The Baseline Playbook: What Works (and What Does Not) in Commercial Existing-Building Studies

Nathan Caron, Erika Page, and Geoffrey Cooper, DNV

ABSTRACT

Commercial existing-building baseline studies provide critical inputs for program planning, evaluation, and savings assumptions, but they have lately become harder to execute. Recent studies have faced low response rates, stale or incomplete contact information, difficulty reaching knowledgeable respondents, and limited willingness to participate in technical interviews or site visits. These challenges do not mean commercial baseline studies are no longer feasible. They do suggest, however, that designs focused too narrowly on statistical targets can struggle in practice when frame quality, respondent routing, and recruitment requirements are not treated as core design issues.

This paper reviews four recent commercial baseline efforts in the Northeast and Mid-Atlantic to identify design choices that affected feasibility and interpretability. The studies used parcel-based, utility-premises-based, and multi-source frames; mixed-mode recruitment; and, in several cases, tiered data collection through surveys, technical interviews, and on-site or virtual verification.

The paper is intended for evaluators, utility and program administrator staff, and technical reviewers responsible for designing or reviewing commercial baseline studies. Across the four cases, the key lesson is that baseline studies differ not just in sample design but in how teams build the frame, reach knowledgeable respondents, structure follow-on data collection, and report results. By comparing those design choices explicitly, the paper offers a practical framework for selecting among different baseline study approaches while making tradeoffs in rigor, burden, and feasibility easier to understand.

Introduction

Based on the authors' recent project scoping discussions, many commercial baseline studies are now expected to support not only traditional baseline estimation but also planning questions related to electrification and changing end uses. Recent field experience has also shown that general-population commercial recruitment can be difficult, particularly when sample frame information is dated or the sampled contact is not the person best positioned to answer technical questions.[1] Under those conditions, choices about frame construction, respondent routing, and tiered follow-on data collection affect what can realistically be measured, how representative the achieved sample will be, and how results should be interpreted. This paper reviews four recent studies to document those choices and the tradeoffs they created.

This paper is written for evaluation teams, utility and program administrator staff, and stakeholders who design, fund, or review commercial existing-building baseline studies. It is

[1] In the context of sample design, a sample frame is the operational version of the population: the list of units that can actually be sampled for study. In the studies discussed throughout this paper, frames included tax-parcel records, utility billing or premises records, and multi-source files that combined utility, parcel, and third-party business data.

intended to help those readers weigh competing priorities, including defensible population estimates, useful technical detail, reasonable respondent burden, and field procedures that remain workable under current recruitment conditions.

The baseline studies presented in this paper followed a process similar to that illustrated in Figure 1, though their sample frame development and recruitment stages varied. We do not intend to offer a single best design but rather to provide a decision framework with clearly articulated tradeoffs. This framework enables future baseline teams to select an approach consistent with objectives, constraints, and field conditions. The paper organizes the comparison around five design choices that repeatedly shaped study feasibility and interpretation:

- Defining the sampling unit and building the frame
- Stratifying for both statistical precision and recruitment feasibility
- Designing mixed-mode recruitment and respondent routing
- Deciding whether to use tiered data collection
- Reporting weighted results, dispositions, and domain estimates clearly

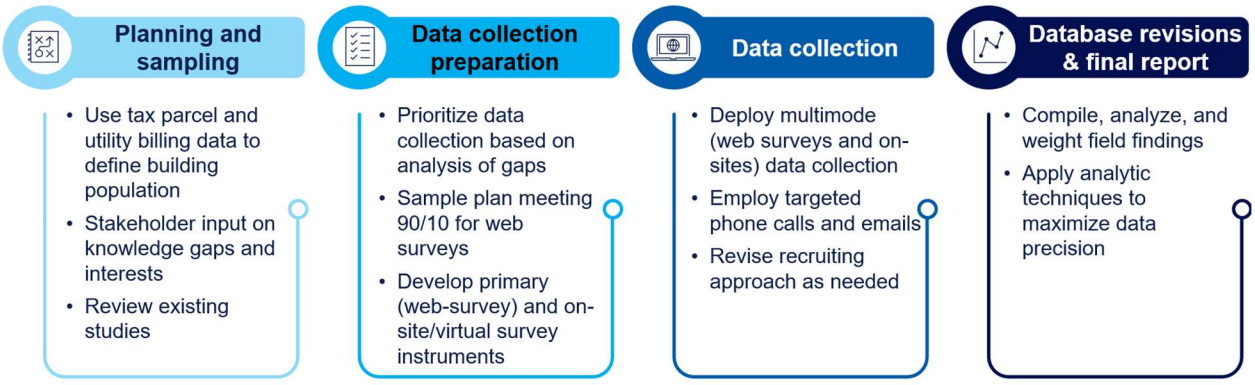

Figure 1. Example of C&I existing building baseline study process flow. Source: Authors' synthesis of the four study designs reviewed in this paper.

