NANA

GUIDE · OCTOBER 2025



SURVEY DESIGN FOR ASSOCIATIONS

A practical guide to writing member surveys that produce year-over-year comparable data.

Inês Carvalho-Reyes

RESEARCH AND BENCHMARKING

ABOUT THIS GUIDE

This guide shows association staff how to write, field and report a member survey that can be compared with itself year after year. It is written for research directors, membership managers and executive directors who run a survey once a year with a small team and no statistician. It draws on the practice of NANA's research team, which has run the State of Associations survey since 2024.

Its four rules are short. Fix a question's wording the first time you ask it, and size the sample for the smallest segment you will report. Choose a scale on purpose, and print the margin of error beside every headline figure.

HOW TO USE THIS GUIDE

Read sections 1 to 3 before you write a single question, because they fix the decisions that cannot be changed later. Sections 4 to 7 cover the response scale, the pilot, the margin of error and the calendar, and they can be read in any order. Section 8 lists the mistakes we see most often in a first tracked survey. Section 9 shows the whole design at work in the State of Associations, and the appendix gives you a register and a change log to copy into your own files.

AUTHORS

Inês Carvalho-Reyes

Guide · October 2025. Published by NANA. Cover photograph: generated for this guide.

CONTENTS

1	Why comparability is worth protecting	4
2	Writing questions that survive multiple survey years	6
3	Setting a sample size	9
4	Choosing a response scale	12
5	Piloting before fielding	15
6	Reporting a margin of error	18
7	Timing and reminder cadence	21
8	Common mistakes in a first tracked survey	24
9	The State of Associations survey in practice	27
10	Where to go next	30
	Appendix: the fixed-design register and change log	30

1

WHY COMPARABILITY IS WORTH PROTECTING

Your board will ask why you cannot fix one awkward question. By the end of this section you can answer with a number, and you can list the five things your own survey must hold still.

The value of a member survey compounds over time. One year's results tell you what members think today. Three or five years of the same questions, asked the same way, show whether sentiment on an issue is rising, falling or holding, and a board usually gets more from that line than from any single number. Comparability is fragile, though. A survey redesigned each year to fix small wording problems, add a timely question or shorten a section produces snapshots that cannot honestly be set side by side, even when the topic looks the same.

This guide rests on one discipline: decide what stays fixed before the first survey goes out, and protect it when a later team is tempted to improve the wording. NANA's State of Associations survey has kept a stable core question set since its 2024 launch. It adds questions at the edges and never rewords one that anchors a trend. The payoff shows in the headline series. Median membership growth read 2.1 percent in both editions, and median renewal read 93 percent in both, so a reader can compare the two years knowing the questions underneath did not change.

Table 1. The State of Associations headline series, 2024 and 2025.

EDITION	RESPONDING ORGANIZATIONS	MEDIAN MEMBERSHIP GROWTH	MEDIAN RENEWAL
2024	690	2.1%	93%
2025	720	2.1%	93%

Read the table the way a board member would. Growth and renewal both held, which is a finding, and no one has to wonder whether a rewording moved them. The sample grew by 30 organizations between the two years with no change to the wording. A growing sample on a fixed instrument is what makes the series worth quoting, and a one-point move in either column would still have to be read against its margin of error (section 6).

WHAT HAS TO STAY FIXED

Five things have to hold still for a trend to mean anything: the wording of each tracked question, the response scale attached to it, and the population

the survey is drawn from. The other two are the fielding window and the reminder cadence, because who answers depends on when and how often they are asked.

Everything else can change. You can add a question, retire a question, redesign the invitation or move the survey to a new tool, provided none of the five moves. The appendix gives you a register that lists all five in one place, so a new hire can see in a minute what they are not allowed to touch.

WORKED EXAMPLE: ONE ADJECTIVE, SIX POINTS

The Council of Early Childhood Educators has 11,000 members and runs a member survey every year. In 2023 it asked, "How satisfied are you with the value of your membership?" and 64 percent of respondents chose satisfied or very satisfied. In 2024 a board member found the phrase vague, and the team changed the question to "How satisfied are you with the value your membership delivers for your center?" The 2024 result was 71 percent, and the annual report announced a seven-point gain.

Mei-Lin Harbaugh, the council's Director of Data and Analytics, was not sure the gain was real. In spring 2025 she split the sample. Of 2,200 responses, 1,100 respondents were randomly shown the 2023 wording and 1,100 were shown the 2024 wording. The old wording scored 66 percent and the new wording scored 72 percent, a six-point gap between two halves of the same membership answering in the same month. Each half carried a margin of error of about 2.8 points, so the gap was well outside the noise.

The lesson is arithmetic. If wording alone is worth about six points, then roughly six of the seven points reported in 2024 came from the edit, and the real change in satisfaction between 2023 and 2024 was near one point. The council now keeps both wordings in a bridge table, reports the new wording going forward, and labels the series break on every chart.

A split-sample test, as the council ran, is the honest way to repair a break once it has happened. It costs a full survey cycle and a larger sample, which is the reason to avoid the break in the first place.

