Methodological Assurance Review Panel

Census 2031 Key Population Groups: Identification framework and prioritisation method

Purpose

1. This paper outlines the framework to identify key population groups and the prioritisation method for Census 2031.

Timing

2. Work is at an early, but substantive stage of development. We are seeking input from MARP on the approach taken.
3. We are aiming for an approved method and list of key population groups by Autumn 2026 to support other deliverables across the census programme.

Request for advice and assurance

4. Members of the board are invited to:
    i. Advise on whether the overall framework is conceptually sound and appropriate.
    ii. Identify any significant methodological risks, gaps or assumptions.

Executive Summary

2. The census is a survey of all the people and households across England and Wales, run every ten years by the Office for National Statistics (ONS). The next census takes place in 2031. It provides a detailed picture of the population, informing decisions about funding, public services, transport, healthcare, education, and local planning.
3. A core objective of the 2031 Census Strategy is to ensure that: “everyone in England and Wales can take part, using completion routes that are respondent-centred, reliable and accessible.” (ONS, 2025)
4. Achieving this requires a clear, evidence-based approach to understand where the standard Census approaches are at most risk of being less effective and where adaptations to the standard approach are required.
5. This paper sets out a single, consistent approach to identifying and prioritising key population groups (KPGs). KPGs face barriers to participation and standard census approaches are less effective without adaptation.
6. The KPG framework builds on the successful Census 2021 approach, which achieved an overall response rate of 97%. For Census 2031 we are introducing a programme level list of KPG. Our approach is to be holistic, provide structure and transparency to:
    i. identify groups with historically low response rates and data quality
    ii. integrate a range of quantitative and qualitative data sources
    iii. prioritise where targeted interventions are most needed
    iv. support earlier and coordinated planning across the census programme

Background

Census non-response and its consequences

7. While overall Census 2021 response rates were high, some groups remained under-represented. For example, response rate in students' halls of residence were 78%.

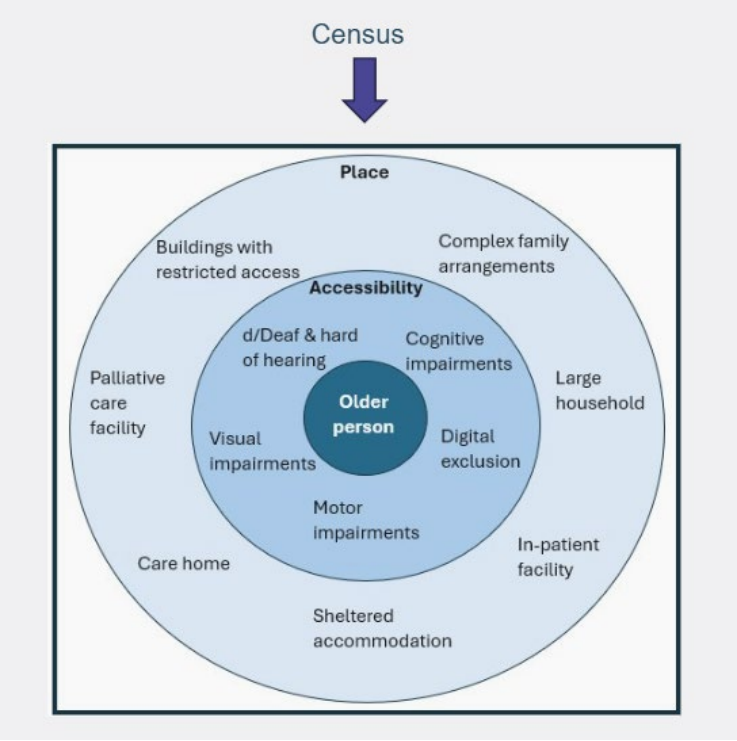
8. Undercount and poor-quality data for specific groups can:
    i. bias population estimates and outputs
    ii. affect resource allocation and service planning
    iii. reduce the visibility of inequalities

