Analysis Plan

Evaluation Name: Descriptive Evaluation of EPA's Biennial Hazardous Waste Report | **Evaluation Code:** 2602

Date Finalized: 11/5/2026

Project Description

The U.S. Environmental Protection Agency (EPA) has an imperative to protect human health and the environment, which includes overseeing, alongside states, the management of the nation's hazardous waste facilities, which are regulated under the Resource Conservation and Recovery Act of 1976 (RCRA), as amended by the Hazardous and Solid Waste Amendments of 1984 (HSWA).

As part of its authority to control hazardous waste "from cradle to grave", the EPA oversees the National Biennial RCRA Hazardous Waste Report (Biennial Report, or BR). The BR includes a set of forms describing the nature, quantities and disposition of hazardous waste generated at their facility. Facilities who meet the Federal definition of a RCRA Large Quantity Generator (LQG) or Treatment, Storage, and Disposal Facilities (TSDF) are required to submit the BR.[1,2] The report is completed on a two-year cycle and is due on March 1 of every even year (e.g., reports due by March 1, 2026 described activities from calendar year 2025).

EPA has indicated that the process of filling out and completing the BR is a significant burden both for facilities who are responsible for preparing and submitting it (i.e., BR preparers), as well as regulators (i.e., state and EPA staff) who process and review the data. EPA wants to increase the efficiency of BR reporting and alleviate the administrative burden on both form submitters and state and EPA staff by adjusting the BR forms and process, as well as understanding how they can leverage data from the Hazardous Waste Manifest System to improve data integrity of the BR.

Using BR data spanning three reporting years (2021, 2023, 2025), alongside other relevant datasets (e.g., Manifest data), OES is collaborating with EPA to conduct a descriptive evaluation that summarizes trends in BR reporting behavior. There are multiple goals of this evaluation: to (a) summarize trends in BR preparers' characteristics and reporting behavior (e.g., report return rates), (b) investigate how BR data aligns with Manifest data, and (c) generate recommendations that EPA can use to make improvements to the next BR reporting cycle, due in March 2028. For example, these findings may reveal common issue areas for BR preparers, or point to additional system validations or autopopulation of form fields that could be introduced to ease form completion.

Preregistration Details

This Analysis Plan will be posted on the OES website at oes.gsa.gov before outcome data are analyzed. Prior to posting this Analysis Plan, OES had access to the following non-publicly available datasets, sent by EPA:

1. BR Report data for 2025 (partial)

[1] More detailed definitions of a federal LQG or TSDF can be found in EPA's booklet 'RCRA Subtitle C Reporting Instructions and Forms' (p. 46-47).
[2] A given facility may generate waste or receive waste, or both.

2. BR Validations data for 2023
3. BR Logging data for 2025
4. BR Review data for 2021 and 2023

OES used this data for data pre-processing and preparing the analysis code only, and strictly did not examine outcome data. Additionally, we had access to publicly available Manifest data for 2021, 2023, and 2025 from EPA's RCRAInfo platform here, and BR data for 2021 and 2023 here.

Research Questions

Analyses will focus on estimating counts, proportions, rates, averages, medians, distributions, and pre-specified associations.

The evaluation is designed to describe BR reporting behavior, validation activity, report returns, and the relationship between BR data and Manifest data across BR reporting years 2021, 2023, and 2025.

The primary research questions are as follows:

Questions related to BR preparers:

1. What are the characteristics of BR preparers? For example:
    a. Facility type (e.g., LQG or TSDF)
    b. Geography (e.g., EPA region)
    c. Shipment frequency
    d. Hazardous waste volume

Questions related to validations:

1. How many preparers likely encounter errors or warnings when preparing the BR?
2. How often do preparers provide written justifications in response to different warning categories?
3. What are commonly used justifications for different warning categories?

Questions related to report returns/corrections:

1. How frequently were BR submissions returned by regulatory reviewers to preparers for revisions? Did this rate change over time?
2. Are there certain preparer characteristics that are associated with higher report return rates? For example:
    a. Facility type
    b. Geography
    c. Shipment frequency
    d. Hazardous waste volume
    e. Whether or not preparer reports annually
    f. Rate of electronic manifest usage (vs. mix or paper only)
3. What are the most common reasons a report might be returned?

Questions related to Manifest data and how it corresponds to BR data:

1. Which fields are used in the BR and Manifest systems to report similar or comparable information?

Data and Data Structure

This section describes variables that will be analyzed, as well as changes that will be made to the raw data with respect to data structure and variables.

Data Source(s)

The analysis will use a combination of administrative data provided directly by EPA and publicly-available datasets available from RCRAInfo, outlined in Table 1.