> “Our members compare this year's numbers with last year's whether we invite them to or not. The least we can do is make sure the questions match.
>
> LUCINDA FARROW-MBATHA, EXECUTIVE DIRECTOR, FEDERATION OF COMMUNITY LIBRARIES

> “We changed one adjective in 2024 and spent most of 2025 proving it had moved our score by six points. I would rather have lived with a vague question.
>
> MEI-LIN HARBAUGH, DIRECTOR OF DATA AND ANALYTICS, COUNCIL OF EARLY CHILDHOOD EDUCATORS

2

WRITING QUESTIONS THAT SURVIVE MULTIPLE SURVEY YEARS

Write a question once and it should still read the same in 2030. By the end of this section you can test any draft against six rules and rewrite it before a member sees it.

A reusable question carries no trace of the year it was written. "Given this year's dues increase, how satisfied are you with member value?" fails next year, when there is no increase to mention. A better version asks about member value on its own terms, so it means the same thing whether or not anything unusual happened.

Avoid questions that ask two things at once. "How satisfied are you with communications and events programming?" forces a member who holds two opinions into one answer, and you cannot say which topic the result describes. Each tracked question should isolate one idea, so a respondent reading it years from now understands it without this year's context.

Once a tracked question has been fielded, treat it as fixed. If you find a real flaw, such as an ambiguous phrase that members keep misreading, retire the question and start a new trend line with a clearer replacement. Do not edit the wording and keep reporting against the old baseline as if nothing changed.

SIX RULES FOR A TRACKED QUESTION

Apply these rules to every draft. A question that fails one should be rewritten before the pilot, not after the first field.

- Ask one thing. If the question contains "and" or "or" joining two topics, split it.
- Name no year, price, event or person. A question that mentions a specific dues increase or a departing executive expires with that event.
- Use the respondent's words. Say "dues", not "membership fee", and "chapter", not "regional affiliate".
- Define the reference period. "In the last 12 months" survives; "recently" does not.
- Offer every honest answer. If a respondent could reasonably say "not applicable" or "I don't know", provide the option, because forcing an

answer corrupts the count.

- Keep the question neutral. A question that signals the approved answer produces approval.

Table 2. Rewrites that make a question reusable.

DRAFT QUESTION	PROBLEM	REUSABLE VERSION
Given this year's dues increase, how satisfied are you with member value?	Names a one-time event	How satisfied are you with the value your organization receives for its dues?
How satisfied are you with communications and events?	Asks about two topics	Ask twice: one question on communications, one on events
Do you attend chapter meetings regularly?	"Regularly" means something different to each respondent	In the last 12 months, how many chapter meetings did someone from your organization attend?
Wouldn't you agree that the new portal is easier to use?	Leads the respondent	How easy or difficult is it to find a resource in the member portal?
How has our new executive director done so far?	Names a person and a moment	How confident are you in the association's current leadership?

KEEP THE NUMBER OF TRACKED QUESTIONS SMALL

Every tracked question is a commitment, so tracking should be rationed. Ask what decision a board would make differently if the answer moved by five points. If nobody can name one, the question is context rather than a trend, and it belongs in the rotating part of the survey where it can be reworded freely.

A useful split is two tiers. The core tier holds the questions you will track for at least three years, written once and never edited. The rotating tier holds timely questions about a new program, a pending vote or a current event, and it can change every cycle because nobody expects it to line up with last year.

WORKED EXAMPLE: A COMPOUND QUESTION SPLIT IN TWO

The Society of Rural Hospital Administrators has 1,450 members and 19 staff. Its first tracked survey, fielded in 2024, asked, "How satisfied are you with the society's webinars and conferences?" The answer was 68 percent satisfied, and the staff had no idea which of the two programs was carrying the number.

In the pilot, a member told the team she attended every webinar and had never been to a conference, so she did not know how to answer. The team split the question into two and fielded both from then on. Webinars scored 79 percent and conferences scored 52 percent, and the 68 percent in the original wording had been an average of two very different experiences. The conference score became the first item on the next year's program committee agenda.

Notice that the split cost nothing except one extra question. It would have been far harder to do in year three, after a trend line had been built on the compound wording.

> “The combined number told us we were fine. The two numbers told us exactly where we were not.
>
> MARGUERITE TILLERY-OSEI, DIRECTOR OF PROGRAMS, SOCIETY OF RURAL HOSPITAL ADMINISTRATORS

RETIRING A FLAWED QUESTION

Sometimes a question turns out to be broken despite the pilot. When that happens, write down what went wrong, retire the question on the register, and start a new question with a new label and a new baseline year. Do not fold the new data into the old series, and do not delete the old data, because the old series is still a true record of what members said to that wording.

Add a note to the next report: the question was retired, here is why, and the replacement begins a new trend. Members who follow your data will respect the plain account far more than they would a series that quietly changes shape.

- [ ] Every tracked question asks about one topic only.
- [ ] No tracked question names a year, a price, an event or a person.
- [ ] Each question defines its reference period in words.
- [ ] Each question offers "not applicable" or "don't know" where an honest respondent might need it.
- [ ] The core tier is listed in the register and the rotating tier is marked as such.
- [ ] Each tracked question carries a baseline year and a label that will never change.

3
SETTING A SAMPLE SIZE

A 40 percent response rate can still leave your smallest region with 11 answers. By the end of this section you can calculate the responses each segment needs and decide which segments are too small to report.

