases, then the small area estimate will reflect the EPOP estimate and not the model

Small Area Estimation – Example
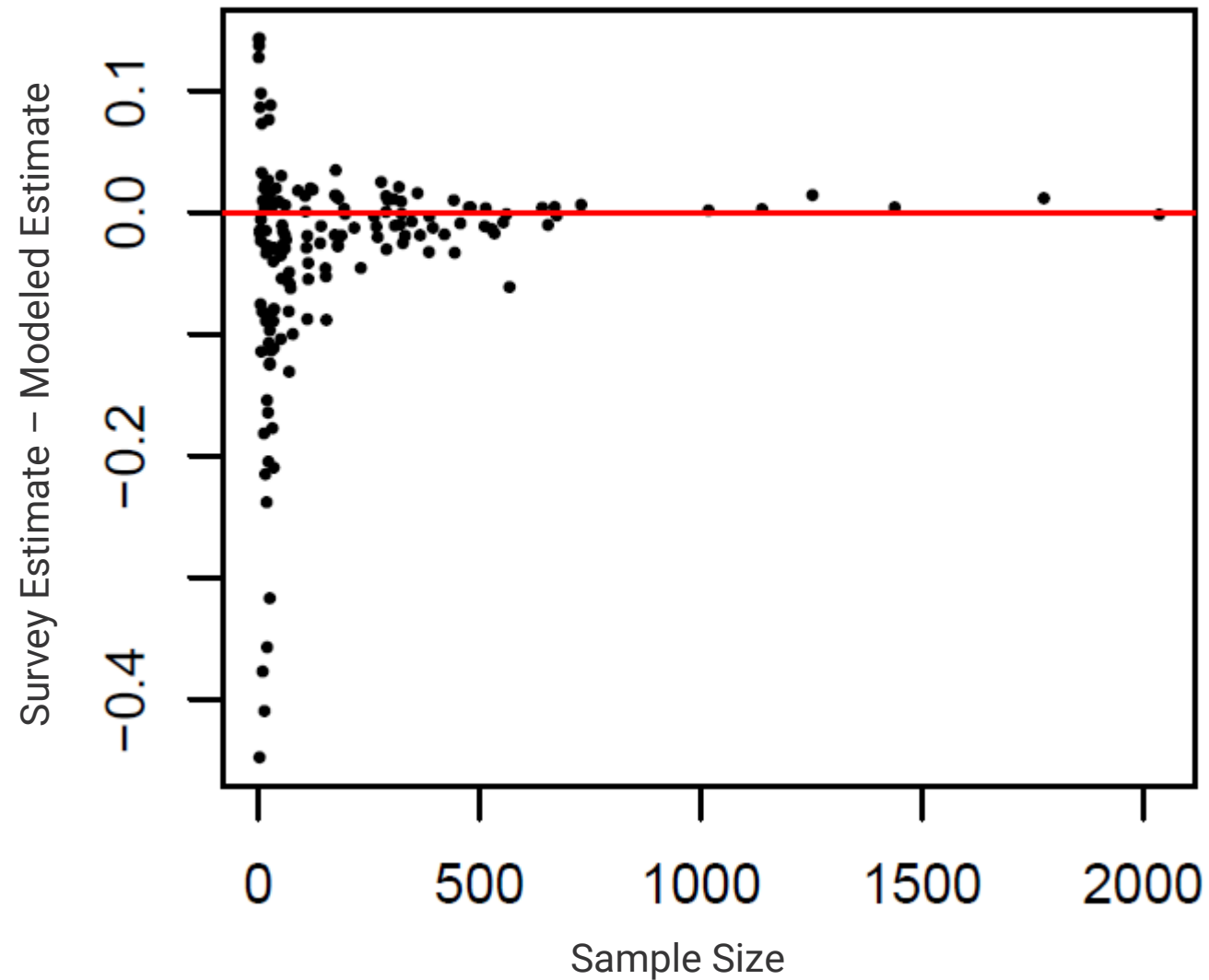
Accessing EPOP Data – Publications, findings, and survey

The EPOP Survey Project Website

- EPOP Survey data publication and presentations will be available or linked on the project website: https://epop.norc.org/us/en/epop/media/publications.html

- This becomes a resource for the community:
  - To share their findings
  - To learn about findings of others