Table 1

Data Sources

Dataset (BR years covered)	Description	Example data elements	Retrieval method
BR form data (2021, 2023, 2025)	This data includes site-level and waste-stream-level information submitted through the BR. We will mainly use this to examine BR preparer characteristics and quantities (e.g., amount of waste generated in a given year).	Handler identifiers, reporting year, geography, generator status, waste quantities, waste codes, source codes, form codes, management methods, and shipment or receipt information	BR data for 2021 and 2023 is publicly available on RCRAInfo here. BR data for 2025 will be shared with OES by EPA in November 2026.
Manifest data (2021, 2023, 2025)	This data includes shipment-level and waste-line-level information. We will mainly use this to summarize (on a year level) quantities for comparison with BR data (e.g., total waste shipped).	Generator ID, receiving facility ID, shipment and receipt dates, manifest submission type, waste quantity, units, waste codes, management method codes	Publicly available on RCRAInfo here
Validations data (2023, 2025)	This data captures instances where a BR preparer provided a written explanation in response to a BR validation warning. We will mainly use it to count instances of written justifications in response to warning validations, and categorize common justifications.	Handler ID, reporting year, BR form, validation issue description, issue explanation, timestamps	Transfer from EPA
Logging data (2025)[3]	Log of instances when industry users or regulatory	Handler ID, reporting year, user type (i.e., regulatory or industry),[4]	Transfer from EPA

[3] Logging data are not retained beyond two years, so this dataset is only expected to be available for the 2025 reporting year.
[4] The logging data were shared with OES by EPA; therefore, no public data dictionary is available. According to EPA, the USER_TYPE variable corresponds to "industry" when the preparer submits the BR, while USER_TYPE = "regulatory" indicates that the state authority accessed the submitted report to review and approve it. After successful submission and eventual approval of a BR, the logging data would generally be expected to show at least two rows per handler: one "industry" entry and one "regulatory" entry. Multiple rows per handler may indicate additional interactions during the review process, such as errors or deficiencies identified by the regulatory authority that required correction by the handler, followed by resubmission and subsequent re-review or approval by the regulatory authority.

	users ran the BR validation process.	user ID, and date/time of validation activity.	
Review data (2021, 2023, 2025)	Data showing individual reviewer actions on submitted BRs to ascertain whether a given BR was for example rejected or resubmitted prior to final acceptance of the report.	Indicator for whether a report was rejected or returned, associated comments, handler ID, year, timestamps, action/status fields	Transfer from EPA

Additional external or constructed data may be merged in for selected analyses, including:

- State-year indicators for whether a state used RCRAInfo for BR submission in a given reporting year (shared by EPA)
- State-year indicators for whether a state required annual hazardous waste reporting in a given reporting year (shared by EPA);
- Rural-urban classification data, such as the USDA 2020 Rural-Urban Commuting Area (RUCA) Codes, using site ZIP code or county information (publicly available here)
- Indicators showing what EPA region each state is in (publicly available here)

Data integration and linkage:

The above datasets will be linked using different indicators in each dataset, highlighted below.

- BR data: Link to the Manifest data via a multistep process, to account for different types of waste streams and storage facilities, variables for matching involve the BR form data's HANDLER_ID, RECEIVER_ID, REPORT_CYCLE, and SHIPPED_TONS. Contingent on data quality, more variables may need to be drawn on during the pre-processing for the analysis.
- Manifest data: The multistep process of linking the BR data to the Manifest data, will involve the following variables HANDLER_ID, SHIPPER_ID, YEAR, and RECEIVED_TONS. Contingent on data quality, more variables may need to be drawn on during the pre-processing for the analysis.
- Validations data: Link to BR data via the unique HANDLER_ID in both datasets.
- Logging data: Link to BR data via the unique HANDLER_ID in both datasets.
- Review data: Link to BR data via the unique HANDLER_ID in both datasets.

The primary unit of analysis for most descriptive summaries will be the **handler-year**, defined by unique HANDLER_ID and REPORT_CYCLE or YEAR. Some analyses will use other units, including waste-stream-level BR records, manifest waste-line records, validation-warning records, or report-return records.

Outcomes to Be Analyzed

A. Characteristics of BR preparers

For each reporting year,we will describe the population of BR preparers using the following outcomes:[5]

1. **Number of unique preparers**
    - Count of unique HANDLER_ID values.
2. **Preparer type**
    - Count and proportion of preparers by handler type (e.g., LQG,[6] non-LQG, TSDF)
3. **Geography**
    - Count and proportion of preparers by state and EPA region.
    - Count and proportion of preparers by urban/rural classification.
4. **Shipment frequency**[7]
    - Shipment and receipt frequency: Mean, median, and range of shipment and receipt counts per BR preparer, reported as absolute counts and proportions.
    - Distribution of shipment and receipt activity: Distribution of counts and proportions across BR preparers, including the interquartile range.
5. **Hazardous waste volume**
    - Total, mean, median, and distribution of reported hazardous waste quantities generated, managed, shipped, and received, using BR variables including GENERATION_TONS, MANAGED_TONS, SHIPPED_TONS, and RECEIVED_TONS.
    - Quantities summarized nationally and by state or region, where appropriate (see 'Waste quantity aggregates' in Table 2 for further information).

B. Validation-related outcomes

For reporting years with validation-related data, we will analyze:

1. **Likely validation encounter**
    - For 2025 only, an indicator for whether a preparer likely encountered at least one validation warning or error, defined as appearing more than once in the logging data where user type equals industry.
    - This measure is expected to overstate true validation encounters because preparers can click "Validate" multiple times even if no warning or error was displayed.
2. **Number of industry validation runs**
    - For 2025 only, the number of times each handler appears in the logging data as an industry user.

---

[5] Note that BR preparers may not meet the thresholds to be required to submit a BR each reporting cycle. For example, a facility may meet the definition of a LQG in 2021 and therefore submitted a BR for that year, but may not be observed in later reporting cycles because they no longer met the LQG status.

