DRAFT

"README.DOC" File
for the Quarterly Data Extract (QDE)
from the
FDA ADVERSE EVENT REPORTING SYSTEM (FAERS)

U.S. FOOD AND DRUG ADMINISTRATION (FDA)
CENTER FOR DRUG EVALUATION AND RESEARCH (CDER)
Office of Surveillance and Epidemiology (OSE)

Revised April 2013

IMPORTANT: This document describes significant changes resulting from FDA's transition from Legacy AERS (LAERS) to the new FDA AERS (FAERS) database at FDA. We have added data to the FAERS database structure and have made minor changes to existing field contents. Users of the QDE should review all of the changes to the new extract before loading the files into their systems.

TABLE OF CONTENTS

A. INTRODUCTION

You are reading the README.DOC file that accompanies the Quarterly Data Extract from the FDA Adverse Event Reporting System (FAERS). FAERS is a computerized database for the spontaneous reporting of adverse events and medication errors involving human drugs and therapeutic biological products. FAERS began on September 10, 2012, and replaced the Adverse Event Reporting System (also known as Legacy AERS), which was decommissioned on August 27, 2012. (For more information on FAERS, please see: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm.)

Each extract covers reports received by FAERS during one quarter of the Year. An exception is the first FAERS extract. It will be a non-standard quarter from August 28, 2012 to December 31, 2012, the time period from the decommissioning of LAERS until the end of the fourth quarter of 2012. This first extract may also include Periodic reports from the July/August 2012 time period, as FDA had not yet entered these reports into LAERS when the transition to FAERS occurred.

The data are provided in two distinct formats in the extract:

1. ASCII files, in which data elements are separated from each other by a

'$' sign ("$ delimited"). Please refer to the ASC_NTS.DOC file for additional information on this format.

2. XML files conforming to the guidelines of the International Conference on Harmonisation (ICH) concerning transmission of Individual Case Safety Reports (ICSR). Please refer to the XML_NTS.DOC file for additional information on this format.

Although an effort has been made to make the ASCII consistent with the XML output file, some of the data elements represented in XML are not represented in ASCII, and a few of the data elements represented in ASCII are not represented in XML. This is due to the very different natures of the two formats and to the fact that ICH E2b/M2 specifically defines the allowable data elements for XML. Furthermore, neither the ASCII nor the XML is intended to include all possible data fields. We expect that as the international standard evolves and as electronic submission of reports becomes a reality, additional data elements will probably be added to the XML format.

If you wish to obtain individual case reports through the Freedom of Information Act (FOIA), please refer to these reports by the case report number only. The request must identify each case report you are interested in receiving. The case report number is included in the field/data element labeled: (1) “CASE” in the AERS ASCII format, (2) "CASE_ID" in the FAERS ASCII format (data files starting the 4th quarter 2012), or (3) “safetyreportid” in the SGML/XML format of the quarterly data files.

B. CLINICAL CAVEATS

Disclaimer: Submission of a safety report does not constitute an admission that medical personnel, user facility, importer, distributor, manufacturer or product caused or contributed to the event. The information in these reports has not been scientifically or otherwise verified as to a cause and effect relationship and cannot be used to estimate the incidence of these events.

There are some important things to keep in mind when reviewing or analyzing FAERS data:

For any given report, there is no certainty that a suspected drug caused the reaction. This is because physicians are encouraged to report suspected reactions; however, the event may have been related to the underlying disease being treated, or caused by some other drug being taken concurrently, or simply occurred by chance at that time.

Accumulated reports cannot be used to calculate incidence (occurrence rates) or to estimate drug risk. Comparisons between drugs cannot be made from these data.

C. CAVEATS for Users who are Converting from LAERS to FAERS

The following notes are intended to assist users in transitioning their IT systems from LAERS to FAERS. Users should take into account the following IT and data changes while importing the new FAERS QDE files:

While the previous LAERS database was Individual Safety Report (ISR) based, the new FAERS database is Case/Version based. In LAERS, a Case consisted of one or more ISRs (that is, Initial and Follow-up reports), and each ISR number represented a separate version of a case. So a case could contain multiple ISRs within it, and the latest ISR represented the most current information about a particular case (for example, Follow-up 1 would have the most up-to-date information about a case containing an Initial ISR and a Follow-up 1 ISR). However in FAERS, the ISR concept is no longer being employed, and instead, each unique submission of a case received is given a version number (for example, Case# 1234567, Version 1). The first Version received (formerly called Initial) will be version 1; the second Version received (formerly called Follow-up 1) will be version 2, and so on. The latest version of a case represents the most current information about a particular Case. As a result, the FAERS QDE output files will always provide the latest, most current, Version of a Case available at the time the QDE is run.