- The EPOP:2022 questionnaire is also available for use: https://epop.norc.org/content/dam/epop/researchers/pdf/epop-2022-questionnaire.pdf
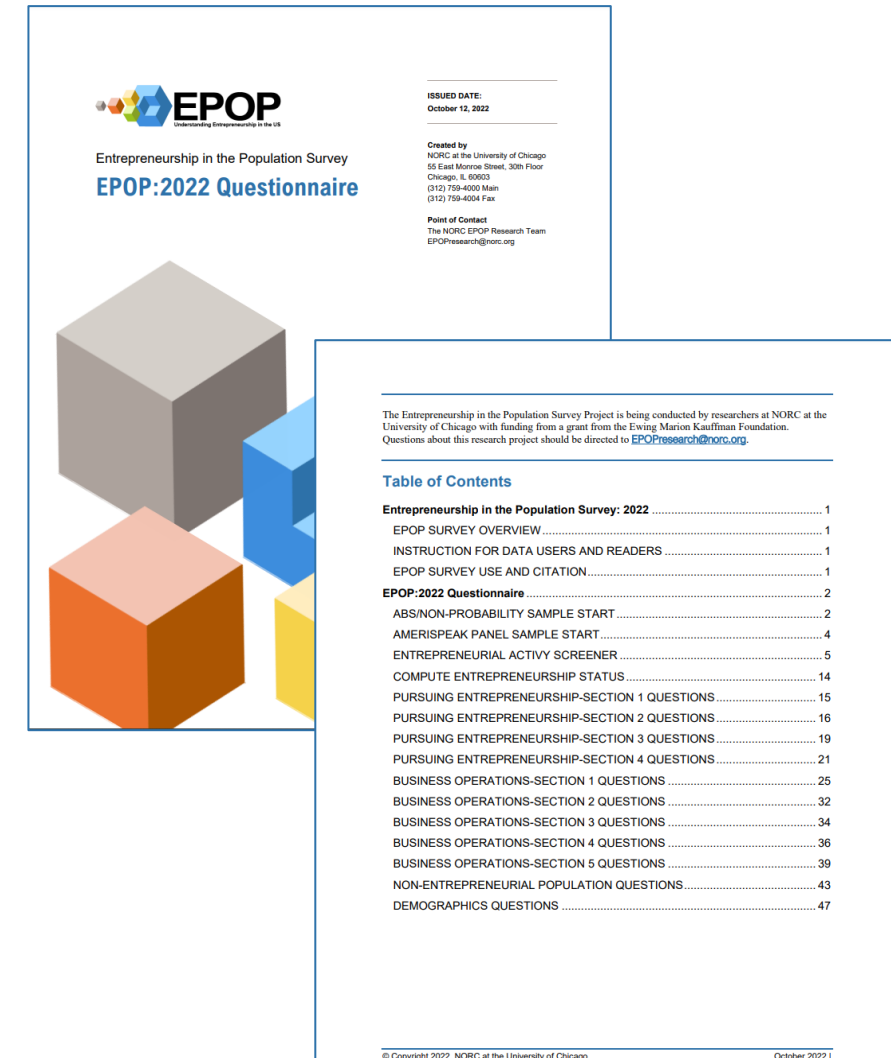
Getting Support

To get support or ask questions about the EPOP Survey data files:

- Send an email message to **EPOPresearch@norc.org**

- Provide details about your request or question

- Should you need follow-up support and would like a call back be sure to include your phone number and when to reach you

- The EPOPresearch Inbox is monitored during standard business hours M-F and you can expect a response within a day

Future EPOP Survey activities are underway to build on the current data release

- Small Area Estimates for each U.S. state and the top 50 Metropolitan Statistical Areas – *Available Soon!*

- Currently in data collection for "EPOP-IN," a special survey to measure the entrepreneurial activities of Indiana residents