[6] Facilities that exceed LQG generation thresholds solely due to a qualifying episodic event may retain their existing generator classification and are generally exempt from federal BR requirements on that basis alone. Facilities that otherwise qualify as LQGs during the reporting year may still be required to report.

[7] Some BR preparers may only report on shipments they export (e.g., LQGs who generate waste that they then ship), some only report on shipments they receive (e.g., TSDFs who treat waste they receive from off-site), and some do both (generating and receiving).

- We will summarize the total number of validation runs and the number beyond the first run.

3. **Written validation justifications**
    - For 2023 and 2025, count and proportion of preparers who provided at least one written justification in response to a validation warning.
    - Count and proportion of justifications by validation issue category using ISSUE_DESCRIPTION.
4. **Validation justification content**
    - Categorized response types as Warning or Error based on ISSUE_EXPLANATION, summarized by warning category (see further explanation of this under 'Validation justification categories' in Table 2).
    - Categories will be developed using a documented coding process described in Exploratory Analysis.

C. Report-return and correction outcomes

For reporting years 2021, 2023, and 2025, we will analyze:

1. **Report return indicator**
    - Indicator for whether a preparer had a report rejected by a regulatory reviewer, defined primarily as presence in the review data with STATUS = "Rejected" for a given handler-year.
2. **Returned-by-preparer indicator**
    - Indicator for whether a report was returned to the preparer or reopened based on preparer request, defined as presence in the review data with STATUS = "Returned", where EPA guidance indicates this typically reflects proactive preparer-requested correction.
3. **Report return rate**
    - Proportion of preparers whose reports were rejected by reviewers, by reporting year.
    - Proportion of preparers whose reports were returned or reopened, by reporting year.
4. **Number of return or rejection events**
    - Count of rejection or return events per handler-year.
5. **Repeat return patterns**
    - Count and proportion of preparers with returned or rejected reports in more than one reporting year.
    - Count and proportion of preparers with returns in an earlier year but not in a later year.
6. **Reasons for report returns**
    - Categories of return reasons based on the COMMENTS field in the review data (further explained under 'Report return reason categories' in Table 2).
    - Rank-ordered frequency and percentage of return reason categories.

D. Relationship between BR and Manifest data

Identifying BR Handlers in Manifest Data: We will use a sequential approach to link BR handlers to manifest records. First, we will match the BR HANDLER_ID to the Manifest GENERATOR_ID. For manifest records with missing generator_id values, we will then attempt to identify additional links by matching the BR

HANDLER_ID to the Manifest DES_FACILITY_ID. For records that remain unmatched, we will explore additional matching approaches using relevant variables and quantities available in the datasets, informed by the RCRAInfo data dictionary and in consultation with EPA. We will document the matching criteria and assess the plausibility of proposed matches. After these steps, we will review the remaining unmatched records with EPA to determine whether further linkage methods are warranted and how unresolved records should be handled in subsequent analyses.

After matching, we we will analyze overlap and differences between BR and Manifest data, including:

1. **Comparable fields**
    - Inventory of fields that capture similar or potentially comparable information across the two systems, including handler/generator ID, receiving facility ID, waste codes, management method codes, quantities, shipment or receipt status, and dates.
2. **Manifest-BR quantity comparisons**
    - For preparers with associated manifest records, comparison of BR shipped tons to Manifest shipped tons aggregated by generator-year.
    - Comparison of BR received tons to manifest received tons aggregated by receiving facility-year.
    - Absolute and percentage differences between BR and manifest quantities, where comparable.
3. **Manifest availability**
    - Create an indicator for whether a BR preparer appears in Manifest data as a generator during the relevant reporting year, and provide the count and proportion of these preparers.
    - Create an indicator for whether a BR preparer appears in Manifest data as a receiving facility during the relevant reporting year, and provide the count and proportion of these preparers.
4. **Electronic manifest use**
    - If available in the Manifest data, the proportion of manifests submitted electronically versus paper or other submission types, by handler-year.

Imported Variables

The following variables may be added to the raw administrative data through merges or constructed crosswalks:

1. **Annual reporting indicator**
    - A state-year indicator for whether the preparer's state required annual hazardous waste reporting in the relevant reporting year.
    - This will be constructed using EPA-provided information on state reporting requirements.
2. **RCRAInfo use indicator**
    - A state-year indicator for whether the state used RCRAInfo for BR submission in the relevant reporting year.
    - This indicator will be used to identify which states should have validation-related data.
3. **EPA region**

- The BR form data contains a REGION variable, that – though not defined in the publicly available data dictionary – seems to correspond to the EPA region. As an additional quality check, EPA region will be assigned to generators based on state using publicly-available data (described above).

4. **Urban/rural classification**
    - Urban/rural classification will primarily use USDA RUCA codes, which classify geographic areas based on urbanization and commuting patterns. ZIP-code-level RUCA classifications will be merged with BR handler records using standardized LOCATION_ZIP values where available. If ZIP codes are unavailable or cannot be reliably matched, alternative geographic identifiers (e.g., county FIPS codes) may be used where an appropriate classification is available. Records without a reliable geographic match will be flagged as unclassified.
5. **Manifest submission type or electronic-use indicators**
    - If available in the Manifest data, variables such as ORIG_SUBMISSION_TYPE or ORIGIN_TYPE will be used to classify manifests as electronic, paper, or other submission types.