Discussion

Defining key population groups

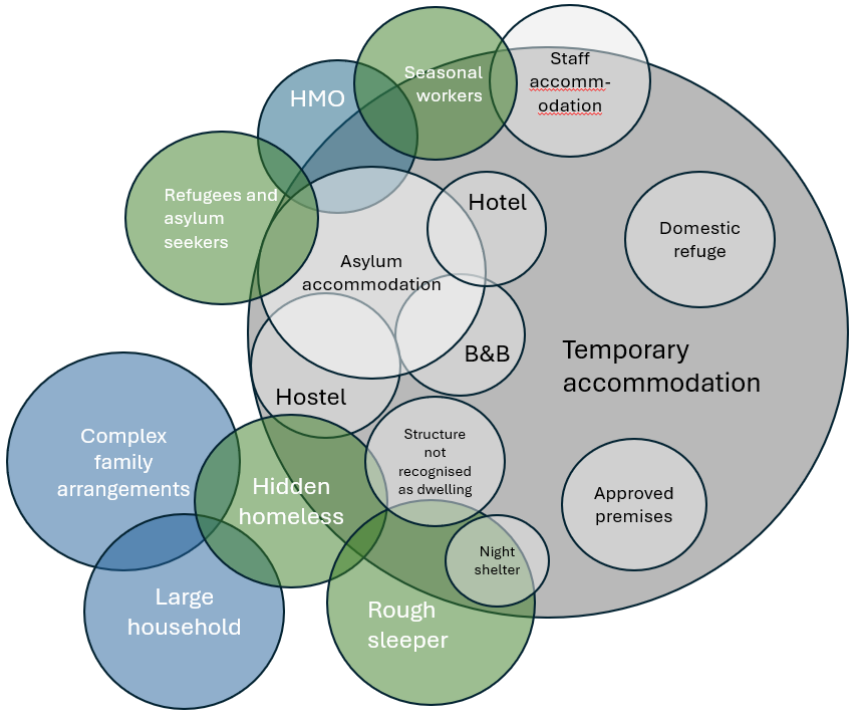
9. Key Population Groups (KPGs) are defined as groups that are less likely to respond to census, resulting from barriers to participation, where the standard census approach is less effective without adaptation.
10. KPGs can be place-based (e.g. people residing in care homes), situational (e.g. people without internet access) or characteristic (e.g. young people).
11. Individuals may face multiple barriers simultaneously, which often interact to increase the risk of nonresponse. As illustrated in figure 1, a respondent may experience accessibility barriers (e.g. health related or digital challenges), as well as place-based barriers (e.g. restricted access in institutional settings). These intersecting barriers reinforce the need for census design and delivery to consider the full user journey rather than isolated obstacles.

Figure 1. Intersecting barriers to enumeration for an older person



12. Individuals may therefore be members of more than one KPG simultaneously and should be understood as overlapping risk profiles rather than discrete or mutually exclusive populations (figure 2).

Figure 2 . Overlapping membership of KPGs



KPG approach

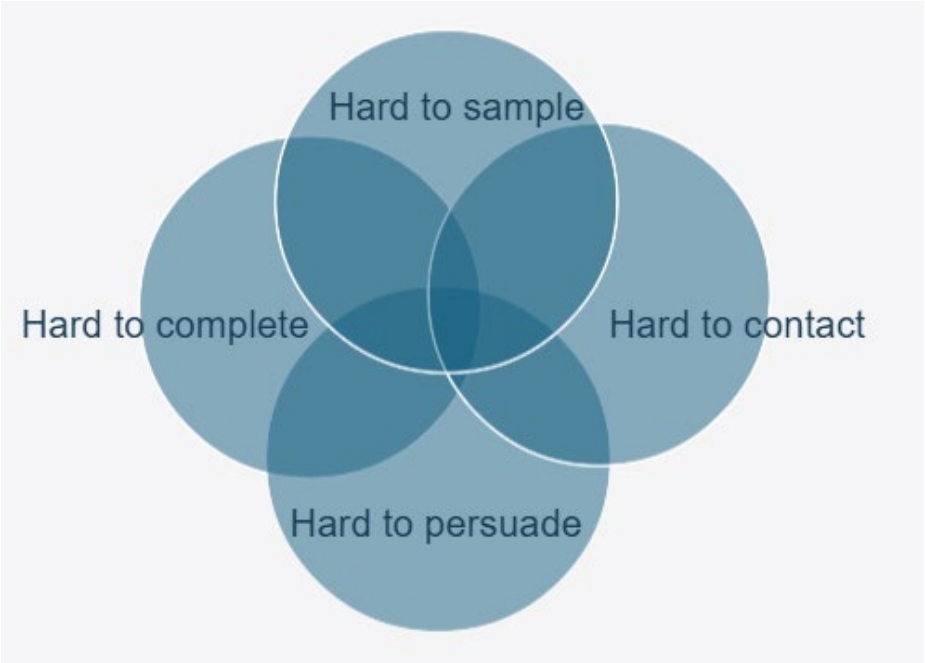
13. The KPG approach consists of four stages, identification, prioritisation, research and intervention design (figure 3).