Study Set and Comparison Plan

The study set spans three frame types often used in commercial baseline work: parcel-based frames (two Massachusetts studies), a utility billing frame (Connecticut), and a multi-source frame (New Jersey), combining utility data with third-party business/building sources. Table 1 provides an overview of the studies. To keep the paper actionable, we limit comparisons to process dimensions that drive feasibility and interpretability: sampling unit alignment; frame assembly, de-duplication, and quality controls; stratification and target setting; recruitment cadence, respondent routing, and disposition management; follow-on conversion for higher-rigor tiers; and reporting conventions tied to weighting.

Table 1. Overview of Study Purpose, Geography, Frame Type, and Sampling Unit Across the Reviewed Baseline Studies

Study	Purpose	Geography	Frame Type	Sampling Unit	Stratification
MA C&I Baseline	General population baseline study	Massachusetts	MA statewide tax data	Tax parcel	Building type (CBECS) and size
CT C&I Baseline	General population baseline study	Connecticut	Utility premise consumption data	Premises	Building type (CBECS) and size
MA Large Building Electrification	Large building baseline study	Massachusetts	MA statewide tax and billing data	Tax parcel	Building type (CBECS) and size
NJ C&I Baseline	General population baseline study	New Jersey	Utility premises consumption data overlaid with tax parcel data	Premises	Building type (CBECS) and size

Playbook Dimension 1: Define the Sampling Unit and Build the Frame

Differences Across Studies

In the Massachusetts studies, teams treated the parcel as the sampling unit, using assessor attributes (tax use category and square footage) as core frame variables. The Connecticut baseline study defined the sampling unit as a premise (business at an address) and labeled it as a "business" in the report. In the New Jersey study, the team collapsed accounts to unique street addresses after geocoding and address standardization, then appended building type from multiple sources to build a premise-level frame. These unit choices are not interchangeable because they determine (1) which respondent should answer, (2) how shared systems and multi-tenant sites appear, and (3) how teams interpret coverage.

Tradeoffs Observed in Practice

Frame choice creates practical tradeoffs because no single data source captures building use, energy consumption, ownership, tenant occupancy, and knowledgeable contacts equally well. The best frame therefore depends on what the study needs to estimate and who can provide the required information.

Parcel-based frames

Parcel-based frames support consistent building-use and square-footage stratification when assessor attributes remain up to date. In contrast, third-party business-listing datasets often provide an industry sector (NAICS code) for business activity rather than building use, but these datasets may also offer useful size proxies such as location sales or number of employees. Parcels encompass entire buildings, and they more reliably capture central systems separated onto different utility accounts from other energy consumption. However, parcel data typically offers only the building owner's contact information, which may not identify the most

knowledgeable equipment respondent. In the Massachusetts study of large C&I customers, the team encountered problems with consistent analysis units, especially when a parcel included several buildings used by the same customer for different purposes (DNV, 2025b). Outdated information and its coverage can also cause issues when municipalities do not frequently update and share tax data. Massachusetts and New Jersey collect and publish municipal data centrally, so this was not a problem. Not all states do so, which may affect whether teams can use tax data in a baseline study.

Billing-based frames

Billing-based frames provide a direct link to energy use (a practical proxy for size stratification) and reflect the population of interest for utility stakeholders; however, these frames have drawbacks. First, teams may spend substantial time and effort requesting, receiving, and standardizing these records. During this effort, listed businesses may have closed or moved, causing a lower response rate if teams address recruitment communications to the wrong business. Second, the listed customer is the billpayer, who may not be the correct contact to provide information on building attributes. For some businesses, a third-party vendor may manage the business's utility bills, but that vendor is entirely unrelated to the business. Third, if teams sample from customer data in multi-tenant buildings, they may fail to reach contacts responsible for central systems, which only an owner or property manager may access. Finally, unless utilities append and store building-use information in their systems, teams will need to append the information from a third-party source, likely tax data or a business information vendor.

Multi-source frames

Multi-source frames can improve contact quality and building metadata, but they increase the reconciliation burden and require teams to establish explicit rules for deduplication, source priority, and treatment of conflicting attributes to preserve defensibility. For example, in the New Jersey study, the team used billing data, but the utilities did not store building use or business type for those accounts (DNV, 2024a). To enable stratification, the team appended information from tax data, a third-party commercial real estate database, and a third-party business listing database to achieve the greatest coverage possible. Additionally, for the largest 200 premises by electric consumption (critical for collecting information on the greatest amount of energy consumption), the team manually confirmed the building use, since experience from prior studies indicated that the largest consuming premises were sometimes not eligible, being things like manufacturing or reflecting non-building energy use. This process helped the team avoid wasting resources on recruitment and data collection.

Playbook Guidance

The following guidance reflects the observed tradeoffs across the study set:

The sampling unit should be selected based on the systems the study needs to characterize. If shared systems, central plants, or whole-building attributes are in scope, a building- or parcel-oriented unit will usually be easier to interpret than an account-oriented unit. However, that choice also requires explicit respondent-routing procedures and clear rules for handling parcels that contain multiple buildings or multiple uses.

Teams should determine early how they will stratify the population by size and use. Tax data can provide square footage and building-use information where it is available and current, while billing-based frames provide consumption as a useful proxy for size and program relevance. In many cases, billing records will still require enrichment with building-use and contact information before recruitment begins.