Transformations of Variables

Table 2 describes the variables that will be created or transformed for analysis. Others may be created or transformed during analysis as needed.

Table 2

Transformation of Variables

Variable	Description
Handler-year identifiers	A unique handler-year identifier will be created by combining HANDLER_ID and reporting year, using REPORT_CYCLE in the BR data and YEAR in validations, logging, and review data.
BR preparer indicator	A preparer will be defined as a unique HANDLER_ID appearing in the BR form data for a given reporting year. The primary preparer-year count will be the number of unique handler-year records.
Preparer type	Preparer type will be based on the EPA-provided generator or handler status variable where available. If needed, we may also use available fields such as CALCULATED_GENERATOR_STATUS, form submission patterns, or inclusion flags to construct categories.[8] Where a handler submits only 'Waste Generation and Management' (GM) forms, we may classify the handler as a generator-only preparer.[9] Where a handler submits 'Waste Received from Off-Site' (WR) forms or has received quantities, we may classify the handler as a receiver or destination facility-related preparer.[10] Where a handler has both generation/shipment and receipt/management activity, the handler may be classified as both generator and receiver/manager.

[8] This variable indicates whether a preparer met one of three nationally defined values of generators: LQGs, Episodic LQGs, or Not a LQG. See further information here.
[9] Each GM form generally represents a reported hazardous waste stream. One facility can submit multiple GM forms for different waste streams.
[10] Each WR form generally represents a reported hazardous waste stream received from an off-site source. One facility can submit multiple WR forms for different waste streams received from one or more off-site facilities.

Waste quantity aggregates	For each handler-year, we will aggregate reported quantities from the handler-year-shipment level BR form data: ● total_generation_tons = sum(GENERATION_TONS) ● total_managed_tons = sum(MANAGED_TONS) ● total_shipped_tons = sum(SHIPPED_TONS) ● total_received_tons = sum(RECEIVED_TONS) Aggregations will be conducted separately by reporting year. Because each BR form is expected to correspond to a single waste stream's annual activity, we will assess whether records are unique at the expected form level and flag duplicate records for further investigation as potential data quality issues. For quantity totals, we will sum across distinct records as appropriate for each quantity field. The National Biennial Report (NBR) inclusion flag will be used to distinguish federally required waste streams from other reported waste. For comparisons with Manifest data, we may consider both federally required and excluded waste, as excluded waste may still be relevant to shipment counts.[11] Where relevant, we will create versions of these aggregates restricted to records where the corresponding National BR inclusion flag equals Y, such as GEN_WASTE_INCLUDED_IN_NBR, SHIP_WASTE_INCLUDED_IN_NBR, RECV_WASTE_INCLUDED_IN_NBR, or MGMT_WASTE_INCLUDED_IN_NBR.
Shipment and receipt counts from Manifest data	For each reporting year, manifest records will be restricted to shipments or receipts occurring in the relevant calendar year. For generator-side analyses: ● Use GENERATOR_ID as the handler identifier. ● Count unique manifests using MANIFEST_TRACKING_NUMBER. ● Sum manifest quantities using QUANTITY_TONS or another EPA-recommended standardized quantity field. For receiving-facility analyses: ● Use DES_FACILITY_ID as the handler identifier. ● Count unique manifests using MANIFEST_TRACKING_NUMBER. ● Sum manifest quantities using QUANTITY_TONS or another EPA-recommended standardized quantity field. Where Manifest data are at the waste-line level, shipment counts will count unique manifest tracking numbers rather than waste-line rows, unless the analysis specifically concerns waste lines.[12,13]
Electronic manifest use	If manifest submission type variables are available, we will create handler-year-level measures including: ● total number of manifests; ● number of electronic manifests;[14]

[11] The NBR inclusion flag indicates whether a reported waste stream is included in the BR dataset for purposes of federal reporting requirements. Waste streams excluded from the NBR may still be relevant for analyses of waste quantities and shipment activity.
[12] One manifest can contain multiple waste lines, so we count each manifest only once when calculating the number of shipments, rather than counting every row in the dataset.
[13] Manifest records may report waste quantities by volume rather than weight, and density information is not required for all records. Where quantities must be converted to weight and density is missing, the data-processing method may assume the density of water. This assumption may introduce measurement error because actual waste density varies by waste type. Consequently, discrepancies between manifest and BR quantities may partly reflect differences in units or density assumptions rather than differences in the underlying waste quantities. We will document the conversion approach and interpret quantity comparisons with this limitation in mind.
[14] EPA's e-Manifest program now treats both Hybrid and FullyElectronic ORIG_SUBMISSION_TYPE values as "electronic manifests".

