

SURVEY FUTURES

SURVEY DATA COLLECTION
METHODS COLLABORATION

Methods for respondent-driven sampling with probability-based seeds: Lessons learnt from the NatCen panel feasibility study

Luciano Perfetti Villa* • Olga Maslovskaya • Curtis Jessop • Carina Cornesse

2nd Survey Futures Early Career Researchers Conference • Southampton, UK • 5th February 2026



University of Essex



University of
Southampton



Economic
and Social
Research Council



Introduction: RDS in Survey Futures

- **Research Strand 1** of Survey Futures: Enhanced Sampling Frames and Procedures (led by Prof Paul Smith):
 - **Sub-project 3: Respondent-driven Sampling Methods.**
 - Evidence review produced summarising current knowledge and practice <https://surveyfutures.net/wp-content/uploads/2026/01/report-10-respondent-driven-sampling-evidence-review.pdf>
 - Survey Practice Guide is currently being developed
- **Phase 2 Project:** *Under-Represented Population Sub-Groups in Social Surveys: Methods for Respondent-Driven Sampling with Probability-Based Seeds*

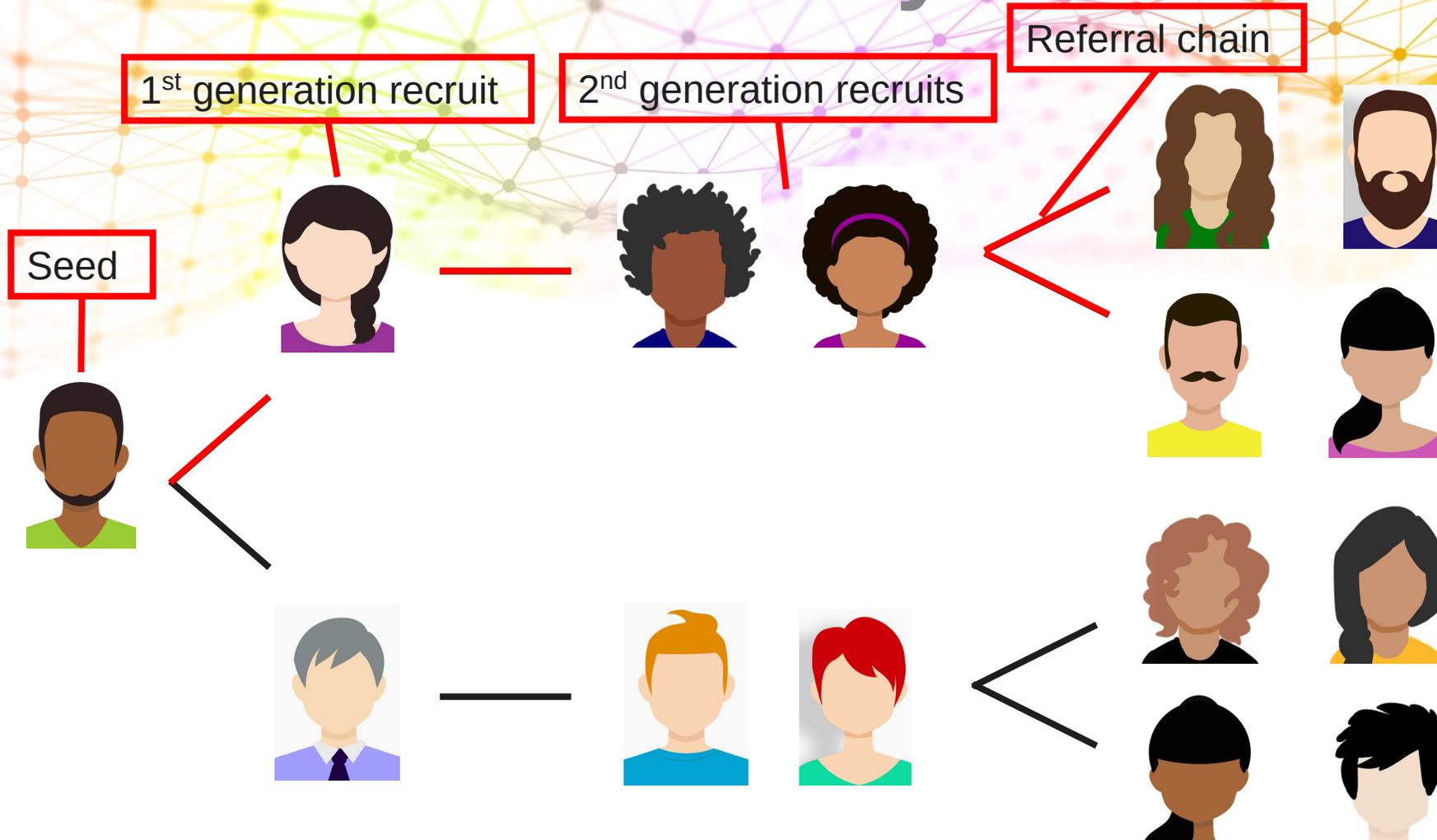
Introduction: Problem

- Some population sub-groups - consistently under-represented or excluded from survey samples or may appear in insufficient frequencies
- UK context: individual-level sampling frame is usually unavailable and over-sampling from the PAF or screening can be very expensive
- **BUT:** ensuring inclusivity for general population surveys is paramount for robust survey research
- Representativeness of social surveys is crucially important too
- RDS may be a suitable approach for addressing challenges related to sample boosting, inclusivity and representativeness

Introduction: RDS

- **Chain-referral sampling method** used to reach hidden or hard-to-reach groups when no sampling frame exists (Heckathorn, 1997); shown to be effective for recruiting harder-to-survey sub-groups (e.g. migrants, ethnic minorities, LGBTQ+ communities)
- Builds on **snowball sampling**, but with a more structured recruitment process supported by **mathematical modelling**, enabling more reliable population estimates.
- **Equilibrium assumption** refers to the point at which the sample composition becomes stable after several recruitment waves and is no longer influenced by the initial participants
- Relatively unexplored approach in the UK: exceptions are ONS pilot studies with specific groups (people living on boats, Gypsy and Traveller communities and York residents); cognitive interviewing was conducted by Welsh Government to test RDS questions.