When no single source provides adequate coverage, a multi-source frame may be justified, especially for large or strategically important sites. In those cases, the study should document source-priority rules, de-duplication procedures, and how conflicting attributes were resolved. The same documentation should identify known limitations in frame recency, contact quality, and building-type classification, since those limitations can affect recruitment and analysis.

Playbook Dimension 2: Stratify Sample for Recruitment Feasibility and Statistical Efficiency

Differences Across Studies

Across the study set, sample stratification served two purposes: improving statistical efficiency for priority domains and managing feasibility for strata known to be hard to recruit. The MA C&I baseline study revised completion targets after a soft launch when initial response rates were lower than expected, demonstrating that target-setting must remain adaptive (DNV, 2024b). The Connecticut study found that recruiting buildings in the largest strata would not be successful using the email and postcard methods aimed at smaller businesses, requiring more resource-intensive efforts to identify the correct contact at the facilities, including phone recruiting and support from utility contacts (DNV, 2025a). The other Massachusetts study and the New Jersey study avoided a mid-study change, with upfront plans to conduct more extensive research on contact information for large facilities and to make repeated recruiting phone calls as needed (DNV, 2025b, 2024a).

Common Challenges

A common challenge across the studies was defining a stratification scheme that appeared statistically reasonable on paper and satisfied stakeholder requests but later proved unfeasible in the field after recruitment began. This resulted in fewer completions than desired in the strata that matter most for program impact (often large and complex sites). Another challenge was over-stratification, designed to meet stated study goals but resulting in a final sample that could not support stable stratum-level estimates, requiring post-stratification. A third challenge was not effectively communicating upfront the various levels of precision that might be expected for metrics reflecting a range of variability or that apply only to a subset of the sample, given that the studies sought to capture information across many metrics. Precision targets during planning assumed one or two levels of moderate variability and applicability to the entire sample, leading stakeholders to see wider-than-expected margins of error at a given sample size for certain metrics.

Playbook Guidance

Based on the observed challenges across the studies:

Stratification should account for recruitment feasibility as well as statistical efficiency. Large and complex sites often drive energy consumption, savings potential, or policy relevance, but they may require a different recruitment approach than smaller businesses. For that reason, teams should treat large-site coverage as a dedicated operational design problem, not simply as another allocation cell.

A soft launch can help test whether planned targets are realistic before the full field effort begins. If early response patterns show that certain strata are underperforming, the team should revisit targets quickly, document any changes, and report those changes transparently. This is especially important when stakeholders expect stable estimates for priority segments.

Precision planning should also distinguish between metrics that apply to the full sample and metrics that apply only to a subset of respondents. During planning, teams should identify which metrics the precision targets are intended to cover and explain why some estimates may have wider margins of error than a generic sample-size calculation suggests.

Playbook Dimension 3: Plan for Mixed-Mode Recruitment to Maximize Response Rates

Differences Across Studies

All studies used mixed-mode outreach, but they differed in how respondents were routed to the correct instrument and how large sites were handled. The CT C&I baseline study used postcards, repeated emails, and phone outreach, with utility assistance for very large sites, and documented that targeted recruitment for large sites required substantial staff effort (DNV, 2025a). The MA C&I baseline study, focused on the general C&I population, began with owner postcard outreach and then routed to tenants/managers using appended contact data and targeted calls, with the initial survey as a warm recruitment gateway to technical n-depth interviews (IDIs) and on-site visits (DNV, 2024b). The other MA study, focused on large C&I customers, used a targeted engagement strategy that relied on the deliberate identification of knowledgeable site contacts who could support detailed discussions of building systems (DNV, 2025b). The NJ C&I baseline study emphasized pre-field steps intended to improve recruitment yield, including recency filters, assessment of contact fill rates, manual screening of large-consuming customers to exclude non-building (ineligible) records, and third-party contact enrichment (DNV, 2024a).

These mixed modes contributed to increased response rates for various segments of the target population in each study; however, as illustrated in Table 2, three of the four studies still experienced generally low response rates, particularly within specific targeted segments. The comparative insights provided by Table 2 emphasize that while all studies adopted mixed-mode outreach, overall response rates remained modest: both the MA C&I Baseline and CT C&I Baseline studies demonstrated comparatively low performance due to issues identified under Playbook Dimension 1. For the MA study, employing tax-parcel as the sampling dimension was suitable for capturing information about central systems, but only if the study could successfully reach building owners and appropriately route respondents to tenants or managers, which frequently proved challenging. In the CT study, frame recency emerged as a critical limitation: in order to minimize sample development costs and accelerate recruitment, the study was provided with a pre-existing billing dataset, but this data was several years old (collected pre-COVID). Consequently, although most premises still existed, the business occupants and contact

information were often outdated, negatively impacting response rates, as highlighted by the relatively low figures in Table 2. In contrast, the NJ study achieved a response rate nearly double that of the MA and CT studies, by using a dataset compiled only a few months prior to recruitment, thus ensuring greater accuracy of contact information. Notably, the MA Large Building Study stands out in Table 2 with a high 21% response rate, attributable to its exclusive focus on large buildings and the consistent application of resource-intensive recruitment methods for all sites, which proved more effective than limiting such methods to only the large-stratum sites, as was the case in the other studies.