	• number of paper or non-electronic manifests; • proportion electronic. The exact classification will depend on values available in ORIG_SUBMISSION_TYPE, ORIGIN_TYPE, or related fields.
Likely validation encounter	From the Logging data, for the 2025 reporting year, we will create the following variables: • industry_validation_runs: number of logging records for a handler-year where user type = industry; • likely_validation_encounter: indicator equal to 1 if industry_validation_runs > 1, and 0 otherwise; • additional_industry_validation_runs: industry_validation_runs - 1, with a minimum of 0. This measure will be interpreted cautiously because repeated validation runs may occur even if no warning or error was displayed.
Regulatory validation or review activity	From the Logging data, for the 2025 reporting year, we will create: • regulatory_validation_runs: number of logging records for a handler-year where user type = regulatory; • multiple_regulatory_validation_runs: indicator equal to 1 if regulatory_validation_runs > 1. This measure may be used as a secondary or sensitivity measure for report-return activity. We expect to observe at least two rows per BR preparer in the logging data (one related to validation activity from an industry user, and one from a regulatory user). Therefore, if we observe more than two rows per BR preparer where the user type is industry, it is an indication that the preparer may have seen validation warnings.
Written validation justifications	From the Validations data, we will create: • provided_validation_justification: indicator for whether a handler-year has at least one validation record; • num_validation_justifications: number of validation records for the handler-year; • num_validation_categories: number of unique ISSUE_DESCRIPTION categories for the handler-year. We will also tabulate counts by ISSUE_DESCRIPTION.
Validation justification categories	From the Validations data, we will develop categorical codes for ISSUE_EXPLANATION responses. These codes will be finalized during analysis on review of the data. Potential categories may include, for example: • quantity difference due to density or unit conversion; • legitimate change in production, operations, or waste generation; • site closure, startup, episodic event, or one-time cleanup; • manifest data incomplete or not applicable; • waste managed on site; • state-specific reporting difference; • correction made or data entry issue acknowledged; • other explanation; • unclear or insufficient explanation. If a single explanation contains multiple reasons, we will assign either the primary reason or allow multiple codes, depending on the coding approach selected before analysis. The selected approach will be documented in the analysis code and Record of Analysis.
Report return indicators	From the Review data, we will create the following handler-year-level indicators:

	• report_rejected: indicator equal to 1 if any record for the handler-year has STATUS = "Rejected"; • report_returned: indicator equal to 1 if any record for the handler-year has STATUS = "Returned"; • any_report_return_or_rejection: indicator equal to 1 if either status is present; • num_rejection_events: number of records with STATUS = "Rejected"; • num_return_events: number of records with STATUS = "Returned". The primary outcome for reviewer-required corrections will be report_rejected.
Report return reason categories	From the Review data, we will categorize COMMENTS using a documented coding scheme. Potential categories may include: • missing or incomplete form; • quantity discrepancy; • generator status issue; • waste code issue; • management method issue; • site identification or address issue; • receiving facility issue; • manifest-related discrepancy; • state-specific requirement; • administrative or certification issue; • other; • unclear or missing comment. Categories may be developed using keyword searches, manual review of a sample of comments, or a combination of automated and manual coding. The final coding scheme will be documented.
Manifest-BR comparison variables	For handler-years that can be linked across BR and Manifest data, we will create comparison variables such as: • manifest_shipped_tons: total tons from manifest records where the handler is listed as generator during the reporting year; • br_shipped_tons: total shipped tons reported in the BR; • shipped_tons_difference = br_shipped_tons - manifest_shipped_tons; • shipped_tons_absolute_difference = abs(br_shipped_tons - manifest_shipped_tons); • shipped_tons_percent_difference = (br_shipped_tons - manifest_shipped_tons) / manifest_shipped_tons, where manifest shipped tons is greater than zero. Parallel variables will be created for received tons, such as: • manifest_received_tons; • br_received_tons; • received_tons_difference; • received_tons_absolute_difference; • received_tons_percent_difference. We will avoid calculating percentage differences where the denominator is zero. Such cases will be categorized separately.

Transformations of Data Structure

The raw data are expected to have different structures across sources. We will create several analytic datasets summarized below.

1. **Comprehensive BR data file**

The main analytic file will be a handler-year-level BR dataset with one row per HANDLER_ID and reporting year. This file will merge the BR datasets from 2021, 2023 and 2025, and in addition to including

transformed and calculated variables, will append variables from other datasets where it makes sense to do so. For example:

- BR preparer characteristics;
- BR quantity aggregates;
- manifest shipment and receipt aggregates;
- validation justification summaries;
- logging summaries for 2025;
- review summaries;
- imported state, region, annual reporting, RCRAInfo-use, and urban/rural variables.

This file will be used for most descriptive summaries and association analyses.

2. **Comprehensive Manifest file**

The handler-year-level Manifest dataset will be appended with calculations and summaries resulting from our data transformation, where it makes sense. Some links to the BR may be made where it makes sense.

3. **BR waste-stream file**

A waste-stream-level BR file may be retained for summaries involving waste codes, source codes, form codes, management methods, and quantities by waste stream.

4. **Manifest waste-line file**

A manifest waste-line-level file may be retained for summaries of manifest waste codes, quantities, management methods, and shipment characteristics.

5. **Validation-record file**

A validation-record-level file will be used to summarize validation warning categories and written explanations.

6. **Review record file**

A report-return-record-level file will be used to summarize return/rejection comments and construct return reason categories.

Data Exclusion

We will use the following exclusion rules.

1. **Reporting years outside scope**
    - Records outside reporting years 2021, 2023, and 2025 will be excluded from the primary analysis.
2. **Manifest dates outside relevant reporting year**
    - Manifest records will be included in a reporting-year analysis only if the relevant shipment or receipt date falls within the corresponding calendar year.
    - For generator-side shipment analyses, SHIPPED_DATE will be used where available.
    - For receiver-side receipt analyses, RECEIVED_DATE will be used where available.
3. **Missing handler identifiers**
    - In the BR form data, records missing the relevant handler identifier will be excluded from analyses requiring linkage by handler ID.
    - These records may be summarized separately if they are substantively meaningful and sufficiently numerous.