- Four more years of data collection, allowing for the ability to track changes over time in entrepreneurial attitudes and behaviors
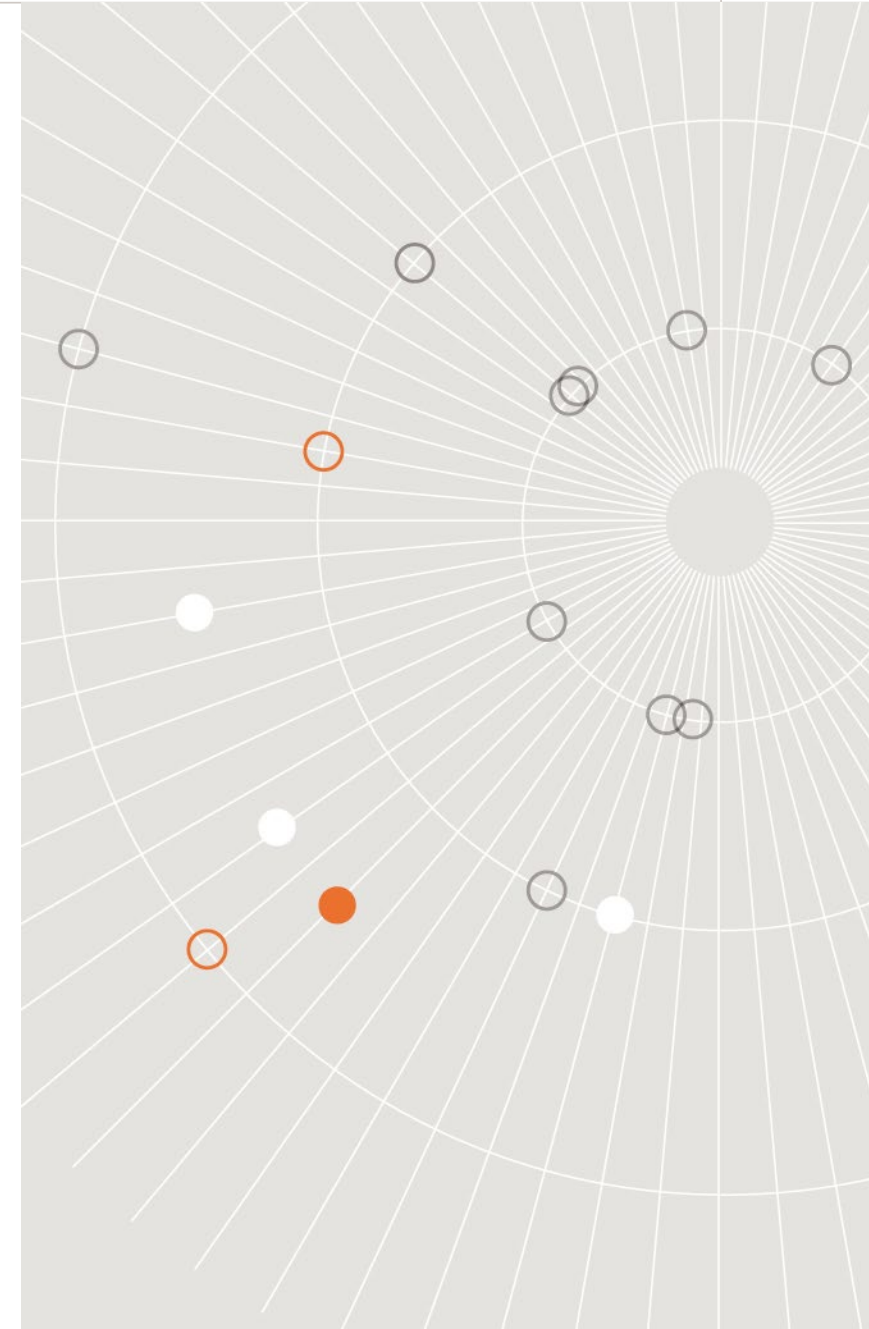
Planned Changes for Year 2

Changes to better capture "serial entrepreneurship" and "portfolio entrepreneurship"

