2026 Bellringing Census

Document History

24 March 2026	Mark Ainsworth	Initial Draft - Objectives, Context, Outline Design including ‘gotchas’
25 March 2026	Vicki Chapman	Changes to geographical scope definition (no longer just UK)
28 March 2026	Mark Ainsworth	Revisions to methodology and gotchas, esp. data protection.
21 April 2026	Mark Ainsworth	Addition of Pilot Objectives section, following meeting with Bryn Reimer

Background

In the past 5 decades there have been 3 (almost 4) major surveys, lead by the CCCBR, conducted of the towers across the affiliated societies list.

Year	Design	Results available	Notes
1972	Thousands of towers, general information on band	Statistical summaries in paper form, for overall total of all responses and split by guild, all in form of 1970s-style computer printout. Photos of Total summaries in Google Drive.	Distribution and response-gathering process unknown.
1988	Random sample of 500 (~10%) of towers, activities in ‘sample week’	Multiple summary articles written for publication in RW, as scanned-in PDF of typewritten text.	These guys knew what they were doing. Multiple different survey designs for different roles.
2007-8	Thousands of towers, activities in ‘sample week’	Detailed survey responses at tower level, in Microsoft Access database and Excel spreadsheets. Some basic summaries. All held in Google Drive. No record of planning decisions.	Lack of clear sampling strategy rendered data unsuitable for estimating size of ringing population.
2015	10% sample of bellringers, to be randomly sampled from Association membership lists	None	Aborted shortly before planned launch due to objections around data protection / personal information

In support of the CCCBR’s Ringing 2030 strategy objectives, especially the intention to be recruiting significant numbers of new younger recruits, we want to undertake a new census to give us a current view of ringing.

Objectives

Primary Objective:
Estimate the health of the exercise by estimating the:

- Total number of active church bell ringers
- Composition of active ringing population: Age, Gender, Motivation, Ringing Ability, when/how learned, and other key metrics.

Secondary Objectives:

- Estimate state of health of towers
- Total number of active handbell ringers
- Ensure results can be interlinked with other known relevant data sources, including: Pealbase (no. of peals rung, no. of peal ringers, first peals); Bellboard (no. of pieces they've performed in that are posted on Bellboard); Dove (information about towers); Guild membership rolls; UK national census …
- Types of ringing activities participants engage in
- Activity levels and frequency
- Reasons for lapsed participation (for follow-up studies)
- Ensure findings are communicated back to the ringing community in a way that supports understanding, engagement, and future development at national and regional level

Detailed version of the truth - what is the state of play. How many ringers are there? What are the demographics? Dream of one person one bell. Use the information to plan retention, development things. Seek to understand the dynamics of retention, satisfaction.

The Survey Design Context

The Challenge

The essential challenge is posed by the first objective: to estimate the true number of bellringers, and information about them. This needs to be done to a precision high enough that the new initiatives to grow the population of ringers can be measured.

This represents a serious challenge as a surveying task, due to *Non-Response Bias*. Not every tower in which ringing occurs will respond to a survey. Not every active bellringer will respond to a survey. And the patterns of non-response are very likely to be connected to things about those towers/bands or ringers of interest. For example, keener ringers are more likely to respond. Thriving or larger towers are more likely to respond.

This completely messes up the validity of anything we estimate - we can't accurately say how many ringers there are IN TOTAL from survey-completing ringers without knowing how many of non-responding ringers there are, nor how many active towers there from the responding towers

without knowing about the non-responding towers. We can't say what proportion of ringers are keen because the keen ones are more likely to respond. And so on.

This has - at times - been well understood in the past. The 1988 survey was carefully designed to overcome this bias, by doing a full 'census' of activity in a single week at a random sample of 500 of the towers, leaning on the fact that Dove's Guide fully enumerates every ringable tower. Great efforts were made by 77 volunteer 'chasers' to ensure every one of the targeted towers provided a response. From these towers, a 33% sample of the individuals at those towers were also asked to complete an individual-ringer survey. Surveys were also issued to association and church representatives. These simple methods yielded counts and estimates which could easily be scaled up to provide anestimate given the technologies available at the time. Given that all of these surveys were filled in by handwritten pen on printed paper, posted back and forth, and needed to be manually coded into a computer, a fractional sample method was absolutely necessary for the task to be practicable.

The planned-for 2015 survey planned to address this issue by doing a 10% sample of Association membership lists, but this proved to be infeasible given Data Protection concerns.

The Opportunity.

Enough of the general population is now digitally-engaged - with a computer or smartphone - that it is realistically possible to do a near-complete survey of the ringing population using fully digital means. The marginal cost of someone completing a survey is zero. Notwithstanding the vagaries of having network signal inside a tower, it is almost universally feasible for a ringer to complete a survey on their phone between touches.