4. **Duplicate records**
    - Exact duplicate records will be removed where identifiable.
    - For manifest shipment counts, duplicate waste-line rows within the same manifest will not be counted as separate manifests.
5. **States without RCRAInfo validation coverage**
    - For validation-related analyses, preparers in states that did not use RCRAInfo for the relevant reporting year will be excluded from denominators where validation data could not have been captured.
    - These exclusions will be explicitly documented in tables and figure notes.
    - States that transitioned into RCRAInfo during the study period will be handled according to their state-year RCRAInfo-use status.
6. **Non-final 2025 data**
    - If 2025 data are received before EPA's final data publication or finalization process is complete, analyses using 2025 data will be labeled preliminary until EPA confirms that the data are final.
    - Final reported results will use the most recent data available before the Record of Analysis is completed.
7. **Implausible or invalid quantity values**
    - Negative quantities or other impossible values will be flagged.
    - We will consult EPA or relevant data documentation before excluding or recoding such values.
    - If excluded, the number and type of exclusions will be reported.
8. **Comment fields unrelated to substantive report returns**
    - For analyses of return reasons, comments that are blank, administrative-only, or not interpretable will be categorized as missing, administrative, or unclear rather than excluded.

Treatment of Missing Data

The analysis will not use imputation for missing administrative data. Missing values will be handled as follows:

1. **Descriptive denominators**
    - For each descriptive statistic, denominators will be reported or made clear in table notes.
    - Where a variable is missing for some records, percentages will generally be calculated among records with non-missing values for that variable unless otherwise specified.
2. **Unknown categories**
    - For categorical variables such as preparer type, urban/rural classification, annual reporting status, or manifest submission type, missing or unresolved values may be retained as an "Unknown" category where useful for interpretation.
3. **Missing quantity values**
    - Missing quantity values will not be assumed to be zero unless EPA documentation indicates that missingness corresponds to no reported activity.
    - Where appropriate, we will distinguish between true zero values and missing values.

4. **Missing manifest matches**
    - A BR preparer with no matching manifest records will be interpreted as having no observed manifest records in the provided manifest data for that reporting year, not necessarily as having no manifest-required waste activity.
    - This distinction is important because some waste may be managed on site, exempt from manifest reporting, recorded under another identifier, or absent from the provided data for other reasons.
    - Missing matches between the BR form and Manifest data can be traced back to missing generator identifiers in the Manifest data. We will take several steps to fill any missing values (including matching on DES_FACILITY_ID and other variables, if non-matches remain).
5. **Missing comments**
    - Missing or blank COMMENTS fields (incl. which contain a string but no content, e.g. simply a full stop or dash) in the review data will be categorized as missing and excluded from analyses of substantive return reasons, while still included in counts of return/rejection events.
6. **Missing validation explanations**
    - The validations dataset is expected to include only warnings for which an explanation was provided. However, some rows in the dataset are not actual explanations (e.g., they contain a string in the form of solely a full stop or dash, essentially leading us to classify this as a non-explanation). Therefore, absence from the validations dataset will not be interpreted as absence of a validation warning.

Descriptive Statistics, Tables, and Graphs

The evaluation will include pre-specified descriptive statistics, tables, and graphs, described further below (note that some may be adjusted for readability in final reporting).

Table A

BR Preparer Characteristics By Reporting Year

Suggested rows:

- number of preparers (N_Handlers_unique);
- number and percentage by preparer type (Preparer_type_{}_N & Preparer_type_{}_proportion);
- mean and median generation tons (generation_tons_mean & generation_tons_median);
- mean and median shipped tons (shipped_tons_mean & shipped_tons_median);
- mean and median received tons (received_tons_mean & received_tons_median);
- mean and median managed tons (managed_tons_mean & managed_tons_median);
- mean and median manifest shipment count (shipment_count_mean & shipment_count_median).

Suggested columns:

Characteristic	2021	2023	2025

Additionally, we will offer summaries of some of the BR preparer characteristics in separate tables, by State, EPA region and urban/rural classification. Suggested columns for the respective tables:

	2021		2023		2025	
State	**N Preparers**	**Percent Preparers**	**N Preparers**	**Percent Preparers**	**N Preparers**	**Percent Preparers**

	2021		2023		2025	
EPA Region	**N Preparers**	**Percent Preparers**	**N Preparers**	**Percent Preparers**	**N Preparers**	**Percent Preparers**

	2021		2023		2025	
RUCA	**N Preparers**	**Percent Preparers**	**N Preparers**	**Percent Preparers**	**N Preparers**	**Percent Preparers**

Figure A

Number of BR Preparers By Reporting Year

Bar chart showing the number of unique preparers in 2021, 2023, and 2025.

Figure B

Geographic Distribution Of Preparers

A map showing the percentage of BR preparers by state for each reporting year, using a continuous color gradient to represent the percentage. Where useful, maps may be presented as separate panels by reporting year to facilitate comparisons over time.

Figure C

Distribution Of Hazardous Waste Volumes

Histogram, box plot, or density plot showing generation, shipped, managed, and received tons. Because hazardous waste quantities may be highly skewed, we may use a log scale or present medians and percentiles.