Figure 3. KPG process

Identifying KPGs

14. The approach to identifying KPGs is based on research to identify barriers to survey completion, notably, the ‘hard-to-count’ framework developed by Tourangeau (2014). The hard-to-count framework uses each stage of survey design and delivery to identify barriers (figure 4).

Figure 4. Hard to count framework



i. hard to sample: groups do not appear on the address frame.

ii. hard to contact: physical accessibility barriers, such as institutional barriers to directly accessing individuals.

iii. hard to persuade: motivational barriers, such as low trust in government.

iv. hard to complete: barriers to completion including language, literacy and health.

15. This framework is widely used internationally, including in the UN Recommendations for the 2030 Round of Population and Housing Censuses (2025) and by the United States Census Bureau (2020).

16. Internationally, countries apply broadly similar, barrier-based approaches to identifying and supporting hard to count populations. For example, the United States segments groups by challenges such as locating, contacting, persuading and interviewing, while Australian Bureau of Statistics and Statistics Canada focus on targeted field operations, multilingual support and community partnerships to improve accessibility and response.

KPG identification framework

17. The identification of KPGs follows a staged approach to ensure clarity and robustness (figure 5).

18. **Stage 1** identifies the risks of non-response, data quality and scale for each group.

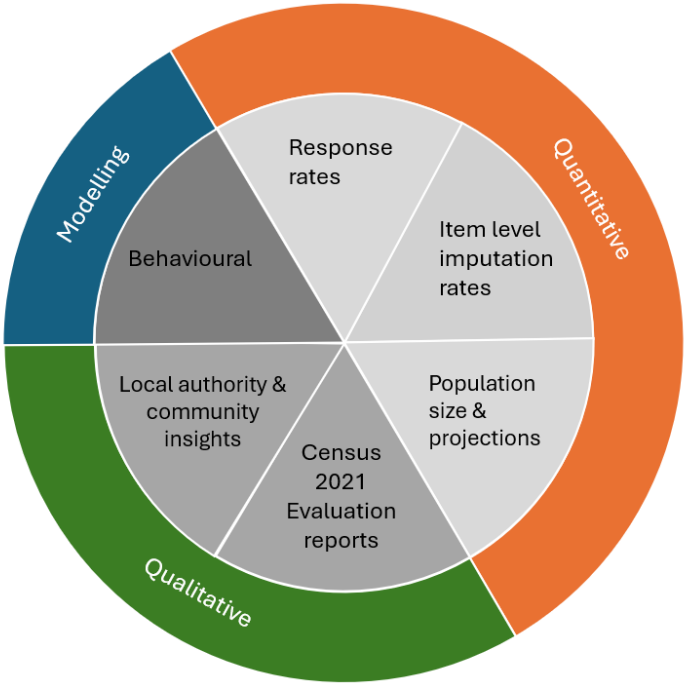
19. using Census 2021 data to identify groups with low response rates and high item level imputation rates, highlighting a risk to the quality of data collected.