- Evidence review and parallel study in Germany (led by Cornesse)

How does RDS usually work?



Source: Cornesse (2024)



Univ

Important: It is the participant who does the recruiting NOT the survey agency!

Advantages of online RDS

- Use of RDS to boost an under-represented groups using seeds from a general population survey sample has not yet been explored in the UK
- Fast turn-around, can help **boosting** sub-groups effectively
- Easier than non-online RDS – no need to handle offline coupons
- Possible to preserve **the link** between each panelist & their recruited network members (e.g. via personalised pre-connected survey links)
- If a number of assumptions are met, data approximate a probability-based sample without needing a sampling frame – ensures **inclusivity** and improves **representativeness**

Our Feasibility Study



- **Methodological goal:** to explore feasibility and evaluate the quality of RDS method
- **Purpose:** to employ RDS with seeds from a probability-based NatCen online panel to boost probability-based sample
- **Target population:** young people - 18-24 year-old
- **Seeds:** 18-24 year old and 25+
- **Dual incentive scheme:** incentive for participation and per each successful recruit
- **Recruits** continue recruitment with their own coupons created traceable “**waves**” of participants (up to 10 waves).
- **Questionnaire:** 15 minutes online survey with three modules: demographic questions, network and community questions, and substantive questions (to replicate a realistic survey experience)

Main Aims of Feasibility Study

1. Test the **feasibility** of the operational set up, including practical arrangements and systems
2. Examine the data generating and **recruitment process**
 1. Can RDS be used to boost a probability-based sample to large numbers?
 2. Are seeds willing and able to recruit successfully?
 3. Are recruits willing and able to participate successfully?
 4. What proportion of recruitment chains reach 4-6 waves, sufficient to achieve equilibrium?
3. Assess **representativeness** and test whether key RDS **assumptions** hold
 1. Representativeness
 2. Exponential growth
 3. Peer recruitment through existing social ties
 4. Random recruitment within personal networks
 5. Equilibrium

Protocol: Original

Fraud prevention: an important consideration!