Table B

Validation Activity And Written Justifications

Suggested rows:

- number of preparers in validation-covered states (N_Preparers_validationcoveredstates);
- number and percentage likely encountering validations, 2025 only (N_Preparers_validationcoveredstates & Preparers_likely_encountering_validations_percent) ;

- number and percentage providing written justifications (Provided_validation_justification_N & Provided_validation_justification_percent);
- number of validation justification records (num_validation_justifications);
- mean/median number of justification records among preparers with at least one justification (Preparers_providedjustification_numjustifications_mean & Preparers_providedjustification_numjustifications_median).

Suggested columns:

Validation measure	**2023**	**2025**

Table C

Validation Justification Counts By Warning Category

Suggested columns:

Validation warning category	**2023 count** (using N_validation_category_2023)	**2023 %** (using Percent_validation_category_2023)	**2025 count** (using N_validation_category_2025)	**2025 %** (using Percent_validation_category_2025)

Table D

Common Validation Explanation Categories

Suggested columns:

Explanation category	**Count**	**Percent**[15]	**Example interpretation**

Figure D

Most Common Validation Warning Categories

Bar chart showing the most frequent ISSUE_DESCRIPTION categories among records with written justifications.

Table E

Report Return and Rejection Rates By Year

Suggested rows:

- number of preparers (N_HANDLER_ID_unique_BRsubmissionhistory);
- number and percentage with STATUS = Rejected (report_rejection_count & report_rejection_percent);
- number and percentage with STATUS = Returned (report_returned_count & report_returned_percent);

[15] Percent will be out of a total of all validation justification occurrences (noting that some preparers may provide justifications in response to multiple validations, where they are triggered).

- number and percentage with any return or rejection (repeat_returns_2021_2023);
- mean/median number of rejection events among rejected preparers (Rejections_per_REJECTED_handler_mean} & Rejections_per_REJECTED_handler_median}).

Suggested columns:

Report-return measure	2021	2023	2025

Figure E

Report Rejection Rate Over Time

Line or bar chart showing the percentage of preparers with a reviewer rejection by reporting year.

Table F

Report Rejection Rates By Preparer Characteristic

Suggested columns:

Characteristic	Category	Number of preparers	Number rejected	Rejection rate

Characteristics may include preparer type, EPA region, annual reporting state, shipment frequency, hazardous waste volume, and electronic manifest use.

Table G

Common Reasons For Report Returns Or Rejections

Suggested columns:

Return/rejection reason category	Count	Percent

Figure F

Most Common Report-Return Reasons

Bar chart showing rank-ordered return/rejection reason categories.

Table H

Comparable Fields In BR and Manifest Data

Suggested columns:

Concept	BR field(s)	Manifest field(s)	Degree of comparability	Notes

Table I

BR and Manifest Quantity Comparisons

Suggested rows:

- number of handler-years with BR shipped tons and manifest shipped tons;

- mean/median BR shipped tons;
- mean/median manifest shipped tons;
- mean/median absolute difference;
- mean/median percent difference;
- number and percentage with no matching manifest records.

Parallel rows will be included for received tons.

Statistical Analyses

This section describes the statistical models and hypothesis tests that will make up the analysis — including any follow-ups on effects in the main statistical model and any exploratory analyses that can be anticipated prior to analysis.

Exploratory Analysis

We may conduct exploratory analyses to generate additional insights for EPA. These may include:

1. **Additional subgroup analyses**
    - Report rejection rates by state, industry sector, NAICS code, waste code group, management method, or waste property.
2. **Alternative definitions of report-return outcomes**
    - Compare quantity generated by report_rejected in 2025 against the incidence of report returns in the logging data[16]
3. **Alternative validation encounter definitions**
    - Testing different thresholds for repeated validation runs, such as more than two industry validation runs.
    - Comparing logging data validation encounter indicators with presence in validations data (subject to the ability to matching the datasets).
4. **Text analysis of explanations and comments**
    - More detailed natural language processing, keyword extraction, or topic modeling of ISSUE_EXPLANATION and COMMENTS, if data volume and quality support this.
5. **Manifest-BR discrepancy analysis**
    - Identifying characteristics associated with larger differences between manifest quantities and BR quantities (subject to the ability to match the datasets).
    - Examining whether discrepancies are more common for liquid wastes, volumetric units, certain waste codes, or certain management methods (subject to the ability to match the datasets).
6. **Electronic manifest use**

[16] Where a given preparer appears as a row in the logging data (and where "user type" = regulatory), this may infer that a BR submission was returned to the preparer for changes, and then was validated again before the preparer's submission was accepted. We can only obtain logging data for the 2025 BR.

- Using the SUBMISSION_TYPE variable, we will examine whether higher electronic manifest use is associated with lower report rejection rates, fewer validation justifications, or smaller BR-Manifest quantity differences.