Most association surveys size the sample against total membership, but the finding a board wants is usually about a segment: small member organizations, one region, staff with under two years of tenure. A survey with a comfortable overall response rate can still be too thin for the breakdown the board asks for, because that segment is a small share of respondents.

The working rule is to name the smallest segment you will report before you field anything, and to calculate the sample for that segment. If you will report by region, the smallest region needs enough responses to stand on its own. Sometimes that means oversampling it on purpose, with extra reminders to its members, until every segment in the report has a usable count.

THE CALCULATION, STEP BY STEP

You need three inputs: the number of members in the segment, the margin of error you are willing to accept, and a confidence level. Use 95 percent confidence unless your board has a reason to ask for something else. Use a 50 percent assumed answer, which is the cautious case and gives the largest sample.

First compute the sample you would need from an unlimited population, which for a margin of 5 points at 95 percent confidence is 384. For a margin of 10 points it is 96. Then adjust for the real segment size with the finite population correction: divide that number by one plus the figure less one, divided by the segment size. For a segment of 40 members and a margin of 5 points, that is 384 divided by 1 plus 383 over 40, which comes to 37 responses. The correction matters most for small segments, because a region with 40 members cannot return 384 responses and does not need to.

WORKED EXAMPLE: FOUR REGIONS, ONE SURVEY

The Federation of Agricultural Cooperatives has 610 member co-ops and 26 staff. In 2025 its research lead planned a member survey to be reported

by region: Central with 270 co-ops, South with 190, East with 110 and Mountain with 40. The question was what it would take to report each region within 5 points.

Table 3. Responses needed to report each region within 5 points at 95 percent confidence.

REGION	MEMBER CO-OPS	RESPONSES NEEDED	SHARE OF MEMBERS WHO MUST RESPOND
Central	270	159	59%
South	190	128	67%
East	110	86	78%
Mountain	40	37	93%
All four regions, combined	610	236	39%

The last row is the number most people calculate, and it is the misleading one. Reporting the whole membership within 5 points takes 236 responses, which is 39 percent of members. Reporting every region within 5 points takes 410 responses, the sum of the four region rows, which is 67 percent of members. The second goal needs about 74 percent more responses than the first, and the Mountain region needs nearly everyone to answer.

DECIDE WHAT YOU WILL NOT REPORT

The table shows the real choice. You can fund a very high response rate in the small regions, loosen the target for them, or decline to publish a regional figure. Loosening is usually right. For a margin of 10 points, Mountain needs 29 responses, a 73 percent response rate, and East needs 52 responses, a 47 percent response rate.

A fourth option is to pool years. A two-year figure for Mountain is smoother than either year alone, and it must be labeled as a two-year figure. Many of the same co-ops answer in both years, so pooled responses do not behave like a larger sample. Use a pooled figure to describe a region and never to test a change.

State your reporting rule before fielding and keep it. A reasonable rule reads: "We publish a segment figure only when we hold enough responses to report it within 10 points at 95 percent confidence." Anything below that threshold is described in words, never in a percentage.

OVERSAMPLING AND WEIGHTING

Oversampling means asking more of a small segment than its share of the membership would suggest. In a census survey, where every member is invited, oversampling takes the form of extra reminders and a personal call from a regional leader. In a sampled survey, it takes the form of a larger invitation list for the small segment.

When oversampling changes the mix of respondents, a headline figure for the whole membership needs weighting so that each segment counts in proportion to its true size. Without weighting, a heavily oversampled Mountain region would count far more in the overall figure than its 40 co-ops deserve. Write the weighting method into the register, because a change to it is a change to the design.

- [ ] The smallest segment the report will cover is named in the register.
- [ ] The required responses are calculated for each segment, not only for the whole membership.
- [ ] A minimum for publishing a segment figure is written down.
- [ ] A plan exists for the segments that will fall short: extra reminders, pooling years or no figure.
- [ ] The weighting method, if any, is documented and will be repeated.

> “A national average tells me little about a region of 40 co-ops, and a figure built on 11 answers tells me less.
>
> CELESTE VRANAS, VICE PRESIDENT OF MEMBER SERVICES, FEDERATION OF AGRICULTURAL COOPERATIVES

4
CHOOSING A RESPONSE SCALE

Your survey tool picked a scale for you before you wrote a question. By the end of this section you can pick your own, defend it to a board and keep it for years.

A five-point scale, a seven-point scale and a yes or no question measure different things, even on the same topic. A five-point scale with a neutral midpoint lets a member register real ambivalence. A forced-choice scale makes the member lean one way. That helps when you want to avoid a wave of noncommittal answers, and it hurts when it manufactures a clear sentiment that does not exist.

Choose the scale for what each question must measure, not for what your platform offers by default. Then keep the scale fixed for as long as the question is tracked, for the same reason you keep the wording fixed. Switching from five points to seven mid-series breaks the comparison as thoroughly as rewording the question would.