- **Two-factor authentication:** to prevent duplication
- **UK mobile number requirement** (starting with “07”) to verify eligibility
- **Screening questions:** used to confirm eligibility (18-24 living in the UK); the survey was closed for ineligible respondents – *though age could potentially be misreported*
- **Recruitment was stopped** when:
 - Chain length reached 10 participants,
 - Speeding was detected (between the participant and parent),
 - Same email address was used again for incentive payment

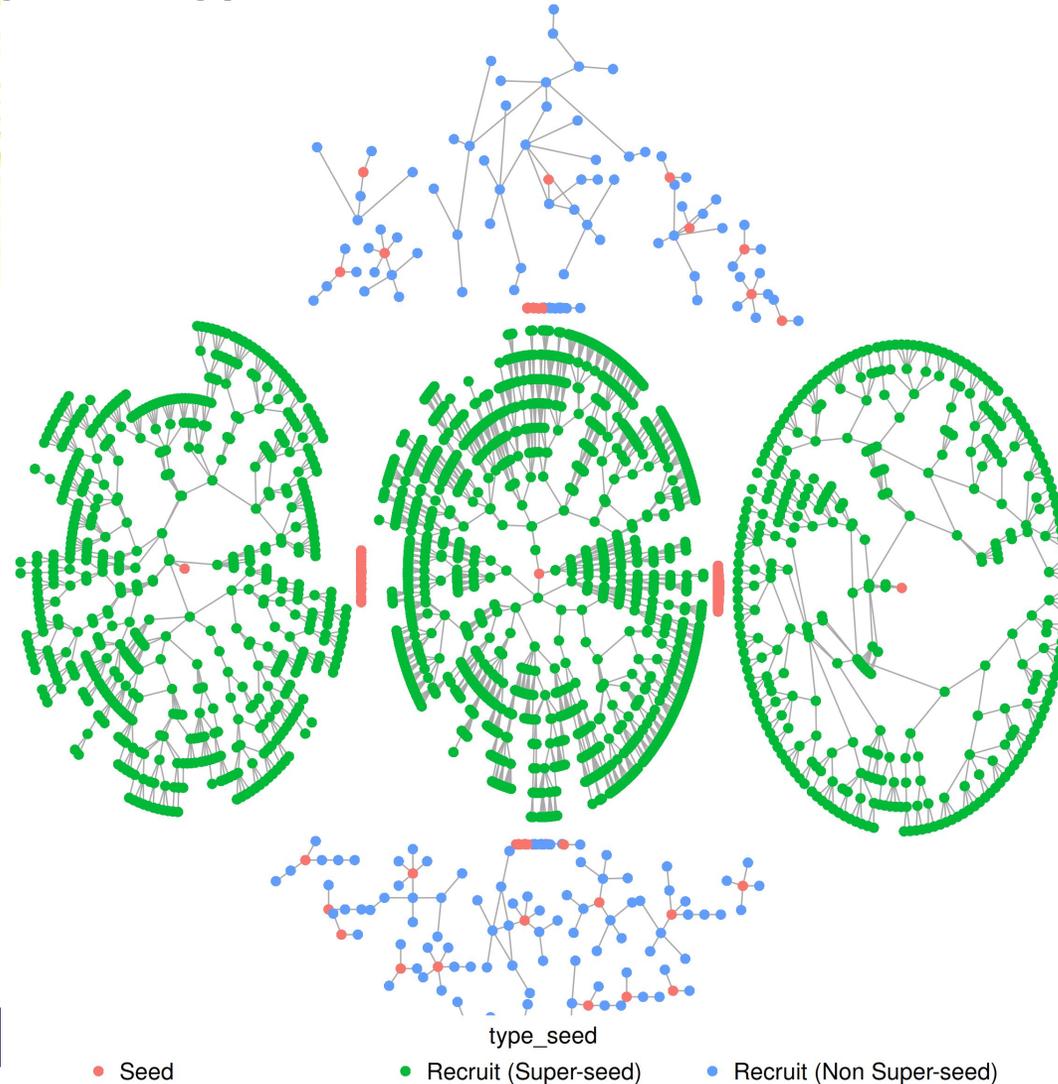
Protocol: Adaptations

Protocol was **adapted during fieldwork** in response to emerging issues (non-compliance and super seeds):