Moreover techniques in statistics, especially bayesian methods enabled by improvements in computing power, enable more powerful and nuanced methods for correcting for bias and using large samples of data, with well-defined good practices emerging in related fields such as national statistics (ONS, the census) and market research / election polling. These techniques have been adopted wholeheartedly in recent years in the election polling industry, where they are called MRPs = Multi-level Regression and Post-stratification. As here, non-response bias is a critical challenge, and these advanced methods allow for better correction for biases in different groups' likelihood to respond to a poll and likelihood to actually vote. This corresponds closely to our ringers' likelihood to respond to a survey and likelihood to actually ring.

Outline Design

Sampling Frames

We envision a two-stage sampling architecture anchored to the only complete dataset we possess: the physical towers. Dove's Guide provides the definitive 'universe' of ringable towers, serving as our primary sampling frame. We propose establishing a stratified sample of target towers. The first data capture layer would be a brief, operational 'headcount' survey completed

by the Tower Captain, giving us the structural baseline of the band. It may be useful to segment the band by age group given that this is (a) likely known by the captain, and (b) likely to be correlated with non-response propensity.

From these responding towers, the second layer of the frame drops down to the individual ringers, who would be asked to complete a more granular personal survey. The critical mechanism to validate this approach is the deliberate use of human enumerators, our “chasers.” We know from prior attempts that struggling or silent towers are the least likely to organically respond, creating a massive statistical blind spot regarding the true demographic floor of the Exercise.

By assigning enumerators to actively follow up with a specific sub-sample of towers, we can drive a near-100% compliance rate within that group. This forced compliance allows us to explicitly map the characteristics of the hard-to-reach, non-responding population of towers. The exact stratification parameters and size of the enumerator-chased sub-sample are decisions the analytical team will need to make, but it's worth being clear about what's at stake in that choice.

A random, representative sub-sample of towers gives us the ability to calibrate non-response rates across different tower types - essentially asking "what is the relationship between observable tower characteristics and likelihood to respond voluntarily?" That relationship can then be applied across all non-responding towers to correct the full dataset.

An alternative approach would deliberately over-index the chased sub-sample towards the towers with the *highest uncertainty* — the small, struggling, or silent ones we know least about. This maximises the informational value of each enumerator hour, but at the cost of making the non-response calibration more dependent on modelling assumptions, since the chased group is no longer representative.

These aren't mutually exclusive - a stratified design could pursue both goals - but they require different analytical handling downstream. The choice should be driven by what the reweighting model actually needs, which is another reason to engage the analytical team early, before the sampling architecture is fixed.

Unlike in 1988, in principle - because of the zero marginal cost from digital tools - we could encourage responses from *every* tower and *every* ringer in the land. Nevertheless we will need to re-apply the sampling techniques used in 1988 because there will still be a significant human effort constraint from the enumerators, and so it’s likely to be more fruitful to have those enumerators reach 100% coverage of all of their small number of towers than be spread thinner over more towers.

Reweighting Methodology

Capturing the data is only the first half of the problem; the second is mathematically correcting for the inevitable biases in our response distribution.

The key insight is that our enumerator sub-sample, by achieving near-100% compliance within a defined group of towers, gives us something rare: a window into the non-responding population. If we can characterise how the hard-to-reach towers differ from voluntary responders — on dimensions like band size, urban/rural setting, number of bells, or association — we can use that relationship to re-weight the full dataset back towards a nationally representative picture.

The ringing data has a natural nested structure: individual ringers sit within bands, bands within towers, towers within association branches, branches within geographic regions. This hierarchy matters statistically. A ringer in a struggling rural six-bell tower is not exchangeable with one in a thriving city twelve — and any model that ignores this nesting risks misattributing variation to the wrong level. The analytical team will be best placed to choose the specific framework, but the problem structure is well-suited to approaches that model this hierarchy explicitly, and that can incorporate prior information — for instance, from the 1988 or 2007 surveys — where it is informative.

The output we're aiming for is not a point estimate but a range: "There are between X and Y active ringers in the UK" with a well-characterised level of confidence. The tighter that range, the more useful it is as a baseline for measuring the impact of Ringing 2030 initiatives. Determining what level of precision is actually achievable — and what sample sizes and enumerator effort that implies — is one of the first tasks for the analytical team.

‘Gotchas’ to consider

A few complications that will need careful thought to handle:

Clarity of tower identity.

Our planned survey platform (QuestionPro) allows for definition of questions where the user has to choose from a predefined list. We have verified that this can be preconfigured with all the many towers from Dove. The alternative, typical approach - letting users type in the identity of their tower as free text - would cause tremendous problems given how vague, ambiguous and duplicated tower names can be.

Itinerant ringers who ring in multiple towers.

