



TRACKER TRANSFORMATION

MANAGING CHANGE IN TRACKERS AND NORMED STUDIES
CHAPTER 3 | MANAGING THE METHODOLOGY



CHAPTER 3: MANAGING THE METHODOLOGY



In chapter 1, we discussed the undercurrents of changing consumer behavior that are threatening trackers and normed studies. In chapter 2, we set out a proven project management framework for structuring work, setting expectations, defining success measures, and — perhaps most importantly — providing critical visibility about what’s going on. In this chapter, we’ll discuss the important research issues to consider.

How likely is it that my data will change?

We will ask this question from a political point of view in chapter 4, but for now let’s consider it from a research perspective.

There are two ways to think about this question. One is that any change, as small or insignificant as it may seem, should be expected to change the data. We can hope otherwise, but it’s not a recognized component of a solid research methodology. People should thus be prepared for change, and the extent to which you prepare them for it is essential to successfully navigating the transition (more on this in chapter 4).

From a research point of view, though, the question of degree isn't terribly important. If we assume that there is a high likelihood of change, it's far more important to understand *what* has changed and *why* it is changing.

THREE FACTORS THAT CAN CAUSE NORMS TO SHIFT:

1. Changes in questionnaire format
2. Changes in sample composition
3. Changes in questionnaire content

Recall that, in chapter 1, we made the case that change is already afoot and appearing in your data. Thus any conversation about change from a research point of view needs to consider the existing data as having its own flaws or weaknesses. Keep these in mind as you consider the points below.

There are three large factors that will potentially create a trend break or cause norms to shift.

1: Changes to the questionnaire format

If your tracker is being adapted to smaller screens, it is very likely that the format of your questions or responses, i.e. how they are displayed on the screen, will need to be adjusted to fit smaller screens. Touch screens especially create new ways for respondents to physically interact with a study.

For simple questionnaires with single- or multi-punch responses, the odds of a major data change should be small. The typical radio buttons or check boxes we see in online questionnaires are completely analogous for small screens. More complex questionnaires will naturally incur greater risk. The grid question should be first on the firing line as it has indisputably proven to be poorly understood by respondents despite its large-screen efficiency. Likewise, changes to scale questions, particularly the use of sliders (which often enable the respondent to provide more finely-grained responses, and thus change the variance of responses) or the inversion of horizontally-arrayed scales to vertically-arrayed ones, may cause shifts in data.

There are two things to take away from any changes in format. One is that you do not need to become an expert. Any decent provider will have a point of view on how to format and display different question types regardless of screen size. Survey research platforms have this

built into their logic. The other is that you shouldn't go into the redesign process from a point of view of risk-aversion to proven techniques. The best course of action is to do a parallel test to determine whether the data will change. We discuss the parallel test in detail below.

2: Changes to the sample composition

There are three ways sample composition can change in a tracker. From least to most consequential, these are (1) changing demographic composition, (2) changing sources, and (3) changing the sampling frame.

Changing a sample's demographic composition - that is, changing the proportions of different types of men vs. women, young vs. old, etc. - is usually manageable through weighting. Weighting is a recommended practice for any tracker, especially if there are characteristics of a sample which are not easily managed in field that are correlated with outcomes of interest. Weighting is not a panacea, especially if the raw sample composition is very different from the target population, but it is easy to ensure the sample is well-aligned with the desired target beforehand.

It is a poorly-understood fact that the sources from which panel companies recruit respondents can and do change over time. It is thus hardly a guarantee that using the same panel company over time ensures the same composition of respondents in terms of their demographics or data quality.

Changing sample sources entails greater risk. Whether you are using a research panel or some other source of respondents, there are plenty of differences in people's behavior that aren't fully captured by standard demographics like age, gender, or household size. Ordinarily, we would not expect big changes — unless there is something about the panel that is correlated with the subject of the tracker. For example, there are many loyalty communities around mobile gaming that offer members game currency for taking surveys. If you are running a study about mobile phone use, you may see differences in behavior. Nevertheless, it is a poorly-understood fact that the sources from which panel companies recruit respondents can and do change over time. It is thus hardly a guarantee, that using the same panel company over time ensures the same composition of respondents in terms of their demographics or data quality.

The third way a sample can change is if there is something about the execution of the study that implicitly changes its sampling frame. In practical terms, opening a study up to completion on a mobile phone invites a large group of people who have foregone participation on desktops, and thus represents the biggest risk to trackers that decide to go mobile. People who are smartphone dominant especially have different behaviors when it comes to shopping and media consumption than those on desktops. They are also, for most brands and retailers, a very desirable subpopulation to understand.

3: Changes to the questionnaire content

Whether it's to reflect the evolution of the market, or behavior you're studying, or to simply streamline and reduce cost, changes to questionnaire content can have a big impact on the data. These types of changes are frequently underestimated in terms of their impact: a single word can literally change a market.