- **Additional age validation checks** were introduced a few days into fieldwork, using three reference events (Brexit referendum, Twin Towers attack and Covid-19 pandemic)
- **Non-compliance checks:** Participants who reported not personally knowing their recruiter were **not allowed to recruit others**
- **Actions taken for identified super seeds** (seeds with unusually large or influential social networks):
 - Chains were closed
 - Automated incentive payments were paused to implement additional checks
 - **Age validation follow-up:** all participants recruited by the three super seeds were asked to confirm their date of birth in a short survey (half of participants reported a date of birth consistent with the age they provided)

Results 2.1: Recruitment process

Recruitment process by chain type



Results 2.1 and 3.2: Data Collection

	Soft Launch	Mainstage
Dates	June 2025	July 2025
Dual Incentives	£5	£10
Codes	3	5
Expected sample size	3,000	1,500

- Achieved sample size: **2,229**
- Response rates (seeds): 39% (321 out of 820)
- **1.1** average recruit per participant
- 3 super seeds identified (for 1 fraud activity identified)
- All results are presented for full sample and then broken down by only super-seeds and without super-seeds

Results 2.1: Participants recruited by different types of seeds

2.1 Can we use RDS to boost a probability-based sample to large numbers?

Seed source	Number of RDS recruits	Percentage	All participants	Percentage
Super seed 1	375	16.8%	376	14.7%
Super seed 2	541	24.3%	542	21.2%
Super seed 3	1,131	50.7%	1132	44.4%
All other seeds	182	8.2%	499	19.7%
TOTAL	2,229		2,549	

Results 2.2 and 2.3: Willingness to invite others to participate

Willing to invite others	Total sample		Super seeds only		Non-super seeds only	
	Total	Percentage	Total	Percentage	Total	Percentage
Yes	1849	90.02%	1523	97.32%	329	66.87%
No	153	7.45%	28	1.79%	125	25.41%
I don't know anyone 18-24	52	2.53%	14	0.89%	38	7.72%
Total	2054	100.00%	1565	100.00%	492	100.00%

- Interestingly, seeds 25+ had higher willingness to invite those 18-24 than those from within the target group (99% against 93%)

Note: Excluded observations with values such as "Don't know/Not applicable/Refused" (n=492).

Results 2.2 and 2.3: Willingness to participate

Reasons to invite others to participate

	Total Sample		Super Seeds		Non Super Seeds		Seeds Only	
	Total	Percentage	Total	Percentage	Total	Percentage	Total	Percentage
Receiving a voucher for myself	208	11.82%	155	10.62%	55	18.15%	30	17.14%
So that others can receive a voucher	249	14.15%	167	11.44%	82	27.06%	46	26.29%
To support scientific research	748	42.50%	669	45.82%	80	26.40%	53	30.29%
Others' curiosity to try new things	113	6.42%	90	6.16%	23	7.59%	10	5.71%
I think they will enjoy completing the survey	430	24.43%	374	25.62%	56	18.48%	31	17.71%
Other (Please specify)	12	0.68%	5	0.34%	7	2.31%	5	2.86%
Total	1760	100.00%	1460	100.00%	303	100.00%	175	100.00%

Results 2.2 and 2.3: Reasons not to invite others to participate

Reasons not to invite others to participate

Reasons not to invite others	Total Sample		Super Seeds		Non Super Seeds		Seeds Only	
	Total	Percentage	Total	Percentage	Total	Percentage	Total	Percentage
I do not feel comfortable inviting others to take part	86	58.11%	7	25.93%	79	65.29%	61	64.89%
I do not see the point	4	2.70%	2	7.41%	2	1.65%	2	2.13%
The effort is too much for me	14	9.46%	5	18.52%	9	7.44%	5	5.32%
The effort will be too much for those I would invite	32	21.62%	13	48.15%	19	15.70%	17	18.09%
Other (Please describe)	12	8.11%			12	9.92%	9	9.57%
Total	148	100.00%	27	100.00%	121	100.00%	94	100.00%

Results 2.2: Recruitment process – Seeds

Are seeds willing and able to recruit successfully?