The 1988 survey acknowledged this as a likely source of over-counting: a well-travelled ringer might appear in the headcount at two or three towers in the same week. In a two-stage design - Tower Captain headcount followed by individual ringer survey - this is partially addressable. Including a "primary ringing tower" or "home tower" field in the individual survey, combined with explicit deduplication logic at the analysis stage, would allow us to identify and resolve most double-counting. The analytical team should design this field carefully: for some ringers, "home tower" is unambiguous; for others (itinerant ringers by definition, or university ringers) it is not,

and those edge cases need a defined handling rule rather than being left to respondent interpretation.
Towers under the care of a non-tower-defined band, such as the London towers rung by the College Youths, are a clear special case here.

University ringers and their towers.

These are real ringers, but their relationship with towers is more moveable. Sampling in a boring, middle-of-term week (3rd week of November) was the solution used in 1988 and likely to be a good idea this time too.

Denominator definition: who counts as a ringer?

Before any weighting or estimation can begin, the team needs an operational definition of what we're counting. This has two dimensions that interact:

The activity boundary. Is the census capturing "people who ring" in a general sense, or "people who rang in the census week"? A census-week measure is cleaner and more objective, but will undercount ringers who are temporarily inactive - on holiday, recovering from injury, or simply away. A broader definition requires more subjective judgement, both from Tower Captains reporting their band and from individual ringers self-reporting.

The skill/engagement boundary. Does a learner who has attended three handling lessons count? Does a highly skilled ringer who only rings for weddings once a quarter? There's no obviously correct answer, but the choice needs to be made explicitly and applied consistently, because the reweighting models will amplify any fuzziness in the denominator.

These two boundaries also interact with a subtler risk: optimism bias in Tower Captain reports. A struggling tower may over-report its active numbers or under-report its reliance on visiting ringers to sustain Sunday service. Cross-validating Tower Captain headcounts against individual ringer responses - particularly on home tower and ringing frequency - gives us a mechanism to detect and quantify this.

The 1988 survey's choice of a mid-term Monday-to-Sunday window (avoiding festivals, holidays, and inter-tower visiting) was a deliberate attempt to capture a 'steady state' week. A similar choice this time, combined with a clear written definition of who the Tower Captain should include in their count, would go a long way to reducing ambiguity at source.

Data Protection and personal data.

This project will handle personal data and must be fully compliant with UK GDPR throughout. The following principles should shape the data architecture from the outset - before any collection begins - and a qualified data protection professional should be consulted to finalise the legal basis and controller arrangements.

The mosaic effect. The ringing community is small, geographically specific, and highly networked. Even where respondents do not provide their name or contact details, the combination of home tower, age bracket, years of experience, and highest method rung can create a unique demographic fingerprint - enough for anyone with local knowledge to re-identify an individual. This means that true anonymisation of *individual survey returns* is functionally impossible at the point of collection, and we should not rely on claims of anonymity as a legal or ethical shield. All individual returns must therefore be treated as Personally Identifiable Information (PII) throughout collection and processing. Aggregate published outputs are a different matter: with appropriate cell suppression (see below), these can be genuinely anonymised.

Lawful basis. The most intuitive basis is explicit consent under Article 6(1)(a) of UK GDPR, and an opt-in consent step is almost certainly appropriate regardless. However, consent carries an unqualified right to withdrawal at any time prior to aggregation, which creates operational and analytical fragility if withdrawals are non-random - concentrated, for example, in a particular tower type or demographic. For a statistical census conducted under a charitable public interest mandate, Article 6(1)(e) (public task) or Article 9(2)(j) (scientific research) may be more robust bases, either alongside or instead of consent. This is a decision for a data protection professional.

Data Controller. The consent declaration, survey instrument, and data architecture must all name a defined Data Controller - the legal entity responsible for how personal data is collected, stored, and used. Whether this is the CCCBR, a specific affiliated society, or another body needs to be resolved before any data collection begins, as it determines liability and retention obligations.

Informed opt-in. Whatever the lawful basis, the survey instrument should present a clear, mandatory acknowledgement before any data is collected. This should state: who the Data Controller is; what the data will be used for; how long identifiable records will be retained; and - if consent is the chosen basis - how a participant can request withdrawal prior to final aggregation. These are not bureaucratic requirements: they are the mechanism by which we maintain trust with the ringing community, on whose goodwill the entire project depends.

Trust in data handling is likely to be a material driver of participation, and therefore of data quality. Communication of data protection arrangements should therefore be considered not only as a compliance requirement, but as a factor influencing response behaviour.

Retention and deletion. UK GDPR requires that personal data is not held longer than necessary. The data architecture should define, in advance, the point at which identifiable individual records can be purged once aggregate outputs have been produced. Given that Ringing 2030 initiatives may generate follow-up surveys, the retention schedule should also consider whether longitudinal linkage of records is intended - if so, this must be explicitly consented to and documented.

