Technical Appendix F

Summary of Differences between RSEI Data and the TRI National Analysis

Table of Contents

1 Introduction

The RSEI model currently uses a data set that EPA uses in publishing the annual Toxics Release Inventory (TRI) National Analysis. This data set is known as the National Analysis data freeze, as it represents a static reference point for the TRI data, which is continually updated and revised. Currently, RSEI uses the National Analysis data freeze. However, both the National Analysis and RSEI revise the data to increase its accuracy, and some of the revisions are performed for one dataset and not the other. The following sections discuss each of the types of the revisions made, and how they differ between RSEI and the National Analysis.

2 Off-site Facility Consolidation

In TRI, transfers to off-site facilities are reported by the facility transferring the chemical. Therefore, TRI contains multiple entries for the same off-site facility. Often, these entries are reported slightly differently by different facilities; for instance, one entry may be Columbia POTW at 1214 Main St., while the same off-site facility may have entry of Columbia Water Treatment Plant at 1214 Main Street. TRI lists entries like these separately. For modeling purposes, RSEI compares the text of all of the off-site facility entries to determine which entries actually represent the same facility, ultimately resulting in a set of unique off-site facilities. However, due to the large size of the original set, the comparison must be done automatically through a computer program, which is not as accurate as using human judgement. RSEI errs on the conservative side to avoid false matches, instead accepting a certain level of non-unique facilities to remain in the final set. For details on the method used, please see Technical Appendix D.

3 Off-site and On-site Facility Latitude/Longitude Revisions

Facilities reporting chemical releases to TRI no longer report their latitude and longitudes. RSEI uses EPA's Facility Registry Service (FRS) as the primary source of locational data. Additionally, RSEI performs manual verification of high-scoring facilities, and in some cases adjusts the coordinates based on satellite data. TRI does not contain coordinates for off-site facilities. It should be noted, however, that in geocoding facilities the quality of the coordinates varies, from an exact match based on street address with verification based on satellite data, to a match based only on a 5-digit ZIP code. See Technical Appendix D for details on the deriving the locational data for both reporting facilities and off-site facilities.

4 Adjustments for Double-Counting

TRI facilities must report any chemicals that are transferred off-site to other facilities. These recipient facilities can dispose of the wastes in various ways, some of which are modeled by RSEI. Some of the off-site facilities that receive chemicals from TRI facilities are treatment, storage and disposal (TSD) facilities regulated under the federal Resources Conservation and

Recovery Act (RCRA). These facilities were required to report their own releases to TRI for the first time in 1998. The new reporting requirement means that there is the potential for doublecounting releases that are transferred by a TRI facility to an off-site facility (and so reported to TRI), then released by a RCRA-regulated TSD (and reported again). The National Analysis uses facilities' RCRA identification numbers to match releases reported by TSD facilities to off-site transfers from other TRI reporters, and omits the matching off-site transfer from the summary. RSEI does not use this same matching routine, but does drop transfers to incineration after 1998 where the receiving facility is also a TRI reporter in the North American Industry Classification System (NAICS) code 562211 (Hazardous Waste Treatment and Disposal).

5 Selection of Industry Classification Codes

5.1 Introduction

The consideration of industrial sectors is an essential component of RSEI. The foremost reasons are that the 6- and 4-digit primary NAICS codes for a facility is used to estimate the stack air modeling parameters for those facilities for which facility-specific information is not available, and 4-digit NAICS codes are used to estimate chromium speciation.

When submitting Form R reports by chemical, facilities are asked to always list the primary NAICS code for the entire facility first, and then list up to five additional 6-digit NAICS codes for other "establishments" (defined as "distinct and separate economic activities [that] are performed at a single physical location") which are associated with reportable releases and other waste management and source reduction activities. At least one Form R report is required for each chemical, but some facilities report multiple Form R reports for a single chemical to reflect the activities involving a TRI chemical at each establishment or group of establishments. This could lead to more than six 6-digit NAICS codes being reported for a single facility.

Facilities sometimes commit errors in following these instructions and in providing valid NAICS codes. When a facility reports more than one primary NAICS code (based on all of its submitted Form R reports), RSEI performs a frequency analysis to assign a single primary NAICS code to that facility. This analysis is performed on the "frozen" data set that is prepared by EPA prior to each year's National Analysis. This data set represents not only the current year of reporting, but also any modifications or data corrections for all previous years.

5.2 Processing TRI Data

For each new year of annual release data (as reported in the TRI National Analysis), the RSEI methodology examines the current Form R reports for each TRI reporting facility (or those reports in the last year of reporting for facilities no longer reporting to TRI). There are normally over a million latest-year submissions, with a smaller number of submissions with NAICS code information on them. The set of latest-year submissions with NAICS code information is examined for the purposes of assigning a primary NAICS code to individual facilities (either current reporters or historical reporters). Some small number of these submissions (usually less

than 1%) have invalid primary codes (shown by the code INVA in the TRI data). Most facilities report only one NAICS code. The rest report multiple NAICS codes, but designate one code as the facility's primary NAICS code. A few facilities report more than one primary NAICS code. In these cases, RSEI performs a frequency analysis and selects the code most often designated as primary to be that facility's primary NAICS code. Similarly, if more than five additional NAICS codes are submitted, RSEI selects the five most frequently submitted (this is very infrequent).

Therefore, the main difference between RSEI's treatment of NAICS codes and that of the National Analysis is that in cases where more than one primary codes is reported, RSEI selects the most frequently reported.