Outcome	Total Sample		Super Seeds		Non Super Seeds	
	Total	Percentage	Total	Percentage	Total	Percentage
Total seeds	321	100.00%	3	100.00%	318	100.00%
Consented to RDS	186	57.94%	3	100.00%	183	57.55%
Recruited at least one participant	185	57.63%	3	100.00%	182	57.23%
At least one recruit started survey	55	17.13%	3	100.00%	52	16.35%
At least one recruit completed survey	49	15.26%	3	100.00%	46	14.47%

Results 2.3: Recruitment process - Recruits

Are recruits willing and able participate successfully?

Outcome	Total Sample		Super Seeds		Non Super Seeds	
	Total	Percentage	Total	Percentage	Total	Percentage
Total coupons	9205		7560		1645	
Started survey	2716	100.00%	2507	100.00%	209	100.00%
Completed survey	2229	82.07%	2047	81.65%	182	87.08%
Willing to recruit others	1666	61.34%	1520	60.63%	146	69.86%
Average coupons per seed	49.49		2520		8.99	
Average started recruits per seed	14.6		835.67		1.14	
Average completed recruits per seed	11.98		682.33		0.99	
Average willing to continue recruiting per seed	8.96		506.67		0.80	

Results 2.4 and 3.2: Recruitment Process - Waves



	Seeds	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
Potential invites	N/A	889	404	265	320	575	860	1105	1425	1620	1735
Completed main survey	317	107	65	70	119	183	232	297	341	371	445
Completion rate	N/A	12%	16%	26%	37%	32%	27%	27%	24%	23%	26%
Potential recruiters	315	106	61	68	118	177	228	291	329	362	N/A
Agreed to recruit	183	84	55	64	115	172	221	285	324	347	N/A
Rate of agreement	58%	79%	90%	94%	97%	97%	97%	98%	98%	96%	N/A
Successful recruits per participant	0.3	0.6	1.1	1.7	1.5	1.3	1.3	1.1	1.1	1.2	N/A
Successful recruits per participant (super seeds)	3.7	1.6	2.9	2.1	1.6	1.3	1.3	1.2	1.1	1.2	N/A
Successful recruits per participant (non super seeds)	0.3	0.5	0.4	0.4	0.6	1.2	0.5	N/A	N/A	N/A	N/A

Results 2.4: Recruitment Process – Chains (Seeds)

Wave	Chains alive	Percentage
1	49	100
2	21	43
3	10	20
4	4	8
5	4	8
6	4	8
7	4	8
8	3	6
9	3	6
10	3	6

What proportion of recruitment chains reached 4-6 waves sufficient to achieve equilibrium?

- 49 productive chains: overall average chain length is 2.14 with maximum of 10 waves; without super seeds: average length is 1.63 and maximum of 7 waves