7. **Longitudinal preparer patterns**
    - Describing whether the same preparers repeatedly experience report returns, validation justifications, or large BR-Manifest discrepancies across cycles.
8. **Potential BR coverage gap analysis**
    - Using Manifest data, we may be able to identify facilities that potentially should have submitted a BR in a given reporting year but do not appear in the BR data. However, this analysis will be exploratory, as Manifest data alone cannot reliably establish BR reporting eligibility. In particular, using the quantities of hazardous waste shipped or received to infer reporting requirements has several limitations:
        i. **Shipment quantities do not equal waste generated:** The quantity shipped during a reporting year may differ from the quantity generated during that year, as facilities may accumulate waste for later shipment or ship waste generated in a previous year.
        ii. **Incomplete coverage of waste management activities:** Manifest data captures documented shipments of hazardous waste but do not necessarily capture all hazardous waste generated or managed on-site. Consequently, shipment quantities may not accurately reflect a facility's total waste generation or management activities.
        iii. **Monthly generation thresholds:** LQG status depends on the quantity of hazardous waste generated in each calendar month, rather than the total quantity shipped over a reporting year. Annual shipment totals therefore cannot reliably establish whether a facility met LQG thresholds.
        iv. **Differences in reporting requirements:** Federal and state reporting requirements may differ, and certain facilities or waste streams may be subject to exemptions or alternative requirements.
        v. **Incomplete or inconsistent identifiers:** Missing or invalid EPA IDs in manifest records may prevent reliable matching to BR handlers, while differences in reporting coverage may result in apparent rather than actual gaps.
    - Accordingly, unmatched facilities identified through this approach will be treated as potential BR reporting candidates, not confirmed cases of missing reports or noncompliance. Where feasible, we will use additional information on generator status, facility regulatory status, and applicable reporting requirements to assess the plausibility of BR eligibility.
9. **Exploratory analysis of validation thresholds (descriptive regression discontinuity type plot)**
    - For selected validation rules with explicit quantity thresholds, we may plot report return rates against the distance between the reported quantity change and the threshold for quantity discrepancies that trigger a validation threshold.[17] For example, for a rule requiring a difference from the previous reporting cycle of both more than 20% and more

[17] There are currently six validations associated with quantity thresholds in the context of amount of waste shipped, received, generated, or managed on site (see specific validations on RCRAInfo here).

than 1,000 tons, calculate the distance from the effective threshold as the absolute change in quantity minus the greater of 20% of the previous cycle's quantity or 1,000 tons.

- Examine whether return rates change discontinuously around the validation error threshold threshold. This analysis will be exploratory and descriptive; we will not estimate a formal regression discontinuity model.
- This analysis may help us assess whether reports just above a validation threshold are more likely to be returned than reports just below it, providing insight into whether specific validation rules are associated with report returns.

Limitations

The study is subject to several limitations and practical constraints, including those related to data availability, data quality, and analytical feasibility, which are discussed below.

1. **Limited observability of validation warnings and errors**

The data does not directly capture every validation warning or error displayed to preparers. The validations dataset captures instances where a BR preparer provided a written explanation in response to a warning, but not cases where the preparer corrected the data instead of providing an explanation. The logging data provide a partial proxy for validation encounters in 2025 only, but repeated validation runs may occur even when no warning or error was displayed. As a result, estimates of validation encounters are approximate and may overstate or understate the true number of preparers who saw validations.

2. **No logging data for 2021 or 2023**

EPA does not retain logging data beyond two years, so logging-based measures of likely validation encounters are only available for the 2025 reporting year. This limits the ability to compare validation encounter rates over time.

3. **State variation in use of RCRAInfo**

Some states did not use RCRAInfo for BR submission during all or part of the study period. Validation-related analyses will not capture preparers in states that used separate state systems. State transitions into RCRAInfo during the study period also complicate comparisons over time.

4. **State variation in reporting requirements and review practices**

States vary in reporting requirements, annual reporting practices, data systems, review processes, and criteria for returning reports. Observed differences in report return rates may reflect differences in review practices rather than differences in preparer data quality.

5. **BR and Manifest data are not fully equivalent**

BR data and Manifest data serve different purposes and are collected at different levels of aggregation. BRs summarize annual hazardous waste generation and management activity, while manifests track individual off-site shipments. Some waste reported in the BR may not have a corresponding manifest, such as waste managed on site or waste exempt from submitting manifests. Conversely, manifest records may not map cleanly to BR waste streams.

6. **Quantity comparisons may be affected by unit conversions and density assumptions**

Manifest and BR quantities may differ because of unit conversions, density assumptions, estimates versus measured quantities, or differences in how waste streams are grouped. This is especially relevant for liquids or wastes reported in volumetric units.

7. **Potential linkage challenges**

Linking BR and Manifest datasets depends on the availability of the relevant identifiers -HANDLER_ID (BR) and GENERATOR_ID (Manifest). Where possible, we will attempt to find matches in the Manifest data for all BR preparers, but issues such as missing IDs or other issues may limit our ability to do so.

8. **Report returns may not capture all correction activity**

The review data capture formal return or rejection actions, but may not capture all informal communication or corrections between regulators and preparers. Some corrections may occur before formal submission or outside the structured data.

9. **2025 data availability and finalization**

The 2025 BR data may not be finalized until later in the project timeline. Analyses using preliminary 2025 data may change after EPA completes review and finalization.

10. **BR data may not represent the full universe of LQGs**

The BR data reflect facilities that submitted reports, not necessarily all facilities that met LQG thresholds during the reporting year. Some facilities may misclassify their generator status, misunderstand requirements, fail to recognize threshold crossings, or be subject to different state requirements. As described above in 'potential BR coverage gap analysis' under Exploratory Analysis, we may conduct an exploratory analysis to identify facilities who potentially should have submitted a BR given reporting year and did not, but this will not be fully reliable.