Table 2. Comparison of overall response rates

Study	Population	Target	Completes	Response Rate
MA C&I Baseline	147,820	1,200	868	1.4%
CT C&I Baseline	109,856	416	286	1.8%
MA Large Building Electrification	1260	112	126	21%
NJ C&I Baseline	94,538	420	596	2.8%

Playbook Guidance: Areas for Improvement and Recruitment Design that Worked Across Studies

Areas for Improvement

Table 3 and Table 4 below show segment-level response rates for the MA C&I Baseline and NJ C&I Baseline studies. These tables illustrate that additional targeted outreach methods may be needed for certain segments of interest, beyond the mixed-mode approach used in the studies. For example, the MA study achieved very few completions among large grocery stores, which were mostly regional or national franchises. The study found that, even with high-touch engagement with local management contacts, these contacts would direct recruiters to corporate decision-makers who would then not respond despite repeated outreach. This was similar to the experience with data centers in New Jersey, which are becoming a key building type of interest due to their high energy consumption. The authors do not have a definitive solution to the problem of low response rates among small-population segments substantially comprising local establishments with non-local parent companies; however, we suggest investigating segment-specific response rates from any relevant prior studies to inform discussions with stakeholders of where this might occur, and to pre-plan additional possible avenues for recruitment in those specific segments.

Table 3. MA C&I baseline study completions by segment

Building Use	Population (tax parcel)	Web survey target	Web survey completes	On-site/IDI target	On-site/IDI completes
Education	2,660	74	30	18	8

Building Use	Population (tax parcel)	Web survey target	Web survey completes	On-site/IDI target	On-site/IDI completes
Food Sales[2]	271	30	4	8	2
Food Service	4,642	74	59	18	14
Health Care	2,705	74	54	20	20
Lodging	13,285	74	75	18	14
Manufacturing	4,862	74	83	20	15
Mercantile	14,967	78	59	18	10
Multifamily	42,074	74	66	18	7
Office	22,952	78	69	20	19
Other	4,104	74	37	18	8
Public Assembly	4,420	74	72	18	15
Public Order and Safety	1,970	69	22	18	4
Religious Worship	3,496	64	61	16	14
Service	7,561	74	75	18	10
Unknown	5,512	67	28	18	6
Warehouse and Storage	10,314	74	62	18	11
Non-Building	2,025	74	12	18	2
Total	147,820	1,200	868	300	179

Table 4. NJ C&I baseline study completions by segment

Building Type	Number of Size Strata	Population (premises)	Sample Size	Completes
Data center	3	595	10	3
Education - college	4	26	7	79
Education - other	6	999	29	
Food service - full service	6	1,754	18	58
Food service - quick service	6	359	16	40
Health care - inpatient	6	1,037	21	27
Health care - outpatient	6	1,753	23	68
Very Large Buildings	1	276	15	N/A[3]
Very Large Data Centers	1	24	5	N/A[4]

[2] For this study, the food sales category captured primarily large grocery stores. Small food sales establishments were generally found under the “mercantile” building use category.

[3] There were 10 respondents in the very large building category, but they were recategorized into specific building use types during data collection.

[4] The study did not receive responses from any large data centers.

Building Type	Number of Size Strata	Population (premises)	Sample Size	Completes
Lodging	6	3,769	35	32
Mercantile - food sales	6	1,041	15	19
Mercantile - mall	6	3,668	15	20
Mercantile - standalone	6	14,243	15	41
Office	6	16,872	60	100
Other	6	16,663	33	25
Public assembly	6	2,473	34	46
Unknown	6	23,610	35	N/A[5]
Warehouse and storage	6	5,376	34	38
Total	93	94,538	420	596

Recruitment Design that Worked Across Studies

Across the study set, a two-track recruitment structure appeared most successful: a less-expensive, scalable, low-touch method for the bulk of small and medium sites, and a high-touch, research-driven track for the largest sites. Figure 2 below, an example from the CT study, illustrates that repeated contacts and multiple contact modes are likely necessary to achieve a substantial number of completed surveys from across the target population; however, mixed-mode outreach is not sufficient for success. As discussed above, studies must use up-to-date, accurate frames to have the best chance of achieving their targeted number of completed surveys, and they should plan ahead for segments that might pose particular recruitment challenges based on prior studies.