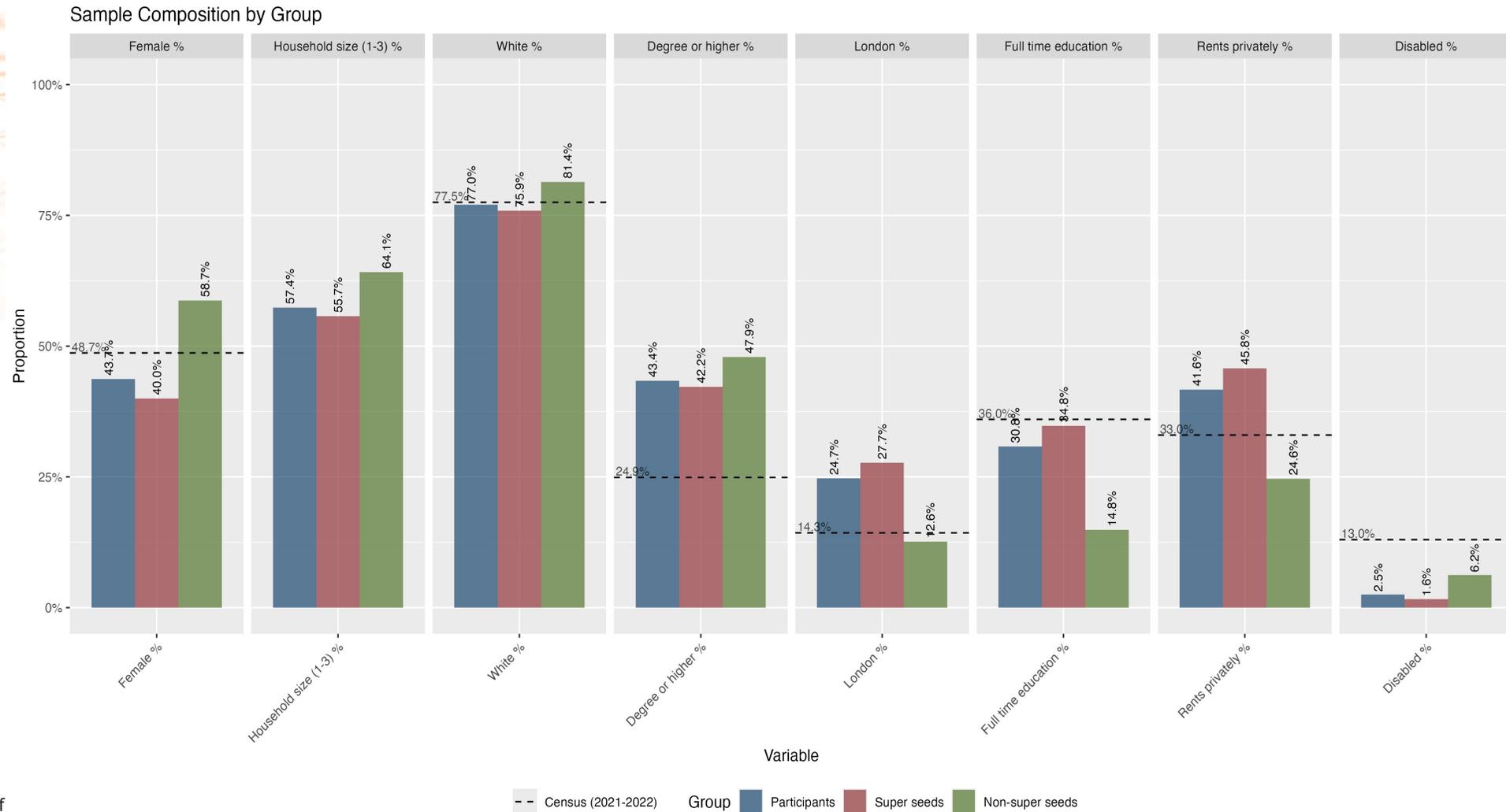
Results 2.4: Recruitment Process – Branches

Branch Length	Total Sample		Super Seeds		Non Super Seeds	
	Branches	%	Branches	%	Branches	%
1	76	5.4%	6	0.5%	70	50.7%
2	42	3.0%	4	0.3%	38	27.5%
3	29	2.1%	14	1.1%	15	10.9%
4	54	3.9%	48	3.8%	6	4.3%
5	97	6.9%	95	7.5%	2	1.4%
6	119	8.5%	115	9.1%	4	2.9%
7	150	10.7%	147	11.7%	3	2.2%
8	185	13.2%	185	14.7%	0	0.0%
9	202	14.4%	202	16.0%	0	0.0%
10	445	31.8%	445	35.3%	0	0.0%
Total	1,399		1,261		138	

What percentage of chains reached 4-6 waves sufficient for reaching equilibrium?