SCALE	BEST FOR	MAIN RISK
Five-point agreement with a midpoint	Satisfaction and confidence questions where ambivalence is real	A large middle can hide a split membership
Four-point forced choice	Questions where the board needs a lean one way	Manufactures a clear sentiment among people who have none
Seven-point	Fine distinctions among a highly engaged panel	Respondents rarely use the outer points, and the labels blur
Yes or no	Facts: did you attend, do you hold a credential	Flattens any question that is really a matter of degree
Zero-to-ten likelihood	Likelihood to renew or recommend	Hard to compare with any other scale, and the cutoffs are arbitrary
Frequency categories	Behavior: how often, how many	Categories must be fixed in advance or counts cannot be compared

Table 4. Common scales and what each one measures.

LABEL EVERY POINT

An unlabeled numeric scale invites each respondent to invent a meaning. Label every point in words: very dissatisfied, dissatisfied, neither satisfied nor dissatisfied, satisfied, very satisfied. If you label only the endpoints, the middle values mean different things to different people, and the spread of answers widens for reasons that have nothing to do with sentiment.

Keep the order and the direction consistent through the survey as well. If most scales run from negative to positive and one runs the other way, some respondents will miss the reversal and answer the opposite of what they intend. A pilot catches this, but only if you watch for it.

WORKED EXAMPLE: WHEN THE MIDPOINT DISAPPEARS

The League of Independent Pharmacists has 3,600 members. Its 2024 survey asked how satisfied members were with the league's advocacy, on a five-point scale with a neutral middle. The result was 58 percent satisfied, 22 percent neutral and 20 percent dissatisfied.

In 2025 the survey platform's default changed to a four-point forced-choice scale, and the team accepted it without comment. The result came back as 71 percent satisfied and 29 percent dissatisfied. The report announced a 13-point gain in satisfaction with advocacy.

Noor Haddad-Ellison, the league's Chief Membership Officer, noticed that the 22 percent who had been neutral had to go somewhere. Thirteen points went to satisfied and nine points went to dissatisfied, which is exactly the change between the two years. The underlying sentiment had not moved. The league will restore the five-point scale in its next field and notes the 2025 break on every chart.

CHOOSE ONCE, THEN LOCK IT

Pick the scale when you write the question, record it in the register with the exact labels, and treat it as part of the question. When a new tool offers a different default, set the scale by hand and check it in the pilot. When a board member asks for a seven-point scale because it sounds more precise, explain that more points do not make a more accurate answer, and that changing now would cost the series its history.

If a scale truly must change, run the old and new versions side by side for one full cycle, as the Council of Early Childhood Educators did with its wording, and publish the bridge. Anything less leaves readers comparing numbers that do not match.

- [ ] Each tracked question has a scale chosen for what it measures, and the reason is on file.
- [ ] Every point on each scale carries a word label.
- [ ] Scale direction is consistent through the whole survey.
- [ ] The survey tool's default has been checked and overridden where it differs from the register.
- [ ] Any planned scale change has a one-cycle side-by-side test scheduled.

> “The platform changed a default, we missed it, and for a year we told our board that members had warmed to us. They had not. We had removed the place where they could say so.

NOOR HADDAD-ELLISON, CHIEF MEMBERSHIP OFFICER, LEAGUE OF INDEPENDENT PHARMACISTS

5
PILOTING BEFORE FIELDING

Fourteen people and one afternoon can save a year of data. By the end of this section you can run a small pilot, read what it shows and decide what to change before the first invitation goes out.

Pilot every tracked question with a small group before it reaches the full membership. Use a chapter leadership call, a handful of staff volunteers or a sample of the target members. The pilot catches ambiguous wording, questions that take too long and scales that do not fit. Those faults are cheap to fix now. After the full survey they either sit inside your trend line for good or force you to discard a year of data.

WHO TO INCLUDE

Aim for 10 to 15 people who resemble your respondents, not your colleagues. Include at least one person from your smallest segment, one from your largest, and one who does not read your newsletter, since the survey will reach people like that. A chapter roundtable is a convenient audience because the group is already gathered and used to giving candid feedback.

Do not include the people who wrote the questions. They know what each item means and cannot see what a stranger would make of it.

WHAT TO ASK THEM

Run the pilot in two parts. In the first part, ask each person to complete the survey on their own and record the time it takes. In the second part, sit with them for 15 minutes and ask three things about each question that gave them pause: what did you think it was asking, which answer did you choose, and what were you thinking about when you chose it.

The second question is the most revealing. When two pilot respondents give the same answer for different reasons, the question is measuring two things at once. When a respondent cannot find an answer that fits, the options are incomplete.

SIGNAL	WHAT IT MEANS	ACTION
Two people read a question differently	The wording is ambiguous	Rewrite, then repeat the pilot on that question
A respondent skips a question	The question may not apply, or may feel intrusive	Add a not-applicable option or reconsider asking
Median completion time is over 10 minutes	The survey is too long for the response rate you want	Move questions to the rotating tier or remove them
Respondents pick the middle option for a mixed reason	The scale is hiding a split	Review whether the question should be two
A respondent reverses a scale	The direction is inconsistent	Align every scale to run the same way

Table 5. What a pilot looks for and what to do about it.

WORKED EXAMPLE: FOURTEEN PEOPLE ON A THURSDAY

The Great Lakes chapter meets for a monthly roundtable on the third Thursday. In 2025 the Federation of Food Bank Networks, which has 96 networks and 24 staff, asked the roundtable's 14 attendees to complete its draft member survey and then stay for a 15-minute debrief.