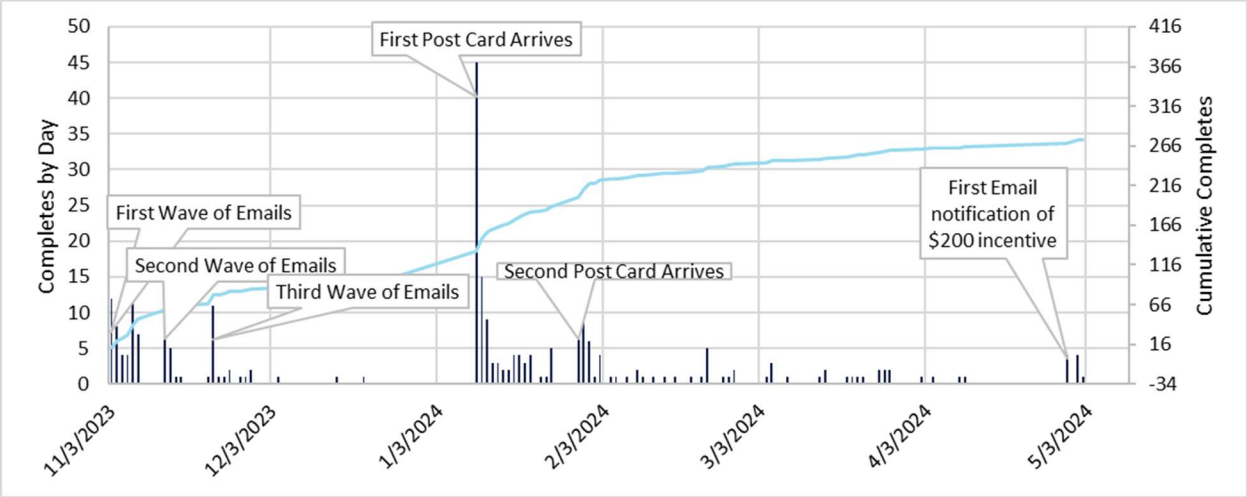

Figure 2. CT timeline of mixed-mode recruitment response

[5] There were 67 respondents in the "unknown" building type stratum, intended to represent all buildings in the population that could not be assigned a type based on data available pre-recruitment. These were recategorized into specific building use types during data collection.

Playbook Dimension 4: Determine Whether the Study Requires Tiered Data Collection and Identify Recruitment Method

Differences Across Studies

Three of the four studies used tiered data collection, but they differed in how they operationalized the tiering. The MA C&I Baseline study focused on the general C&I population and formally defined low-, medium-, and high-rigor data to be collected, directing sites to technical IDIs or on-site/virtual surveys after an initial web survey, depending on complexity. Figure 3 below shows the MA C&I Baseline study recruitment process flow. The CT C&I Baseline Study conducted an initial web survey followed by an opt-in, highly detailed technical survey, but the opt-in survey achieved lower conversion than expected. The NJ C&I baseline study didn't require converting survey respondents into a follow-up site visit. High-complexity sites were identified in advance and recruited only for in-depth interviews/site visits, while lower-complexity sites completed a survey and were not asked to complete a site visit. Figure 4 below shows the NJ C&I Baseline study recruitment process flow.

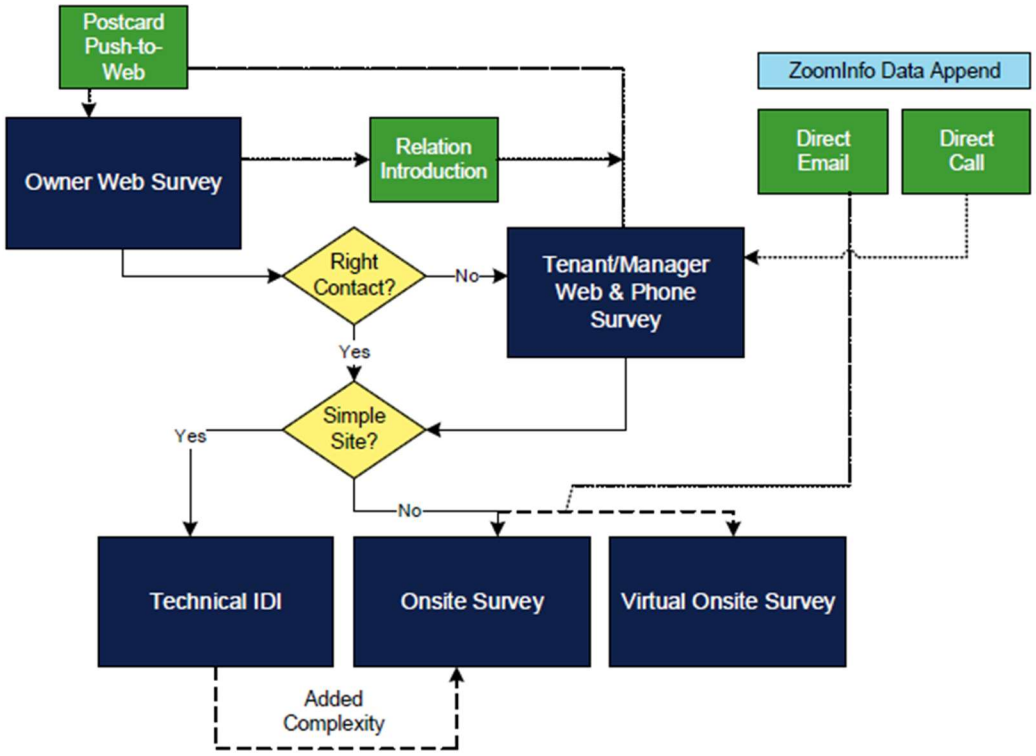

Figure 3. Example of a C&I existing building baseline study process flow: MA C&I Baseline Study