- Branches are starting from the seed and ending in a leaf when no further recruitment occurs – it is one recruitment pathway in the network. Multiple branches could pass through one participant.
- 89% of all successful branches reached at least 4 waves (98% from super seeds sub-sample but only 11% from non super seeds sub-sample)

Results 3.1: Representativeness



Dashed line indicates Census (2021-2022) proportions; Female % obtained from ONS mid-year population estimates 2024

Results 3.1: Representativeness

Distribution of the sample by Ethnicity and type of chain membership

	Total Sample		Super Seeds		Non Super Seeds		Statistical Sign.	Benchmarks		
	Total	Percentage	Total	Percentage	Total	Percentage	P Value	Census 2021	Natcen Panel (weighted)	Natcen Panel (unweighted)
White British (English/Welsh/Scottish/Northern Irish)	1846	72.85%	1472	72.09%	376	75.96%	0.093	65.97%	71.15%	68.55%
Any other White background	114	4.50%	84	4.11%	30	6.06%	0.079	7.64%	6.35%	7.26%
Mixed or multiple ethnic groups	88	3.47%	75	3.67%	13	2.63%	0.315	6.41%	4.06%	5.65%
Asian or Asian British	216	8.52%	158	7.74%	58	11.72%	0.006	17.56%	11.21%	15.32%
Black or Black British	241	9.51%	229	11.21%	13	2.63%	<0.001	1.54%	4.85%	2.42%
Other	29	1.14%	24	1.18%	5	1.01%	0.941	0.87%	2.38%	0.81%
Total	2534	100.00%	2042	100.00%	495	100.00%		100%	100%	100%

Note: Excluded observations with values such as "Don't know/Not applicable/Refused" (n=12).

Significance values measured as comparison between super seed and non super seeds proportions per category.

Results 3.3: Peer recruitment occurs through existing social ties

What is your relationship to the person who invited you to this survey?

	Total Sample		Super Seeds		Non Super Seeds	
	Total	Percentage	Total	Percentage	Total	Percentage
Friend	1542	69.21%	1429	69.84%	113	62.09%
Colleague	284	12.75%	277	13.54%	7	3.85%
Relative/family member	198	8.89%	158	7.72%	40	21.98%
I do not know the person who invited me	96	4.31%	96	4.69%	0	0.00%
Partner	81	3.64%	61	2.98%	20	10.99%
Other	27	1.21%	25	1.22%	2	1.10%
Total	2228	100.00%	2046	100.00%	182	100.00%

Note: Excluded observations with values such as "Don't know/Not applicable/Refused" (n=1).

Significance values measured as comparison between super seed and non super seeds proportions per category.

Results 3.4: Random recruitment within personal networks

Recruitment diagnostics: intra-chain correlation

Group	Variable	ICC	ICC Lower CI	ICC Upper CI	Frequency
Ethnicity	White	0.78	0.59	0.97	1634
	Black	0.32	0.06	0.61	225
	Asian	0.96	0.97	1.00	175
	Other	0.35	0.06	0.63	109
	Aggregate	0.66	0.55	0.80	
Sex	Female	0.13	0.03	0.20	918
	Male	0.13	0.02	0.21	1225
	Aggregate	0.13	0.05	0.18	

Group	Variable	ICC	ICC Lower CI	ICC Upper CI	Frequency
Education	Degree	0.39	0.19	0.47	924
	A-Level	0.00	0.00	0.00	340
	Other/No Qualification	0.38	0.05	0.78	15
	Aggregate	0.25	0.20	0.47	
Disability	With Disability	0.23	0.03	0.38	156
	Without Disability	0.24	0.05	0.40	1944
	Aggregate	0.24	0.11	0.32	

Note: Excluded observations with values such as “Don’t know/Not applicable/Refused”.

- Aggregate ICC is estimated as the weighted average of each category’s ICC and weighted by the percentage of the total sum of the variances (Breen and Ermisch, 2021). Lower and upper bounds of the confidence intervals are estimated using a parametric bootstrap.

Results 3.4: Random recruitment within personal networks

Recruitment diagnostics: intra-chain correlation

Variable	ICC	ICC Lower CI	ICC Upper CI	Frequency
London	0.91	0.92	0.99	567
South East	0.94	0.95	0.99	217
North West	0.93	0.93	0.99	305
Yorkshire and The Humber	0.88	0.86	0.98	159
East Midlands	0.78	0.66	0.97	143
West Midlands	0.88	0.93	0.99	188
South West	0.98	0.98	1.00	118
East of England	0.57	0.26	0.91	117
North East	0.96	0.97	1.00	55
Wales	0.00	0.00	0.00	57
Scotland	0.93	0.94	0.99	125
Northern Ireland	0.03	0.00	0.27	35
Aggregate	0.84	0.89	0.95	

Note: Excluded observations with values such as "Don't know/Not applicable/Refused".