It is possible that a case being updated during a specific quarter to add a subsequent version to the Case will not appear in the QDE output for that particular quarter if it is not yet completed when the QDE output files are prepared. This case will appear in a subsequent output, most likely the following quarter. We do not expect this scenario to affect a significant number of cases each quarter. Another scenario may occur where a version 2 of a case may result in not receiving a version 1 of the case. This would occur if version 1 and version 2 are both received and updated within a quarter. In that case, only the latest, most current, version 2 will appear in the QDE extract. We expect this scenario to happen infrequently.

In exceptional circumstances, cases may not show up in the extract in which they are expected. The reason is the switch to a Case/Version based system and the version update process. In this situation the cases will show up in the next QDE extract. In each version of a case, users can search for when the initial version of a case was received by FDA by using the <receiptdate> XML tag or the INIT_FDA_DATE ASCII tag.

While Legacy extracts provided prior to FAERS forced day or month data (for example, 201202 = 20120201 or 2012 = 20120101) where none was provided, FAERS extracts will provide only the exact data submitted in the case (for example, 201202 = 201202 or 2012 = 2012). Therefore, partial dates may be present in certain date fields (for example, in XML: reactionstartdate, reactionenddate, drugstartdate, drugenddate – in ASCII: FROM_DT, TO_DT, START_DT, END_DT). We did this to increase data quality, as previously there was a loss of accuracy in some scenarios.

In the old Legacy extract, we had a field DOSE_VBM containing the verbatim text for dose, frequency, and route (if provided), exactly as they were entered on the report. In the new FAERS extract, we have added new fields (that is, DOSE_AMT, DOSE_UNIT, DOSE_FREQ, and ROUTE) that capture the individual data elements for the dose, frequency, and route and will be provided in addition to the DOSE_VBM field. The DOSE_VBM field will only be displayed if data cannot be completely captured in the separate distinct dose, frequency, and route fields.

In general we made the new extract the same as the old; however these are some of the few exceptions. It is important to note that the CONFID field will no longer be included in the ASCII extract because no matter what is provided in the Case, FDA does not share this information with the public. We also removed the IMAGE field. We removed the field DEATH_DT (Date patient died) to comply

with privacy guidelines. You should account for these changes if you were previously importing these data elements into your database.

The transition to the new FAERS dictionaries will result in some minor data changes in the Medicinal Product values. For example, in the <medicinalproduct> tag the FAERS dictionary upgrade may result in minor differences in values in some cases (for example, LAERS=BUPROPION HCL, FAERS= BUPROPION HYDROCHLORIDE).

In the ASCII Report Source file, the source from the latest version of a case will be provided in the RPSR_COD field. In the FAERS extract, if there is more than one report source, all report sources will be displayed in a single-line listing, separated by commas. This is different from the LAERS extract, in which each report code appeared on a separate line. See example below.

In the LAERS extract:
ISR$RPSR_COD
8024929CSM
8024929FGN
8024929OTH

In the FAERS extract:
primaryid$caseid$rpsr_cod
34947572$3494757$CR,HP

In the ASCII Outcome file, all of the outcomes from the latest version of a case will be provided in the OUTC_COD field. In the FAERS extract, if there is more than one outcome, all outcomes will be displayed in a single-line listing, separated by commas. This is different from the LAERS extract, in which each report code appeared on a separate line. See example below.

In the LAERS extract:
ISR$OUTC_COD
8017085OT
8017085HO
8017085DE

In the FAERS extract:
primaryid$caseid$outc_code
34947572$3494757$HO,LT

In the ASCII AGE field, users may occasionally encounter large numeric age values because some manufacturers submit the patient age in DAYS, even for adult patients. For example, a 30-year (YR) old will be displayed as a 10,950-day (DY) old. Therefore, users should note the AGE_UNIT accompanying the large AGE value and account for ages not listed with the year (YR) age unit.

In the XML file, the <duplicatesource> tag will display the coded name of the manufacturer sending report, if available. If there is no coded name, we will display the verbatim name of organization sending the case. Similarly, in the ASCII MFR_SNDR field, the coded name of the manufacturer sending the case will be displayed, if available. If there is no coded name, we will display the verbatim name of the organization sending the case. We are doing this to promote consistency between the ASCII and XML file outputs.

