

## Combining Electronic Diary Seizure Data with Visit-Based Seizure Data in the MONEAD Study

Julia Skinner and Ryan May, Emmes Corporation

### ABSTRACT

Electronic diary (eDiary) mobile applications can be a useful tool to collect subject-reported data outside of scheduled study visits. However, integrating eDiary data with data stored in a traditional electronic data capture (EDC) system can present unique challenges. eDiary applications sometimes function under different business rules than traditional EDC systems. Collecting data in different software systems makes linkage of records more difficult, even if a common key exists. This paper describes methods used in the Maternal Outcomes and Neurodevelopmental Effects of Antiepileptic Drugs (MONEAD) study to combine daily subject-reported seizure data recorded in an eDiary application with clinically assessed seizure information collected in the study EDC system.

### INTRODUCTION

Electronic subject diaries are a commonly used method to collect subject-reported outcomes (1,2). These instruments allow patients to report data directly and closer to the medical moment, when subject recollection is most accurate (1,2). Electronic subject diary (eDiary) mobile applications are an option that facilitates data collection by subjects and eliminates the need for data entry by clinical staff. These applications have additional benefits over paper diaries, namely the presence of an audit trail and the ability to design the application to improve subject compliance and accuracy (2,3).

There are several challenges to eDiary use from a statistical programming perspective. First, there are often large amounts of missing data. Electronic diaries have been shown to have higher rates of compliance than paper diaries (3,4,5,6), but missing data is still a significant challenge. Data entered by subjects are also more likely to contain data entry errors and inconsistencies than data entered by trained clinic staff. This problem can be mitigated but not eliminated by review of the eDiary data during clinic visits and by the querying of discrepancies by data management staff (1,7). Finally, eDiary applications are often separate software packages with fundamentally different architecture than a study's traditional electronic data capture (EDC) system (2). eDiary applications are designed for frequent data collection and must be usable by subjects with no specialized knowledge of clinical research, while traditional EDC systems are used by clinical research professionals who have undergone training on the software. The incorporation of eDiary data requires careful planning and consideration throughout the entire study.

The MONEAD study is a prospective, observational, multi-center investigation of the effect of anti-epileptic drugs on pregnant women with epilepsy and their children. The study includes control groups of non-pregnant women with epilepsy, as well as healthy pregnant women without epilepsy and their children. Pregnant subjects are enrolled during pregnancy followed until 9 months postpartum, and non-pregnant controls are enrolled for 18 months. A traditional EDC system is used to collect data obtained primarily via clinic visits occurring every three months and medical record abstraction. In addition, daily, subject-reported data are captured through the use of an eDiary mobile application. The MONEAD eDiary is used to collect daily seizure data for both pregnant and non-pregnant women with epilepsy, as well as medication, mood and sleep data for all maternal and non-pregnant subjects. Subjects were given an iPod Touch at enrollment with the MONEAD eDiary application installed, or had the option to install the application on their iPhone or other iOS device. Clinic staff reviewed eDiary entries with the subject at every clinic visit as an adjudication mechanism to ensure optimal eDiary compliance and to ensure that all relevant events were properly recorded.

The effect of pregnancy on seizure activity is a primary outcome measure in the MONEAD study, and the combined data from both traditional EDC and eDiary sources is a critical component of this analysis. The programming required to combine data from both sources had several challenges. These include the linking of records between the two systems, how to handle missing seizure records in the eDiary, and