- Aggregate ICC is estimated as the weighted average of each category's ICC and weighted by the percentage of the total sum of the variances (Breen and Ermisch, 2021). Lower and upper bounds of the confidence intervals are estimated using a parametric bootstrap.

Summary of Results

1.1 Operational set up, practical arrangements and systems	YES
2.1 Can RDS be used to boost a probability-based sample to large numbers?	YES
2.2 Are seeds willing and able to recruit successfully?	YES
2.3 Are recruits willing and able to participate successfully?	YES
2.4 What proportion of chains reached 4-6 waves?	Seed chains - 8%, branches – 11% (non super seeds)
3.1 Representativeness	Closer to pop estimates for some characteristics
3.2 Exponential growth	YES when super seeds are included
3.3 Peer recruitment occurs through existing social ties	Only 4.3% reported not knowing the person who recruited them
3.4 Random recruitment within personal networks	YES diversity in recruitment for some characteristics but not for ethnicity or region
3.5 Equilibrium	YES for some characteristics but more work needed

Conclusions

- This was the first and successful feasibility study, demonstrating the potential of RDS to boost general population surveys.
- Making RDS work is not a trivial task; further development work is needed to address major issues such as fraud before it can be recommended as a reliable technique for boosting probability-based samples.
- It is broadly possible to boost probability-based sample BUT can the study be replicated with the same protocol? Further research on validity of protocols is needed.
- Representativeness: early results are reassuring but more work is required.
- Analysis of the overall sample shows more positive results.
- Many studies do not include separation, but it is important to be cautious and acknowledge data generating processes.
- Further work on super seeds is needed.

Recommendations

- Use responsive design and remain flexible.
- Different sub-groups may respond differently to the same protocol so piloting and testing are crucial!
- Maintain real-time monitoring of data collection, including during weekends!

Key learnings for replication of our study:

- Implement stricter protocol against fraud from the start, including:
 - **Non-compliance checks** (e.g., participants who report not knowing their recruiter)
 - Stricter **age validation checks** from the onset
- Introduce gradual changes to incentives and the number of coupons each participant can use
- Conduct real-time monitoring of recruits per respondent to prevent super seeds from dominating recruitment; controlling only the number of waves is not sufficient

Documents and Event

- Survey Futures Survey Practice Workshop “*Respondent-Driven Sampling: Advantages, Limitations, and Practical Considerations for Recruiting Hard-to-Survey Populations*” will take place **online on Wednesday the 11th of February 2026 from 13.00 to 16.30 UK time**. Please register here: <https://www.eventbrite.co.uk/e/survey-futures-survey-practice-workshop-tickets-1979986597212?aff=oddtcreator>
- RDS evidence review <https://surveyfutures.net/wp-content/uploads/2026/01/report-10-respondent-driven-sampling-evidence-review.pdf>
- NatCen Panel RDS Feasibility Study Technical Report <https://surveyfutures.net/reports/>
- RDS Survey Practice Guide – *in progress*

References

- Breen, R. and Ermisch, J.** (2021). Sibling Models, Categorical Outcomes, and the Intra-Class Correlation, *European Sociological Review*, 37(3), pp. 497–504. Available at: <https://doi.org/10.1093/esr/jcaa057>.
- Cornesse, C.** (2024). Surveying Panel Participants' Network Members: A New Data Spaces for the Social Sciences? Presentation.
- Heckathorn, D.D.** (1997). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems*, 44(2), pp.174–199.

SURVEY FUTURES

SURVEY DATA COLLECTION
METHODS COLLABORATION

Thank you!

Luciano Perfetti Villa* • Olga Maslovskaya • Curtis Jessop • Carina Cornesse

2nd Survey Futures Early Career Researchers Conference • Southampton, UK • 5th February 2026