20. incorporating population size and change over time to facilitate the identification of new groups, using population projections and migration data, where available.

21. **Stage 2** identifies challenges to enumeration for each group, in the form of barriers.

22. identifies groups which fall within the ‘hard to sample’, ‘hard to contact’ and ‘hard to complete’ groups. It consists of a comprehensive review of qualitative evidence from previous censuses, including evaluation reports and problem statements, together with insights from Local Authorities and communities.
23. the review identifies ‘operational’ barriers to enumeration; where groups may not appear on the address frame (e.g. homeless people), where physical access to respondents is restricted (e.g. care homes), and where ‘accessibility’ barriers (e.g. low proficiency in English), may limit participation.
24. **Stage 3** The third stage identifies groups which may be difficult to find using standard analytical techniques.
25. machine learning analyses patterns of behaviour in how people interact with the census, grouping people on observed behaviour, rather than prior assumptions. This approach may identify previously unrecognised groups or challenge existing assumptions about current KPGs.

Figure 5. KPG Identification framework



Multifaceted approach

26. The purpose of this multifaceted approach is to ensure that groups are not deprioritised simply because of high past response rates or due to their size, supporting a more forward-looking assessment of where tailored approaches will be required.
27. It is an iterative process. The list of KPGs will evolve over time, based on the emergence of new and updated evidence. Ongoing collaboration between teams will ensure insights are shared as they emerge, enabling any changes to be collectively understood, agreed, and implemented.

KPG prioritisation method

28. Prioritisation identifies where barriers to enumeration intersect to create the need for more significant levels of planning.

29. Each KPG is allocated a prioritisation score based on the data sources used to identify them, except for population projections and modelling, which are used solely to help identify groups. Table 1 sets out the measures used to establish prioritisation scores.

Table 1. KPG prioritisation

	Data	Scoring	Data sources	Purpose
Non-response	Quantitative	Log-scale	Response rates from 2021 Census	Identify KPGs with low response rates
Data quality	Quantitative	Log-scale	Item level imputation rates from 2021 Census	Identify KPGs with high imputation rates
Population size	Quantitative	Log-scale	Census 2021and administrative data	Identify size of KPG
Operational challenges	Qualitative	Number of identified operational challenges	Census 2021 evaluation reports, problem statements, Local Authority and community insights	Identify groups that are hard to sample and/or hard to contact
Accessibility challenges	Qualitative	Number of identified accessibility challenges		Identify groups that may find census hard to complete

30. The measures of non-response, data quality and scale are all based on quantitative measures. Both response and imputation rates are calculated as a percentage varying between 0-100%.
31. Scores are standardised to enable composite scores across both qualitative and quantitative measures to be calculated. The process of standardisation for each measure is included in Annex A.
32. The measure of non-response is based on response rates at various stages of census delivery (census day, census day +2, +10 and final response rates) to help identify barriers at different stages of the census process. Analysing response rates across time avoids potentially misleading conclusions from final response rates impacted by interventions.
33. Item-level imputation rates are used to assess data quality for each KPG, with higher levels of imputation increasing concerns around data quality.
34. Larger populations represent higher levels of risk to overall enumeration because even small gaps in response translate into large numbers of missing or imputed individuals. This means that lower response rates or higher imputation rates in KPGs with large populations have a greater impact on overall data quality.

35. The operational and accessibility challenges seek to measure the barriers to enumeration identified in the hard to count framework.
36. Operational challenges (hard to sample, hard to count) relate to practical difficulties in delivering the census, such as accessing people at the right time or place, navigating institutional arrangements, or managing field operations for mobile or hard-to-locate populations.
37. Accessibility challenges (hard to complete) exist where the standard design of the census does not align with a respondent's abilities or circumstances, for example due to language, disability, digital exclusion, or cognitive demands.
38. Scoring is based on the number of identified barriers for each measure. The lists of identified barriers are included in Annex B and C.
39. Standardised scores for each measure are combined using the weighting framework.