EVERY QUESTIONNAIRE CAN BE SEPARATED INTO THREE COMPONENTS:

1. The essential questions
2. The interesting-but-not-critical context questions
3. The demographic questions

Most tracking questionnaires can be broken down into three parts. In usual order of appearance, these are:

1. **The essential questions**, meaning those essential for calculating the tracked metric, starting with the qualifying question that usually takes the form: Did you do <activity> during < some time period>? These are usually followed by questions about who did the activity and where it was done.
2. **Interesting but not critical contextual questions**, which describe or provide additional context about the essential activity being tracked, yet aren't essential to the counting. What were you doing beforehand? Who were you with while you were engaging in our behavior of interest?
3. **The demographic questions**, which should at minimum include age, gender, household composition, a geographical measure (urbanization or region), and a measure of purchasing power (income, employment status, etc).

When thinking about changes to the questions themselves, it's best to think about what you're trying to achieve. Ask yourself the following questions:

Do you need to shorten the tracker?

Trackers are like closets. They start out tidy and organized, and over time become messy and overcrowded, piled high with yesterday's fashions and toys that the kids don't play with any more. It's important to clean them out every once in a while. This is especially true if you notice respondents are abandoning in greater rates, participating in lower rates, become mobile-friendly, or if you're trying to reduce costs.

Trackers are like closets. They start out tidy and organized, and over time become messy and overcrowded, piled high with yesterday's fashions and toys that the kids don't play with any more.

Start from the end and work backward. Eliminate as many demographic questions as possible if you have data as part of a respondent profile. Notwithstanding the fact that most sample providers have this data on file, researchers still reflexively ask these questions to the great annoyance of respondents.

Challenge yourself to justify the interesting but not critical questions. Figure out who is using them and why. Are they still relevant? Is someone **using the data** to make decisions? Does the insight they provide change much over time? If not, ask them less frequently. Be careful, though, if these appear before the essential questions, as they can have a conditioning impact on responses.

The last things you should touch are the essential questions, for what should be obvious reasons. That said, there are two cases where this is important: marketplace changes, and respondent understanding.

Do you need to better reflect the market?

While the pace may be different, markets evolve naturally over time. Categories grow and shrink; brands come and go. While we generally seek to avoid change in a tracker, we must accommodate it where there is a real change afoot. For market share, and brand-awareness tracking

especially, lists that describe the structure of a market by brand or category (or any other relevant factor) should be updated with some regularity. Once a quarter is probably sufficient, unless the market is changing more quickly.

Do respondents understand and complete the survey as expected?

At its core, survey research is a challenging endeavor. Depending on the subject matter, it can be difficult to get respondents to understand your question, summon the appropriate response, record that response, and then want to continue on. This is particularly the case when we ask consumers to recall distant, non-remarkable, or even (socially-scorned) behavior. If you believe respondents are failing to do one of these things, then it may make sense to change the questionnaire more substantially. This change requires very careful consideration and is best served by qualitative or non-conscious approaches to create greater understanding.

Sample size or tracker frequency

Sample size and tracker frequency, are important to consider as they speak to the great tradeoff between readability of the data and cost.

Classically-trained researchers know that the smaller the sample, the noisier the data. Trackers are generally built to have sample sizes sufficient to read changes in the smallest area of interest. Yet even with an ideal questionnaire (which is hardly common) and a probability-based sample (which is achievable only at a very high cost), a tracker may exhibit disconcerting levels of volatility.

There are solutions, though. One is to define the lowest 'safe' level of granularity for reporting to restrain the urge to read the data more deeply than their explanatory power will allow. This can be expressed in simple rules like, "Don't look at anything with a brand share below 1%." Another is to experiment with rolling up data over multiple waves, which has the same stabilizing effect as increasing sample size at a lower cost, and may be justified depending on how often people change their behavior.

At some point the users of the data need to be made to understand that even the most 'scientifically managed' online study has volatility. We recognize this is easier said than done, but the change process will allow you to re-open this discussion. It may be that a survey isn't the best tool for a particular measurement need. Or, more likely, it will be necessary to revisit and retrain people on how data should and should not be used.

Testing

As mentioned above, the answer to the question “Will the data change?” isn’t nearly as important as knowing that, if it does, you’ll be able to witness it and provide plausible explanations. Some change may even be desirable. What we want to avoid, though, is the kind of change where people scratch their heads and say “Gee, we never saw that coming.”

Think aloud pre-testing

David Bakken, of the Foreseeable Futures Company, and methodology expert, has a valuable technique he calls “think aloud pre-testing.” As the term implies, this entails quite literally talking oneself through the tracker, reading each question aloud, and saying out loud, exactly what comes to mind for each and every question. The deliberate and audible nature of the exercise is guaranteed to expose problems of comprehension, flow, and format in ways that silent skimming will never do.