Additions to capture the most important challenges facing entrepreneurs

A small set of topical questions will change each cycle

- Year 1 focused on the pandemic

- Year 2 items are under discussion
  - One consideration: More detailed questions on the relationship between current entrepreneurial activities and past employment
  - Another consideration: Impact of inflation
  - We're open to ideas!
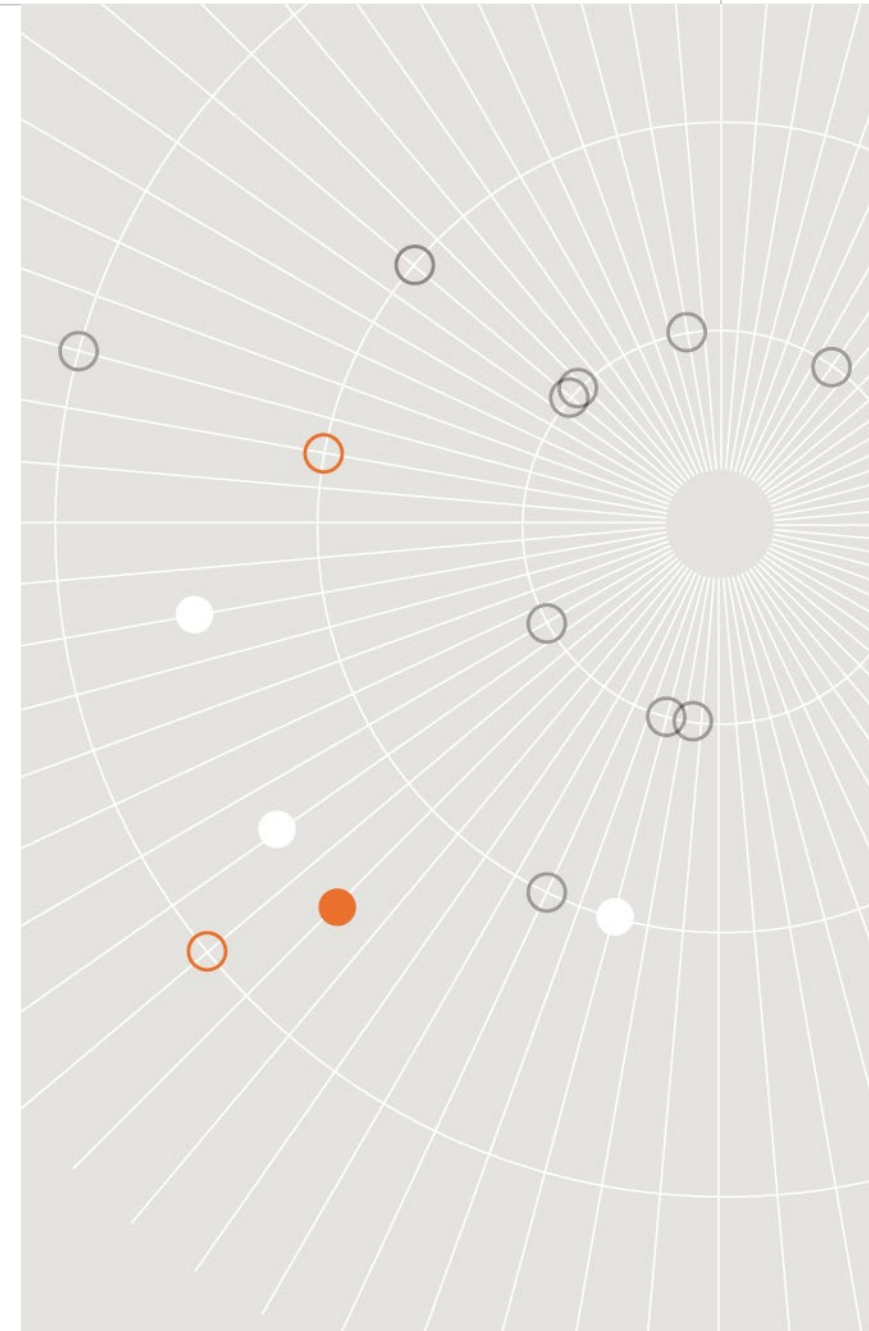
Release Schedule for Future Data Releases