Cell suppression in published outputs. Slicing results by multiple variables - for example, female ringers under 30 in a specific rural branch - can inadvertently recreate individually identifiable records even from an aggregated dataset. The analytical team should establish a minimum cell-size threshold below which results are suppressed or combined, consistent with standard ONS practice.

Non-response as a known consequence. Robust data protection design will itself generate non-response: some ringers will decline to participate precisely because of consent and data handling requirements. This is unavoidable and should be treated as a known, modelable source of missingness - not as a failure of the survey instrument.

Additional Design Considerations: Participation, Trust and Use of Outputs

While the statistical and sampling challenges are central to the success of this census, they are not the only determinants of data quality.

The effectiveness of any design will depend critically on:
- Participation rates across different segments of the ringing population
- Trust in how data is being collected, used, and protected
- Clarity of purpose and perceived value by those being asked to respond

These factors are not independent of the statistical design. Patterns of non-response will be shaped not only by structural characteristics (e.g. tower size, geography, activity level), but also by how the census is communicated and understood within the community.

This raises a set of design questions that sit alongside the analytical framework:
- How should the purpose of the census be framed to maximise broad and representative participation?
- What level of feedback or return of results to participants is necessary to support engagement and goodwill?
- How might different communication strategies affect response patterns — and therefore the assumptions required in reweighting models?
- To what extent should communication and engagement be treated as part of the experimental design, rather than an implementation detail?

There is a risk that a technically robust design could still yield biased or incomplete data if participation is uneven or trust is low. Conversely, strong engagement may reduce the burden on modelling assumptions.

Input from the design group on how best to integrate these considerations — conceptually and practically — would be valuable.

The central challenge is not only to design a survey that *can* produce a valid estimate, but one that the ringing community will engage with in a way that allows it to do so.

Pilot Objectives: Analytical Priorities

The summer pilot serves two distinct purposes: testing the survey instrument itself (question clarity, completion rates, platform behaviour), and generating the empirical estimates needed to make key analytical decisions before the main census goes into the field. This section addresses the second of those purposes — the specific unknowns that most affect the validity and precision of the final estimates.

What the pilot needs to resolve

1. The voluntary response rate as a function of tower characteristics

The entire reweighting strategy depends on being able to model which towers respond and which don't, as a function of observable tower-level covariates. Our working prior is that the dominant predictors are number of bells and degree of urbanisation — as proxies for band size, population catchment, and ringing culture — with settlement density capturing some of the itinerancy dynamic in areas where towers are geographically clustered. The pilot needs to test whether this parsimonious model is adequate, or whether additional covariates (regional association, congregation health, tower ownership) carry independent predictive weight. If response propensity is poorly predicted by observable characteristics, the reweighting model becomes heavily dependent on modelling assumptions rather than data — and the credible intervals on the final estimate will be correspondingly wide.

2. Tower Captain optimism bias

The two-stage design cross-validates Tower Captain headcounts against individual ringer responses. The pilot should explicitly measure the direction and magnitude of any systematic discrepancy — particularly whether struggling towers over-report active band members, and whether towers reliant on visiting ringers accurately characterise that dependence. If optimism bias is large and systematic, the analytical team will need to model it explicitly rather than treat Captain reports as unbiased observations.

3. The itinerant ringer / double-counting rate

Ringers who are active across multiple towers in the same week represent a genuine double-counting risk in any tower-anchored headcount. The pilot should attempt to quantify what proportion of responding ringers identify a tower other than the survey tower as their primary ringing home. Even a rough empirical estimate of this rate — and whether it varies systematically with tower type — will allow the main census to apply a principled correction rather than an arbitrary assumption.

4. Individual response rate conditional on tower participation

The main census design assumes that a meaningful fraction of ringers within responding towers will complete the individual survey. If that conditional response rate is low, or strongly

age-skewed, the individual-level estimates will carry much greater uncertainty than the tower-level estimates. The pilot should measure this rate and its covariates explicitly.

5. Congregation health as a covariate

Tower activity is partly a function of the health of the associated congregation — a declining or absent congregation removes the structural occasion for Sunday ringing and reduces institutional support for tower maintenance and recruitment. This is qualitatively different from the structural tower covariates in that it is time-varying and unobservable from external datasets. The pilot should test whether a simple direct question to the Tower Captain (for example, a single ordinal item on congregation size or trajectory) yields usable signal, or whether the question is too sensitive or subjective to answer reliably. If it can be measured, it belongs in the response propensity model. If not, it needs to be acknowledged as an unobserved confounder.

What "good enough" means

The pilot should generate posterior predictive checks on the width of the credible interval for total ringer count under different assumptions about chaser sub-sample size. The analytical team should agree in advance on a target precision — for example, a 90% credible interval of ±15% on the national ringer estimate — and use the pilot response data to determine whether that target is achievable with the planned enumerator resource, or whether the design needs adjustment before November.