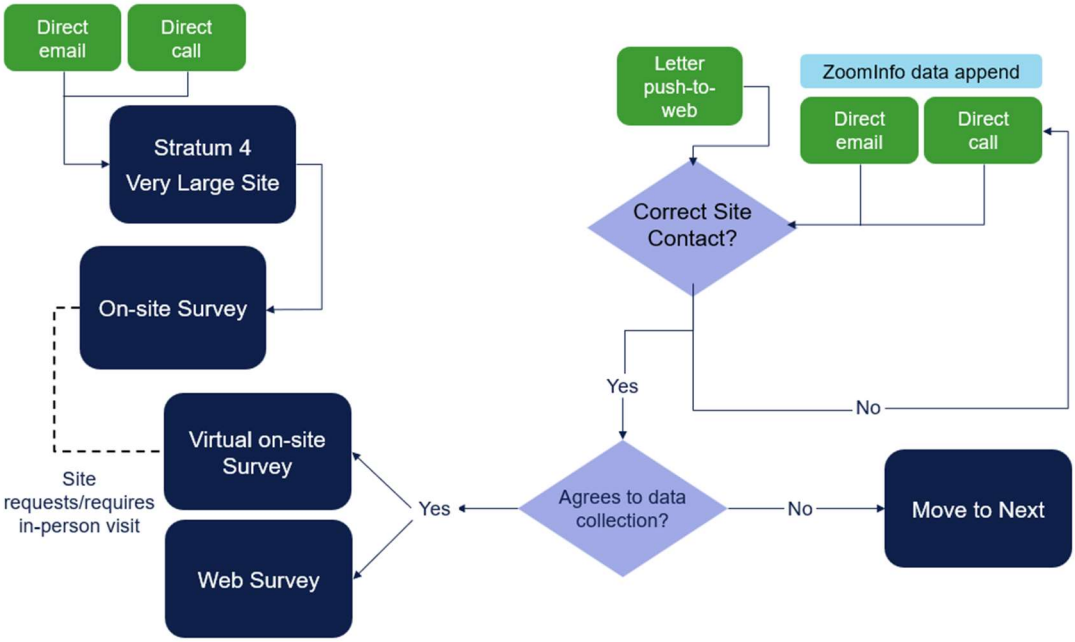

Figure 4. Example of a C&I existing building baseline study process flow: NJ C&I Baseline Study

Common Challenges

Follow-on yields for opt-in data collection after a preliminary survey tended to drop when the initial instrument was long, when follow-up expectations were unclear, or when the first respondent was not the best contact to provide detailed equipment information. Even if a respondent agrees to follow-up data collection, the timing of recruitment is critical: we found that the final conversion rate dropped rapidly over time if a recruiter could not reach out promptly after survey completion to schedule the interview or site visit. Opt-in designs limit initial respondent burden but can underperform without a clear value proposition and a low-friction scheduling pathway. Large and complex sites often require explicit scheduling support and may require different incentives and/or introductions to convert.

Playbook Guidance

The following guidance shows the tiering approaches that were used across the studies:

- Design the initial instrument to serve multiple purposes: collect low-complexity data, determine whether the respondent is qualified for a technical follow-on interview or site visit, and request their permission to schedule that additional data collection.
- Ensure the study has sufficient recruitment resources to reach out to respondents quickly after their initial survey for scheduling.
- Define data tiers and the necessary rigor for what can be self-reported versus what requires trained observation. This supports transparency about why high-rigor measurement is not attempted for the full population.
- Specify triggers for virtual versus in-person verification (e.g., complexity, specific end uses) and document those triggers in the study's methods.

Playbook Dimension 5: Use Clear Reporting Conventions to Prevent Misinterpretation

Study Examples

The following gives two examples from the study set in which the presentation of results could have led to misinterpretation without a clear explanation.

Difference in Weighting Results

One example of where clear reporting conventions, discussed in advance with stakeholders, are critical is around different weighting options. All referenced baseline studies explicitly distinguished business-weighted results ("percent of businesses") from consumption-weighted results (proxy for "percent of commercial space"), which may reduce misinterpretation when large sites drive a high share of load. Table 5 below shows how a large number of small sampling units represents a relatively small amount of population square footage, while a small number of large sampling units represents a large amount of population square footage. The MA C&I Baseline study was the third such study conducted in MA over the past decade, with each study changing the sampling unit based on data availability and study goals. The study's report emphasized that differences in sampling units and modes can produce misleading trend inferences across years and, therefore, limit cross-study trend claims. Figure 5 below illustrates this concern, showing substantial differences in the percentage of sampled units using different fuel types as their primary fuel source, as estimated in 2016, 2020, and 2023.