The pilot found three problems. Three of the 14 read "staff" as paid employees only and the others counted volunteers, which would have made the staffing questions meaningless across the membership. The median completion time was 17 minutes against a target of 10. And one question on regional distribution centers made no sense to the two respondents who ran no such facility.

The team added a definition of staff to the question, moved nine questions from the core tier to the rotating tier, and added a not-applicable option for the distribution question. The final survey took a median of nine minutes in the next test. The response rate that year was 58 percent, which the team credited in part to the shorter instrument.

PILOT, THEN FREEZE

After the pilot, make the fixes, and then freeze. Record the final wording, scale labels, order and routing in the register, and note the pilot date and participants. The frozen version is the baseline. If you are tempted to keep improving the questions after this point, remember that every edit made after the pilot is an edit nobody has tested.

- [ ] The pilot group includes someone from the smallest and largest segments.
- [ ] No question writer took the pilot as a respondent.
- [ ] Completion time was measured and is within the target.
- [ ] Each confusing question was discussed with at least two pilot respondents.
- [ ] The frozen version is recorded in the register with the pilot date.

> “A fourteen-person pilot cost us one afternoon and saved us a year of data. Now nothing goes to the full list without one.

RASHIDA KIMBROUGH, CHIEF MEMBERSHIP OFFICER, FEDERATION OF FOOD BANK NETWORKS

6
REPORTING A MARGIN OF ERROR

A three-point rise looks like news until you print the margin of error beside it. By the end of this section you can calculate that margin, judge whether a year-over-year change is real and word the result for a board.

Every State of Associations release reports a margin of error beside its headline figures, and every association survey that expects a board or outside researchers to take it seriously should do the same. The margin tells a reader how much confidence a number deserves, and whether a two-point change in a satisfaction score is a real shift or ordinary sampling noise. Leaving it out does not make the finding more certain. It hides the uncertainty, and members who know survey methodology will notice and discount the finding.

WHAT THE NUMBER MEANS

A margin of error of 4.8 points at 95 percent confidence says this: if you ran the same survey many times with different random samples, about 95 of every 100 results would fall within 4.8 points of the true figure for the whole membership. It does not say the true figure is certainly inside the range, and it says nothing about problems that are not sampling error, such as a badly worded question or a skewed set of respondents. The margin of error is a floor on your uncertainty, not a ceiling.

The size of the margin depends mostly on the number of responses and only slightly on the size of the membership. That surprises people. A survey of 600 responses from a membership of 20,000 and a survey of 600 responses from a membership of 200,000 carry nearly the same margin, about 4 points.

RESPONSES	MARGIN OF ERROR
100	±9.8 points
200	±6.9 points
400	±4.9 points
600	±4.0 points
800	±3.5 points
1,000	±3.1 points
1,500	±2.5 points

Table 6. Margin of error by number of responses, large population, 95 percent confidence, assumed 50 percent answer.

Notice the shape. Going from 100 to 400 responses cuts the margin in half, but going from 400 to 1,500 buys only another 2.4 points. Past about 600 responses, extra responses are worth far less to a headline figure than the same effort spent on reaching a weak segment.

SMALL MEMBERSHIPS: THE FINITE POPULATION CORRECTION

When your respondents are a large share of a small membership, the margin shrinks further, because you have heard from most of the people you are describing. The correction multiplies the margin by the square root of the share of members you did not hear from. If you hear from all of them, the margin is zero, since a census has no sampling error.

WORKED EXAMPLE: IS A THREE-POINT RISE REAL?

The Society of Dental Hygiene Educators has 1,300 members and six staff. In its 2024 survey, 298 members responded and 58 percent said they were confident in the society's direction. In 2025 it received 312 responses and 61 percent were confident.

At the cautious 50 percent assumption, the 2025 margin works out to about 4.8 points once the finite population correction is applied, and the 2024 margin to about 5.0 points. The margin on the difference between two independent surveys is larger than either one, roughly the square root of the sum of their squares: about 6.9 points. A three-point rise is less than half of that.

The honest sentence for the report is: "Confidence in the society's direction was 61 percent in 2025, compared with 58 percent in 2024. The change is within the margin of error, and we do not read it as a shift." The executive

director may wish the sentence said more. It cannot, and members who know the method will trust the next sentence more because of it.

HOW TO WORD IT IN A REPORT

State the margin the first time the figure appears, and again in any chart title or caption that stands alone. Give the confidence level and the number of responses in a note under the figure. Name the date range of fielding, so the reader knows the season.

When a change falls inside the margin, say so in plain words. When it falls outside, say what it is, and do not describe it as dramatic. Avoid the word "significant" unless you mean it in the statistical sense, since many board members hear it as "important".

- [ ] Every headline figure in the report carries its margin of error.
- [ ] Confidence level and response count are stated under each figure.
- [ ] Year-over-year changes are compared against the margin on the difference, not the margin on one year.
- [ ] Segment figures with fewer responses than the publishing minimum are described in words.
- [ ] The word "significant" appears only where a test supports it.