- **Will the layperson understand the question, or does it use unfamiliar jargon or assume industry knowledge?**
- **Will s/he be able to summon an accurate response, or is it something s/he never noticed or forgot?**
- **Will her/his response fit within the choices on the page?**

Think aloud pre-testing entails quite literally talking oneself through the tracker, reading each question aloud, and saying out loud, exactly what comes to mind, for each and every question.

Resist the urge to do this only with internal team members who are likely fluent with the industry, its jargon, and its context. If necessary, draft friends and family for a test run using the always-popular reward of free lunch.

Full parallel testing

A full parallel test is a **must** for any tracker change. It is a rigorous side-by-side experiment that will allow you to validly compare data from the old tracker against the new design. It is the only way to get a thorough understanding of the changes.

A full parallel test typically proceeds in the following manner:

- **Decide on your test questionnaire.** Your control questionnaire will be the existing tracker. The test will be the new version (or versions, if you are contemplating multiple changes) that is as close to final as possible. Be careful if you are planning changes to both the questionnaire and the sample. You should separate these into sequential tests (first test: old vs. new questionnaire; second test: new questionnaire with old sample vs. new questionnaire with new sample) to maximize your ability to understand changes.
- **Pull one big sample, then split it.** Sequential samples carry too much risk when doing full parallel tests, even if the second is pulled immediately after the first. Pull one giant sample, then randomly assign people to the control or test design. Do the statistics beforehand so that you know the margin of error you'll need to exceed to detect significant differences. Don't skimp on this part, even though this is where real dollars (or pounds, euros, or whatever) will be spent.
- **Field the studies exactly the same way.** It's not more complicated than that. Make sure the studies are treated identically in field. Start and stop them at the same time.
- **Process the data exactly the same way.** Tracker changes help expose ghosts in the machine, particularly any aspects of data coding, editing, or weighting that happen after data collection and before reports are generated. Make sure you've understood the current processing steps and faithfully replicate them (or change them in a controlled fashion) for the new study.

In any parallel test, make sure you process the data and produce output just as you would with the live study. This will expose any 'ghosts in the machine.'

- **Produce full output, including field stats.** Examine each and every data field to ensure proper response capture and coding. In the day when humans created import scripts and had to parse the data, this was a major undertaking. These days, with automated systems that instantly and flawlessly parse and display data, it's a lot easier. Likewise, compute and compare the statistics from the field, like raw incidence, completion rates, and length of interview.
- **Do a complete impact analysis.** Line the control data up against the test data, question by question, and look for differences. Use existing reporting templates where possible as this will be most

people's anchor to truth. Do this step very openly and publicly. Discuss why the differences might have arisen and what that may mean for people using the data. This step is essential to socializing the new 'truth' that you will create with the new tracker design.

What to do if data changes

There are really only two things you can do if the data change. One is to live with the change. The other is to try to adjust the data somewhere to minimize it.

If you and your stakeholders are happy with the changes and can rationalize them in ways that make a sensible new truth, then it makes sense to accept the changes and move forward.

Walking away from trends or norms can be a very difficult thing, though. For this reason, it's not uncommon to see companies try to adjust, either the new data (sometimes called 'backcasting') or the old data (sometimes called backcasting) to bring the two into alignment.

Aligning the data is a very tricky thing to do and never works perfectly for what should be obvious reasons. The whole point of changing a tracker is to amend something that isn't right in the existing data in the first place! The data will never perfectly align. Moreover, the practice of trying to mold the new data to fit the old (which, in our experience, is the most common practice, as it minimizes the collective effort needed to assimilate the new truth) is inherently dangerous. Forcing a system that you've spent time, money, and energy to produce a result that it doesn't produce naturally, means you will need to keep forcing it, wave after wave, to produce an unnatural result. If you have no choice and you need to preserve the trend, the best solution is to refactor the 'old' data to align with the new data. Your parallel test should give you insight as to how to accomplish this.

We reiterate here the importance of socializing the data — of openly, publicly, and frequently discussing and rationalizing what has changed and why. Those who are unfamiliar with consumer research

methodologies are going to have difficulty understanding how there can be two versions of truth. But the more you tell the story, the more everyone (including you!) will internalize and accept it.



Conclusion

While the details of your particular tracker or normed study will vary, this chapter should give you a framework for considering and testing the methodology changes you're contemplating due to broader shifts in consumer behavior. At the end of the day, a good tracker methodology—like any research design—is one that is simple, relevant to the subject matter, thoughtfully planned, and well-executed.



For more information, please get in touch with us
support@aytm.com | +1 415-364-8601 | aytm.com © 2016