Table 5. MA C&I Baseline Study: distribution of parcels and square footage by parcel type

Parcel type	Sample n	Percentage of parcels	Percentage of square footage
Standalone small (0-10,000 sq ft)	455	59%	17%
Standalone large (>10,000 sq ft)	175	18%	42%
Building with shared walls	73	11%	18%
Multiple buildings	94	10%	22%
Don't know/Other	14	1%	1%

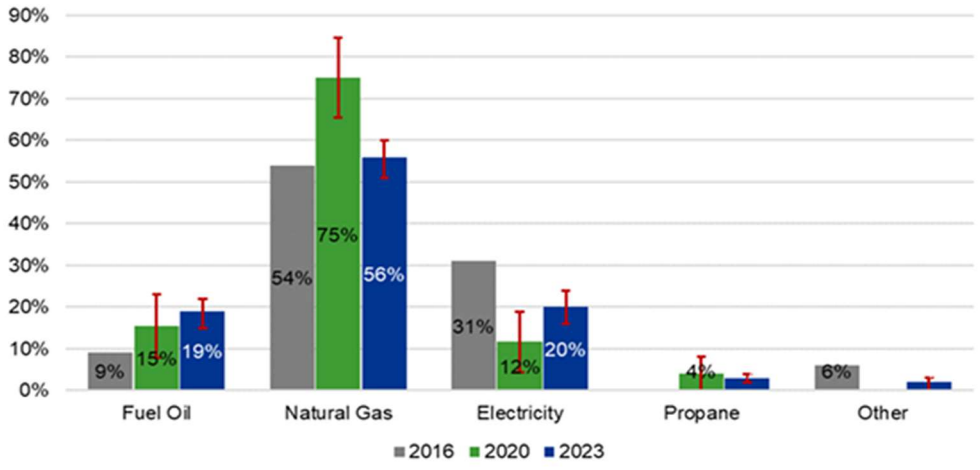

Figure 5. MA C&I Baseline Study: comparison of primary fuel source across baseline studies

Sample Dispositions and Domains Analysis

Another example of where clear reporting conventions are critical is in the discussion of sample dispositions, completion rates, and response rates, especially with respect to original stratum assignments based on information available in the sample frame vs reporting (domain/subcategory) assignments based on respondent feedback or site observations. It is necessary to report the achieved sample sizes in a disposition table with respect to the original sample design, but results will typically be reported by confirmed building type, meaning the sample sizes by segment in the disposition table will differ from those in the analysis tables and figures. Studies can avoid confusion around these differences by clearly discussing the difference in the reporting of dispositions. Table 6 below highlights the possible differences a study might observe, using the NJ C&I baseline study as an example, comparing the number of survey completions by original building type assignment with those by final building type assignment. As discussed in Playbook Dimension 1, the NJ study was particularly susceptible to incorrect building-type assignments in the original frame, due to its partial reliance on industry-sector information from a third-party business listing vendor. It was noted that businesses across many industry sectors are often located in office-type buildings, as illustrated by the shift from 41 "office" completes based on original assignments to 100 "office" completes based on confirmed assignments.

Table 6. NJ C&I Baseline Study: difference in survey completes based on original building type assignment and confirmed building type assignments

Building type	Number of completes based on original stratum assignments	Number of completes based on confirmed building type assignments
Data center	16	3
Education – college	40	79
Food service - full service	46	58
Food service - quick service	10	40
Health care – inpatient	33	27
Health care – outpatient	70	68
Large Buildings	10	N/A[6]
Lodging	114	32
Mercantile - food sales	27	19
Mercantile - mall	36	20
Mercantile - standalone	31	41
Office	41	100
Other	28	25
Public assembly	7	46
Unknown	69	N/A[7]
Warehouse and storage	18	38

[6] There were 10 respondents in the very large building category, but they were recategorized into specific building use types during data collection.

[7] There were 67 respondents in the "unknown" building type stratum, intended to represent all buildings in the population that could not be assigned a type based on data available pre-recruitment. These were recategorized into specific building use types during data collection.

Playbook Guidance

The following guidance supports a clear understanding of baseline study results and is consistent with the reporting practices observed in the study set:

- Define the sampling and reporting units and language used to describe them based on discussions with stakeholders. Ensure all parties understand the implications of different weighting methods (for example, "percent of businesses" versus "percent of commercial space").
- Since the achieved sample will govern inference, separate the intended sample design from the achieved sample in the study's methods narrative. Clearly discuss any differences in building type categorization between the sample design and the final, confirmed building types.
- For multi-phase designs, describe how follow-on selection affects expansion and variance estimation, even if the study defers detailed formulas to an appendix.

Conclusions: Cross-Study Playbook Summary

For evaluation teams and utility stakeholders planning future commercial baseline studies, the main lesson from these four studies is that field feasibility should be built into the study design from the beginning. A design may look efficient from a sampling perspective but still struggle in practice if the frame lacks current contact information, if the sampled unit does not match the person who can answer the study questions, or if large and complex sites require more recruitment effort than the budget and schedule allow.

These studies also point to the value of a two-track approach for commercial baseline work. Broad mixed-mode outreach can help support population coverage, while larger, more complex, or strategically important sites often require more direct and persistent recruitment. Tiered data collection can help manage respondent burden without giving up needed technical detail, but only if the study team decides in advance which data can be self-reported and which data require review or verification by trained staff.