> “Our board stopped arguing about every wiggle once we printed the margin of error beside the number. It gave them permission to ignore noise.
>
> NOOR HADDAD-ELLISON, CHIEF MEMBERSHIP OFFICER, LEAGUE OF INDEPENDENT PHARMACISTS

7
TIMING AND REMINDER CADENCE

The week you field and the number of reminders you send decide who answers. By the end of this section you can lock a window and a schedule, and explain why changing either one changes your results.

Field a tracked survey at the same point in the calendar every year, and lock the window with the questions and the scale. Sentiment moves with the season. A survey fielded right after dues renewal reads differently from the same survey in mid-year. Build next year's schedule around the window instead of fielding whenever staff time allows.

Reminder cadence changes who ends up in your sample. One reminder a week after the invitation reaches members who read email promptly and misses those who get to it on the second or third nudge. A survey that must represent small, thinly staffed members needs more reminders, spaced further apart, to reach staff who check association mail in batches. Document the cadence and repeat it the next year, exactly as you repeat the wording.

PICK A WINDOW THAT AVOIDS YOUR OWN CALENDAR

List the dates that distort answers in your association: the renewal period, the annual conference, the legislative session, the close of a fiscal year. Members who have just renewed rate value differently from members who are about to. Members who have just come home from a conference are warmer than usual.

Choose a window that sits clear of those dates and write it as a rule, not a date. "The first Tuesday of November, for four weeks" survives a change of staff. "When we have time in the fall" does not.

A FOUR-WEEK SCHEDULE

A four-week field with three reminders suits most associations. It gives slow readers three chances to answer without making the survey feel like a campaign.

Table 7. A four-week reminder schedule.

DAY	MESSAGE	NOTES
0	Invitation from the executive director	Personal sender; say how long it takes
7	First reminder to non-respondents	Same sender; short
14	Second reminder, sent at a different time of day	Reaches batch readers
21	Final reminder with the closing date	Name the date and the hour
28	Survey closes	Thank-you note to everyone who was invited

Send each reminder only to people who have not answered. Sending to everyone annoys the responders and inflates unsubscribes. Record the day and hour of each send in the register.

WORKED EXAMPLE: WHOM A SINGLE REMINDER MISSES

The Federation of Community Libraries has 1,900 member libraries, and most employ fewer than ten people. In its 2025 survey it invited all 1,900 and followed the four-week schedule above. The team tracked who answered in each week.

Table 8. Responses by week, Federation of Community Libraries, 2025 survey.

PERIOD	NEW RESPONSES	CUMULATIVE RESPONSES	CUMULATIVE RESPONSE RATE	SHARE OF NEW RESPONSES FROM LIBRARIES WITH THREE OR FEWER STAFF
Days 0 to 7	513	513	27%	18%
Days 8 to 14	209	722	38%	29%
Days 15 to 21	152	874	46%	37%
Days 22 to 28	95	969	51%	44%

Libraries with three or fewer staff made up 18 percent of the first week's responses. Across the full month they made up about 26 percent, because the later waves leaned heavily toward the smallest libraries. A survey that sent a single reminder at day seven and closed at day 14 would have reported 21 percent small libraries, and its satisfaction figures would have leaned toward the larger, better-staffed members.

The team had been tempted to close early in 2024 to meet a board deadline, and it did. The 2024 survey closed on day 14. The small-library share that year was 21 percent, and the change between 2024 and 2025 in the reported figures partly reflected who was in the room.

REACHING THIN-STAFFED MEMBERS

If a segment answers late, give it later reminders rather than more reminders. A reminder sent on a Friday afternoon reaches the batch reader. A reminder from a chapter chair reaches people who know the sender. A short text on a phone or a call from a staff member reaches members who do not open email at all.

Whatever you add for a segment, add it every year. A phone campaign in a good year and none in the next is a change to the design.

- [ ] The fielding window is written as a rule that does not depend on a person's memory.
- [ ] The window avoids renewal, conference and fiscal year-end.
- [ ] The reminder schedule, with days and times of day, is in the register.
- [ ] Reminders go only to people who have not responded.
- [ ] Any extra outreach to thin-staffed segments is documented and will repeat.

> “I answer association email in batches on Fridays, and a four-person staff does not get to every message. The second reminder is the one that got me.
>
> SIGNE THORVALDSEN, EXECUTIVE DIRECTOR, CASCADIA FOREST PRODUCTS COUNCIL

8

COMMON MISTAKES IN A FIRST TRACKED SURVEY

A first tracked survey fails in five predictable ways. By the end of this section you can spot each one and name the cheapest point at which to prevent it.

The first mistake is treating year one as a draft to improve next year, when it is the baseline everything later is measured against. The second is adding tracked questions every year without retiring any, so the survey grows until response rates fall from fatigue. The third is sizing the sample for the whole membership, then learning that the board wanted a breakdown by region or by size, and that the denominator was wrong. The fourth and fifth, covered below, are having no owner for the design and skipping the planning session with the board.

Prevent all five by deciding, before the first field, which questions are tracked, which segments the survey must cover and how long the design stays fixed. A three-year minimum commitment, reviewed only after that period, gives a trend line time to become useful before anything changes.