Data Release	Anticipated Release Period
2022	October 2022
2023	August – September 2023
2024	August – September 2024
2025	August – September 2025
2026	August – September 2026
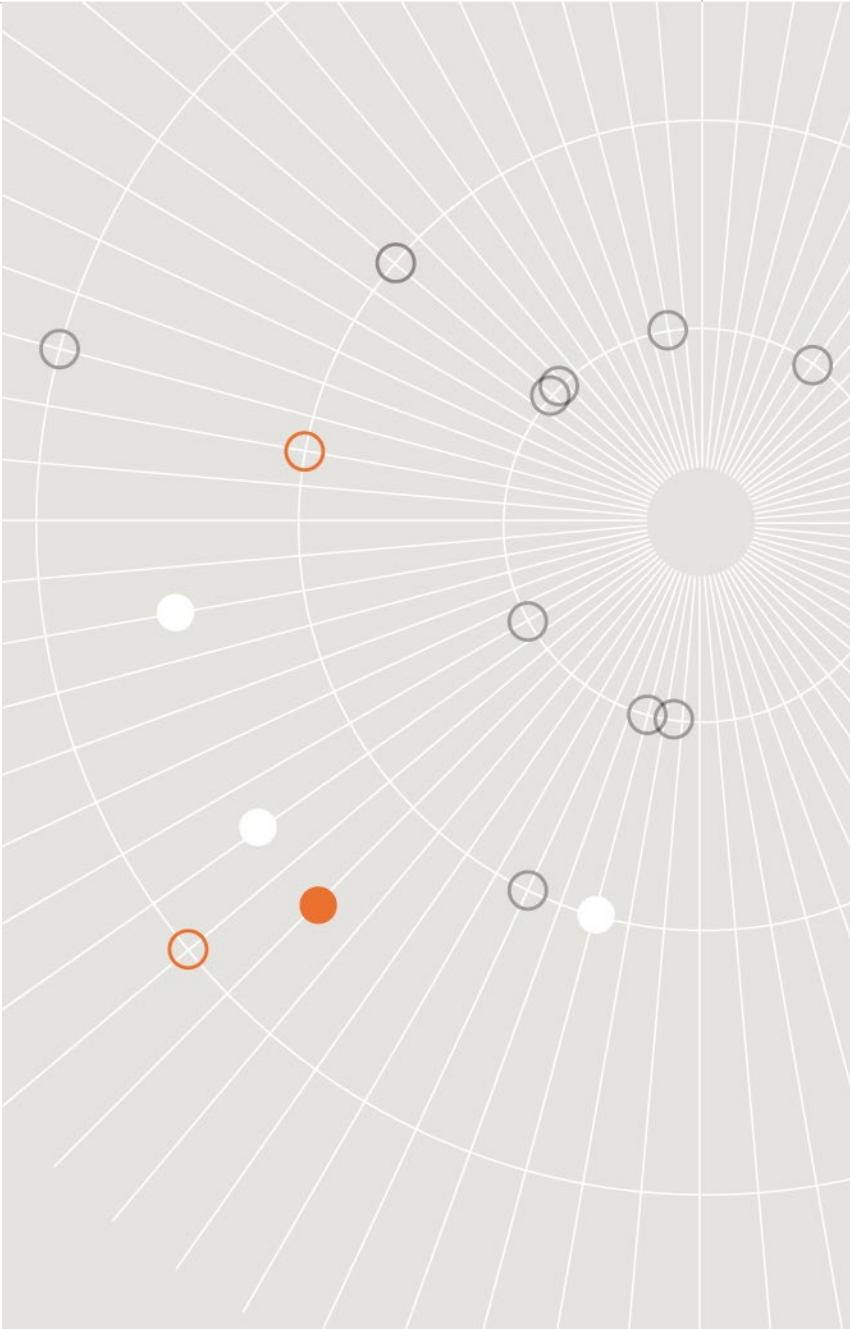
We want to hear from you and grow the community of EPOP Survey data users!

- If you need data use support, let us know

- Share your findings, reports, and presentations that use EPOP Survey data

- Send us your ideas for topical survey modules or survey improvements

- Let us know if you decide to run your own EPOP survey

- EPOPresearch@norc.org

General Questions…
and Answers

EPOP

EPOP

Thank you.

Quentin Brummet
EPOP Principal Investigator
Sr. Research Methodologist
Brummet-Quentin@norc.org

Karen Grigorian
EPOP Project Director
Vice President
Grigorian-Karen@norc.org

CITATION: Brummet, Q., K. Johnson, K. Grigorian, and Y. M. Yang. (2022). Understanding and Using the Entrepreneurship Survey Data. Pre-conference workshop presented at Association for Public Policy Analysis and Management 2022 Conference, November 16, 2022, Washington, DC.

✳ Research You Can Trust™

✳ **NORC** at the University of Chicago