In the ASCII Drug file, there are three fields showing “meaning text” instead of codes. Those fields are RECHAL, DECHAL, and DOSE_FORM. When the extract is run

for the fourth quarter of 2012 (i.e., first FAERS QDE) the Code will be displayed.

D. HOW THE CD-ROM IS ORGANIZED

The main (root) directory on the CD-ROM contains two sub-directories:

1. ASCII, which contains the ASCII data and informational files.

2. XML, which contains the XML data and informational files.

In addition, the main directory contains two general informational files:

1. README.DOC, the informational file you are now reading.

2. SIZEyyQq.TXT, which gives file sizes and record counts for all data files (ASCII and XML) in the extract. (Example: SIZE97Q4 gives file sizes and record counts for the fourth quarter of 1997.)

E. FILE NAME NOMENCLATURE

In the ASCII format, file names have the format <file-descriptor>yyQq, where <file-descriptor> is a 4-letter abbreviation for the data source, 'yy' is a 2-digit identifier for the year, 'Q' is the letter Q, and 'q' is a 1-digit identifier for the quarter. As an example, the ASCII demographic file for the 4th quarter of 1997 is represented as DEMO97Q4.

(The set of six ASCII data files in each extract contains data for the full quarter covered by the extract. These files contain demographic, drug, reaction, outcome, report source, and drug therapy date information, respectively.)

In the XML format, file names have the format ADRyyQq, where 'yy' is a 2-digit identifier for the year, 'Q' is the letter Q, and 'q' is a identifier for the quarter. As an example, the XML file for fourth quarter of 2012 is named ADR12Q4.

F. ASCII FILES

ASCII Data Files:

The ASCII data files are '$' delimited; that is, a '$' is used to separate the data fields. These files can be imported into SAS or into Access or other database program. Some data files (especially DRUGyyQq and REACyyQq) may exceed the maximum number of records that can be imported into spreadsheet programs such as MS Excel (for example in Office 2003, Excel has a limit of 64,000 rows).

1. DEMOyyQq.TXT contains patient demographic and administrative information, a single record for each event report.

2. DRUGyyQq.TXT contains drug/biologic information for as many medications as were reported for the event (1 or more per event).

3. REACyyQq.TXT contains all "Medical Dictionary for Regulatory Activities" (MedDRA) terms coded for the event (1 or more). For more information on MedDRA, please contact: TRW, VAR 1/6A/MSSO, 12011 Sunset Hills Road, Reston, VA 20190-3285, USA; website is www.meddramsso.com

4. OUTCyyQq.TXT contains patient outcomes for the event (0 or more).

5. RPSRyyQq.TXT contains report sources for event (0 or more).

6. THERyyQq.TXT contains drug therapy start dates and end dates for the reported drugs (0 or more per drug per event).

7. INDIyyQq.TXT contains all "Medical Dictionary for Regulatory Activities" (MedDRA) terms coded for the indications for use (diagnoses) for the reported drugs (0 or more per drug per event).

ASCII Informational Files:
---
1. ASC_NTS.DOC shows in some detail the organization and content of the ASCII data files.

2. STATyyQq.TXT gives null (that is, no data) counts and frequency counts for selected fields in the ASCII datasets. (The frequency counts also include the number of null values; however, the percentages shown are for non-null values only.)

G. XML FILES

XML Data File:

ADRyyQq.XML, data in a format generally conforming to the standards of the International Conference on Harmonization (ICH) for the transmission of adverse reaction reports (E2b) and its technical implementation (M2). While the previous LAERS XML files were provided in 3 monthly increments, the new FAERS XML extract will contain the entire quarter of available data in one file.

XML Test Data File:

ADR_TEST.XML, a small sample (about 1000 reports) of XML data that you could use for learning or testing purposes, before trying ADRyyQq.XML.

XML Informational File:

XML_NTS.DOC, which gives an introduction to the E2b standard and detailed notes about its data field descriptions.

H. REVISION HISTORY

April 2013 (fourth quarter 2012 data)

This is the first extract from FAERS and contains data from August 28, 2012 to December 31, 2012. LAERS was shut down on August 27, 2012, and FAERS was deployed on September 10, 2012. Because of this transition, the initial extract is slightly larger.

For LAERS revision history details, refer to files from previous extracts, available at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm.

I. QUESTIONS, COMMENTS

Questions or comments may be directed to the Food and Drug Administration, Center for Drug Evaluation and Research, Office of Surveillance and Epidemiology (301-796-2360 or cderosetracking@fda.hhs.gov).