MISTAKE	WHAT IT COSTS	CHEAPEST PREVENTION
1. Treating year one as a draft	No baseline: year two cannot be compared with year one	Pilot hard, then freeze and record the design
2. Adding questions every year	Longer survey, falling response rate	Ration the core tier; put new questions in the rotating tier
3. Sizing for the whole membership	Segment figures too thin to publish	Name the smallest reported segment before fielding
4. No owner for the design	Every change request is decided by whoever pushes hardest	Name a steward by title with written authority
5. Skipping the board planning session	The board wants a breakdown the sample cannot give	Hold a 30-minute session before writing questions

Table 9. Five first-survey mistakes, their cost and their prevention.

MISTAKE 4: NO OWNER

The fourth mistake sits underneath the others. Nobody owns the design, so every request to change it is decided by whoever shouts loudest in the

meeting. The remedy is the steward described in section 9, a single role with authority to say no.

A steward does not need to be senior. The role needs a clear title, a written list of what is fixed, and the executive director's backing when a board member pushes back.

MISTAKE 5: NOT ASKING THE BOARD FIRST

Many associations field the first survey and then learn what the board wanted from it. This is the fifth mistake. A board that wanted regional results gets a single national figure. A board that wanted to compare large and small members gets a tally that cannot be split.

Hold a 30-minute session with the board, the executive director and the heads of membership and programs before you write any questions. Ask each person to name the decision a result would change, and the group of members they would want the result for. The answers tell you which segments the sample must cover.

WORKED EXAMPLE: A FIRST SURVEY RESCUED IN THE DESIGN STAGE

The Guild of County Records Officers, with 2,700 members and seven staff, planned its first tracked survey in early 2025. The draft held 61 questions and planned a single national figure. The research lead held a planning session with the board, which asked for results by state and by county size.

That request changed the design. The guild dropped the survey to 22 core questions, set a minimum of 30 responses to publish a state figure, and decided that only the nine states with more than 100 members would be reported individually. The other states were grouped into three regions. The survey went out with a fixed wording, a five-point scale, a four-week window beginning the first Tuesday of April, and three reminders.

The April 2025 field drew 1,134 responses, a 42 percent response rate, and the guild could report state figures for all nine large states. It cost one extra planning session. Fixing the same problems after fielding would have meant a second survey.

- [ ] A planning session with the board has named the segments the survey must cover.
- [ ] The core tier is short enough to finish in 10 minutes.

- [ ] The design is committed for at least three years, with a review date in the register.
- [ ] A steward is named by title, with written authority.
- [ ] The first year's version is described to everyone as the baseline, not a draft.

9

THE STATE OF ASSOCIATIONS SURVEY IN PRACTICE

Every practice in this guide appears in one survey you can read for free. By the end of this section you can open the State of Associations methodology section and match each of its choices to a rule in this guide.

NANA's State of Associations survey is fielded each year with Fieldstone Data Partners. Its core question set has not changed since the 2024 launch. Its satisfaction questions use one five-point scale with a word label on every point. The frame is NANA's member organizations, the design is checked against the smallest chapter reporting group, and every headline figure carries a margin of error. The methodology section published beside each report sets out all four, and you can copy its structure for your own first tracked survey.

HOW EACH PRACTICE SHOWS UP

New questions arrive in a separate block each year and are labeled as new, so no reader mistakes a first-year number for a trend. The research team keeps the instructions Fieldstone follows in a single register, which is the register you will build from the appendix. The frame is NANA's member organizations, 750 or more, and the survey invites all of them. Chapter figures describe the responding member organizations in each chapter.

DESIGN ELEMENT	PRACTICE IN THE STATE OF ASSOCIATIONS	SECTION
Question wording	Core set fixed; new questions added in a labeled block	2
Sample size	Invites every member organization; checked against the smallest chapter reporting group	3
Response scale	Five-point scale for satisfaction questions, fixed across years	4
Margin of error	Published beside every headline figure	6
Calendar	Fixed fielding window; report released each March	7
Ownership	NANA's research team assigns the steward role	9

Table 10. The State of Associations design mapped to this guide.

THE ARITHMETIC BEHIND THE CLAIM

The 2024 edition drew 690 responding organizations and the 2025 edition drew 720. Against a frame of 750 member organizations, that is 92 percent and 96 percent. Membership is 750 or more, so the true rates are a little lower, but the survey is close to a census either way.

A near-census changes what the margin of error means. At the cautious 50 percent assumption, the finite population correction cuts the margin on a percentage from all 720 responses to about 0.7 points, and from all 690 responses to about 1.1 points. The margin on the difference between years is about 1.3 points. With margins that small, the live risk is nonresponse: the 30 to 60 member organizations that did not answer may differ from those that did. The report therefore states the response rate beside the margin.

The same arithmetic answers the chapter-level question from section 3. That section showed a 40-member group needing 37 responses, a 93 percent response rate, to be reported within 5 points. The State of Associations response rate is already at that level, so a small chapter group is reported from a near-census and not from a thin sample. Most associations run a survey at 40 to 60 percent, and for them the segment arithmetic in section 3 is the constraint to plan around.