Future baseline studies should therefore start with a practical design review, rather than treating the sample allocation as the main design exercise. Teams should assess the frame's strengths and limitations, identify segments that are likely to be difficult to recruit, define procedures for routing respondents to the right survey path or follow-up contact, plan for rapid follow-up when tiered data collection is used, and label results clearly so readers know whether estimates represent businesses, buildings, floor area, or energy consumption. These steps will not solve the response-rate challenge, but they can make the final estimates easier to interpret and make the study limitations clearer to the people using the results.

Table 9 closes the paper by summarizing the cross-study conclusions in a quick-reference matrix. The table does not recommend a single design. Instead, it links common study objectives and constraints to the design choices that appeared most workable across the reviewed studies.

Table 9. Cross-Study Summary of Design Tradeoffs in Commercial Existing-Building Baseline Studies

Playbook dimension	When this dimension is most consequential	Approaches observed across the study set	Observed challenges or constraints	Mitigations supported by the study set
Sampling unit and frame construction	Shared systems, multi-tenant buildings, master-metered loads, or large customers occupying multiple buildings.	Parcel-based, premises-based, and multi-source frames.	Incorrect respondent; inconsistent unit of analysis; stale or conflicting frame attributes.	Explicit respondent routing; de-duplication and source-priority rules; documentation of unit-of-analysis implications.
Stratification and precision planning	Large or complex sites drive load, savings, or policy relevance.	Size- and use-based stratification; explicit very-large strata.	Statistically reasonable strata not feasible in the field; precision expectations misaligned with applicability.	Treat large-site coverage as an operational design problem; adaptive targets; clarity on which metrics precision targets govern.
Recruitment and respondent routing	Billing contacts or owners are not knowledgeable respondents.	Mixed-mode outreach; two-track recruitment (scalable plus high-touch).	High effort required for large sites; low yield without explicit routing.	Budget and isolate high-touch recruitment; scripted routing; sponsor or utility support for largest sites.
Tiered data collection and follow-on engagement	High-rigor data needed but infeasible at scale.	Tiered designs separating survey-only and high-rigor follow-on.	Low follow-on conversion; sensitivity to timing and respondent burden.	Keep initial survey as short as possible; clearly communicate follow-on expectations; rapid post-survey scheduling.
Weighting and reporting conventions	Large sites dominate consumption and influence interpretation.	Business-weighted and consumption-weighted reporting.	Misinterpretation when weighing is unclear.	Explicit labeling of reporting units; visual aids; separation of intended versus achieved sample.
Role of baseline-adjacent studies	Depth needed on electrification barriers or system complexity.	Targeted large-site studies complement general-population baselines.	Expectation of representativeness from targeted work.	Clear statement of purpose-built scope; positioning as complementary evidence.

Implications for Future Baseline Studies

For evaluation teams, utility and program administrator staff, and stakeholders who design, fund, or review commercial existing-building baseline studies, the main takeaway from these cross-study comparisons is practical: future studies should match the research design to the conditions under which the study will actually be implemented. As response rates decline and access to knowledgeable respondents becomes more constrained—particularly for large and complex sites—baseline outcomes are shaped as much by frame construction, respondent routing, and tiered data collection strategies as by statistical design considerations.

Across the study set, choices about the sampling unit, frame construction, and stratification consistently shaped recruitment outcomes and the interpretation of results. Parcel-based, premises-based, and multi-source frames each proved workable in different contexts, but none removed the need for clear rules around respondent routing, de-duplication, and unit-of-analysis interpretation. Similarly, stratification and precision targets were most effective when they reflected field feasibility, particularly for large sites, where targeted recruitment and adaptive targets often determined whether priority segments were adequately represented.

Finally, the study set highlights the value of complementary research designs. It shows that tiered data collection can effectively balance rigor and feasibility, while targeted baseline-adjacent studies add insight into electrification barriers and system complexity that are hard to observe at scale. Being explicit about what each study is designed to do and the operational modifications required to implement it helps readers interpret and apply the results.

References

DNV. (2024a). 2024 NJ Baseline Study Frame Development and Sample Design (sample design memo dated June 24, 2024).

DNV. (2024b). Massachusetts Commercial and Industrial Existing Buildings Baseline Study (MA22C05-B-BSLN). Final report dated May 6, 2024. Available at ma-eeac.org.

DNV. (2025a). C2201 C&I Baseline Study and Database (Connecticut Energy Efficiency Board). Final Draft report dated December 23, 2025.

DNV. (2025b). Building Electrification for Large C&I Segment (MA24C07-B-TBELECTF). Final report dated November 21, 2025. Available at ma-eeac.org.

AI Tool Disclosure

AI-assisted tools used: Microsoft 365 Copilot.
Used for: Outlining, aggregating material for the final decision matrix, and initial editing.
Extent: Used on text only. We manually developed all figures and tables displaying the study results.
Human review: The authors reviewed and verified all content, calculations, and citations.