Measuring uncertainty

40. Scores are evidence based where possible. However, there are gaps in the evidence across the measures included in the prioritisation strategy. The absence of evidence may be an area of concern, requiring further research to acquire evidence.
41. Where there is no evidence for a specific measure, it is classed as 'unknown' and given a 'neutral' mid-point score. Areas of uncertainty are flagged to highlight evidence gaps. This ensures missing data does not distort prioritisation while identifying areas requiring further research.

Weighting

42. The weighting framework is designed to reflect the relative importance of each measure and ensure that KPGs with small populations are not overlooked. Each of the five measures is assigned a base weight that reflects its significance in predicting enumeration difficulty.
43. The weighting framework is based on the principles of Analytic Hierarchy Process (AHP) (Németh et al. 2019), a structured multi-criteria decision method. AHP is used to derive the relative importance of each prioritisation measure by structuring the problem as a hierarchy and eliciting expert judgement through systematic pairwise comparisons of factors, non-response, data quality, population size, and operational and accessibility challenges. These comparisons capture how much more important one factor is than another in influencing enumeration.
44. The resulting judgements are combined mathematically to produce ratio-scale weights, ensuring that all inputs are synthesised into a coherent and consistent weighting framework. The derived weights are then applied to standardised scores to generate overall prioritisation scores, providing a transparent, evidence-informed, and defensible basis for weighting within the KPG prioritisation process.

Sensitivity testing

45. Sensitivity testing is used to assess the degree to which results change under different assumptions, for example, if greater weighting is given to operational challenges, to understand whether prioritisation outcomes are driven by a particular set of assumptions or remain broadly consistent irrespective of differing assumptions.
46. For each KPG, sensitivity is quantified using the range of prioritisation scores observed across differing assumptions, calculated as the difference between the highest and lowest scores assigned under alternative weighting assumptions. (A more detailed description of sensitivity testing is included in Annex D.)

47. Groups with a small score range are considered robust to weighting assumptions, while those with a larger range are more sensitive and may require closer scrutiny in decision-making. This measure provides a clear and interpretable indication of robustness.

Key limitations and risks

48. The hard to count framework includes ‘hard to persuade’, or motivational barriers, which the KPG framework currently lacks. We will undertake research to identify where specific interventions should be developed to address these barriers.

49. Evidence gaps remain for some KPGs. These will be mitigated through ongoing research. Uncertainty is explicitly captured within the prioritisation framework and used to guide further research.

50. Prioritisation supports planning, but it does not prescribe specific interventions. Prioritisation does not represent a fixed hierarchy of importance, nor does it prescribe specific interventions. Instead, it provides a practical tool for focusing attention and resources where standard census approaches are least likely to be sufficient without adaptation.

51. KPGs represent patterns of risk and challenges to enumeration, not uniform experiences of respondents.

Conclusion

52. This paper sets out a robust and transparent KPG framework and prioritisation method for identifying and prioritising KPGs for census 2031, grounded in the integration of multiple data sources. This is critical to the delivery of high-quality population statistics and estimates that reflect the population and meet user needs.

Simon Watkins, Jess Bloomfield, Census Statistical Design, ONS, May 2026

List of annexes

Annex A – Standardisation of quantitative scores

1. Response rates and item level imputation rates
    i. The approach assigns a score from 0 to 5 based on a response rate. It gives a fixed score for unknown values (mid-score of 2.5), caps scores at 0 or 5 outside defined thresholds, and linearly scales results between those thresholds, with lower response rates receiving higher scores
    ii. If response rate is 94% or higher, score = 0 (lowest prioritisation). If response rate is 80% or lower, score = 5 (highest prioritisation).
    iii. Scores between 0.01-4.99 allocated based on mathematical formula (Score = (*High threshold – Actual value) / (High threshold – Low threshold) × 5*).
    iv. Target rates are based on the quality target of a 94% overall response rate for Census 2021, with a requirement that no local authority area fall below 80%. The same criteria is applied to each KPG.