THE PRESSURE TO CHANGE SOMETHING

The hardest discipline in tracked research is saying no every year. A board member dislikes a question's wording. A new researcher wants to modernize the scale. A committee wants one more segment broken out. Each request is reasonable alone, and granted often enough, the requests erode the comparability the survey exists to produce.

The remedy is procedural. Appoint one person, by title and not by name, as the survey's steward, with written authority to refuse a change until the scheduled design review. At NANA the research team assigns the steward role, which gives every change request one point of accountability across both editions.

HOW THE STEWARD HANDLES A REQUEST

A request arrives in writing with the requester's reason. The steward logs it in the change log (see the appendix) and answers within 10 working days. The answer is one of three: the request is declined because it would break a tracked series, it is accepted into the rotating tier where it changes nothing

already tracked, or it is held for the scheduled design review, with a note on what evidence would justify it.

Most requests land in the second outcome. A board member who wants a question about a pending vote can have it, in the rotating tier, for the year the vote matters. The same member cannot have last year's satisfaction question reworded because it reads awkwardly.

> “
> I will accept a survey I disagree with. I will not accept one that changes its questions whenever it disagrees with me.
>
> HARUKO TANAKA, TREASURER, ALLIANCE OF REGIONAL CREDIT UNIONS

10
WHERE TO GO NEXT

Pick the next step by what you are doing. By the end of this section you know which NANA course, report and event fits a first survey, a survey you are reading, or a research habit you want to build.

If you are designing a first tracked survey, take Survey Design for Associations, the NANA course that builds on this guide. It works through the register in the appendix with your own questions. If your survey will feed an engagement model, the Member Engagement Scoring course, taught by Mei-Lin Harbaugh and Declan Moorfield, shows how survey answers combine with behavior data.

For a comparison point, read the State of Associations report, released each March, and join the member webinar that follows it. The Staff Compensation Survey, released each June, is a second example of a fixed design that has now run twice. If your survey touches dues, the Dues Model Study led by Haruko Tanaka is the model for a defined sample, and the kickoff webinar recording shows how the research team frames it for members.

To meet peers who run member research, bring a question to the Data-Driven Engagement Workshop or to a chapter roundtable, where a 14-person pilot is easy to arrange. The Membership Growth Summit in Atlanta includes sessions on measurement.

APPENDIX: THE FIXED-DESIGN REGISTER AND CHANGE LOG

This appendix gives you two templates to copy into your own files. The register is the single page that lists everything fixed about your survey. The change log records every request to alter it and what happened.

THE REGISTER

ELEMENT	WHAT TO RECORD	EXAMPLE ENTRY
Survey name and baseline year	Name, first year fielded	Member Value Survey, 2025
Population and frame	Who is invited, from which list, as of which date	All 1,900 member libraries on the March 1 membership list
Core question, label and wording	One row per tracked question, exact text	MV1: How satisfied are you with the value your organization receives for its dues?
Response scale and labels	Every point, in order, in words	Very dissatisfied, dissatisfied, neither, satisfied, very satisfied
Reference period	The period each question covers	The last 12 months
Fielding window	A rule, not a date	First Tuesday of October, four weeks
Reminder schedule	Days, times of day, sender	Days 7, 14 and 21; second reminder at 3 p.m.
Smallest reported segment and minimum	The segment and the publishing threshold	Libraries with three or fewer staff; at least 30 responses
Weighting	Method, or none	None
Steward and review date	Title and next scheduled review	Research lead; review after the third edition

Table 11. Fixed-design register template, with one example row for each element.

THE CHANGE LOG

Keep one row for every request, including requests that are declined. A declined request recorded with a reason prevents the same argument from returning next year.

Table 12. Change log template, with an example entry.

DATE	REQUESTED BY	REQUEST	REASON GIVEN	DECISION	REASON FOR DECISION
Nov 4	Board member	Reword MV1 to name the new portal	Members keep asking about it	Declined for MV1; added as a rotating question	Rewording would break the series; the rotating tier holds timely topics

A ONE-PAGE CHECKLIST BEFORE THE INVITATION GOES OUT

- [ ] The register is complete and the steward has signed it.
- [ ] Every core question passed the six rules in section 2.
- [ ] Required responses are calculated for each reported segment.
- [ ] Every scale point has a word label and the tool's default has been overridden.
- [ ] The pilot is done, the fixes are made and the design is frozen.
- [ ] The report template has a place for the margin of error beside every headline figure.
- [ ] The fielding window and reminder schedule are in the calendar of the people who will send them.

GLOSSARY

Confidence level. The share of repeated samples whose result would fall within the margin of error of the true figure. This guide uses 95 percent.

Core tier. The set of questions tracked across years and never edited.

Finite population correction. An adjustment that shrinks the margin of error, or the sample you need, when the membership is small.

Margin of error. The range around a survey figure that reflects sampling uncertainty alone.

Rotating tier. The set of questions that may change each year because nobody expects to compare them across years.

Split sample. A test in which random halves of the respondents see different versions of a question, used to measure the effect of a change.

Steward. The person, named by title, with written authority to refuse changes to the fixed design.



The National Association of National Associations is the professional society for association executives, boards, and staff, serving more than 750 member organizations and 2 million professionals through research, peer community, and advocacy since 1998.