2. Population size
    i. The approach assigns a 0–5 score using logarithmic scaling. It returns a default for unknown values (mid-score of 2.5), caps high values at 5, and otherwise increases the score based on the logarithmic position of the value between defined lower and upper bounds.
    ii. If population is above 1% of total population of England and Wales, score = 5 (highest prioritisation), if 500 or below, score = 0 (lowest prioritisation).
    iii. Scores between 0.01-4.99 allocated based on mathematical formula (*Score = (log10(actual) − log10(lower bound)) / (log10(upper bound) − log10(lower bound)) × 5*).

Annex B – Operational challenges

1. Access restrictions – Where access to buildings is limited or restricted and/or access is only available through gatekeepers
2. Misclassification issues - difficulty identifying dwelling type correctly, where specific to a population group
3. Delivery of questionnaires or UACs where inhabitants do not directly receive mail
4. Mobile population - may not be at address on or around census day
5. Multiple sites/locations across the country - e.g. halls of residence
6. No fixed abode - e.g. rough sleepers
7. Rely on third part support for operations - e.g. homeless charities to help locate rough sleepers
8. Heightened risk to field staff - e.g. accessing buildings not recognised as dwellings

Annex C Accessibility challenges

1. Language
2. Literacy
3. Visual impairments
4. d/Deaf and hard of hearing
5. Cognitive impairments
6. Mental health impairments
7. Digital exclusion
8. Neurodivergence

Annex D - Development of sensitivity scenarios

1. Sensitivity scenarios are designed as coherent alternative perspectives on prioritisation, rather than purely technical variations in weights. Each scenario reflects a plausible strategic emphasis within census planning, ensuring that variation is interpretable and relevant to census design and delivery operations.
2. Core scenarios include:
    i. Data quality-led scenario: places greater emphasis on response and imputation rates, prioritising risks to statistical accuracy.
    ii. Operational/feasibility-led scenario: increases the importance of operational challenges, reflecting delivery constraints and fieldwork complexity.
    iii. Inclusion and accessibility-led scenario: emphasises accessibility and motivational challenges, prioritising groups facing structural or behavioural barriers.
    iv. Population impact-led scenario: increases the weight assigned to population size, reflecting the broader impact of undercount on outputs.
3. These scenarios ensure that sensitivity testing captures the range of reasonable perspectives that may influence prioritisation decisions.

Bibliography

Australian Bureau of Statistics (2021) How data are collected, How the data is collected | Australian Bureau of Statistics (https://www.abs.gov.au/census/guide-census-data/census-methodology/2021/how-data-collected)

Németh, B., Molnár, A., Bozóki, S., Wijaya, K., Inotai, A., Campbell, J. D., & Kaló, Z. (2019) Comparison of weighting methods used in multicriteria decision analysis frameworks in healthcare with focus on low- and middle-income countries. Journal of comparative effectiveness research, 8(4), 195–204. https://doi.org/10.2217/cer-2018-0102

ONS (2025) Census Strategy for England and Wales, available at: Census 2031 Strategy for England and Wales - Office for National Statistics, accessed on: 22/04/26

Ormston, R., Hunt, M., Martin, C., Walker, A., Huskinson, T., Kotecha, R., Lynn, P. (2025) Understanding Survey Nonresponse Behaviours: Evidence and Practical Solutions, Scottish Government.

Statistics Canada (2020) Painting a portrait of Canada: The 2021 census of population. 3. Population - Counting all Canadians , July 16, 3. Population—Counting all Canadians, (https://www12.statcan.gc.ca/census-recensement/2021/ref/98-26-0001/2020001/003-eng.cfm)

Tourangeau, R. (2014) 'Defining hard-to-survey populations', in R. Tourangeau et al. (eds.) Hard-to-Survey Populations. Cambridge: Cambridge University Press, pp. 3–20.

UNECE (2025) Conference of European Statisticians, Recommendations for the 2030 Round of Population and Housing Censuses, Geneva.

U.S. Census Bureau (2020) 2020 Census operational plan (Version 5.0). https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan5-and-memo.pdf