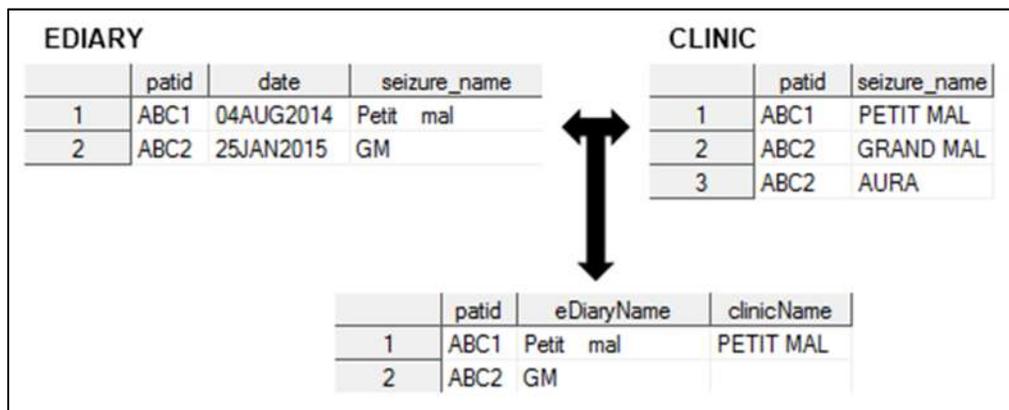
how to treat observations captured via a pop-up window in the eDiary application. The rest of this paper discusses programmatic solutions to each of these issues.

## LINKAGE OF RECORDS

The eDiary used in the MONEAD study is a mobile application developed by a different developer than the main study EDC system (hereafter referred to as “traditional EDC system”). This resulted in some challenges linking the two data sources. When a subject with epilepsy enrolls in the MONEAD study, she is asked about the different types of seizures that she experiences, and instructed to give each type of seizure a unique name of her choosing. This name is entered into the traditional EDC system, and in addition, is entered by the same name in the eDiary, which is used by the subject for reporting daily seizure occurrence. The eDiary software assigns each seizure a unique identification number, but this number is not accessible to the subject and is not stored in the traditional EDC system. Data stored in the traditional EDC system (hereafter referred to as “clinic data”) and eDiary data are entered under a common subject identification number, which facilitates linking of records. However, linkage of detailed seizure classification data collected at clinic visits with the eDiary seizure data requires comparison between free text fields. The SAS® UPCASE and COMPBL functions can be used, respectively, to change all text to the same case and change multiple consecutive blank spaces to a single space:

```
PROC SQL;
  CREATE TABLE seizures AS
  SELECT ediary.patid, ediary.seizure_name AS eDiaryName,
         clinic.seizure_name AS clinicName
  FROM ediary LEFT JOIN clinic
  ON ediary.patid=clinic.patid
     and UPCASE (COMPBL (ediary.seizure_name)) =
     UPCASE (COMPBL (clinic.seizure_name));
QUIT;
```

**Figure 1.** Example of the use of the SQL procedure to join eDiary and clinic data.



Joining data via a free text field is difficult and prone to error. Use of these functions can improve the accuracy to a degree by removing variations in case and spacing. However, data entry errors by the subjects or the clinic staff are common and many cannot be addressed programmatically after the fact. Programmatic identification of discrepancies between the two free text fields began early in the study. This program identifies the seizures reported in the eDiary that cannot be linked to a record in the seizure classification table in the traditional EDC system:

```

PROC SQL;
  CREATE TABLE ediary_only AS
  SELECT ediary.patid, ediary.seizure_name AS eDiaryName,
         clinic.seizure_name AS clinicName
  FROM ediary LEFT JOIN clinic
  ON ediary.patid=clinic.patid
     and upcase(compbl(ediary.seizure_name))=
     upcase(compbl(clinic.seizure_name));
  WHERE clinic.seizure_name IS NULL;

  CREATE TABLE ediary_queries AS
  SELECT ediary_only.patid, ediary_only.eDiaryName,
         clinic.seizure_name AS clinic_names
  FROM ediary_only LEFT JOIN clinic
  ON ediary_only.patid=clinic.patid;
QUIT;

PROC REPORT data=ediary_queries NOWD;
  COLUMNS patid eDiaryName clinic_names;
  DEFINE patid / ORDER "Subject ID";
  DEFINE eDiaryName / ORDER "eDiary Name";
  DEFINE clinic_names / DISPLAY "Name(s) on clinical form";
RUN;

```

**Figure 2.** Output from the REPORT procedure identifying all seizures in the eDiary that cannot be matched to the clinic data. The report also lists all seizure names existing in the traditional EDC system for that subject.

Subject ID	eDiary Name	Name(s) on clinical form
ABC2	GM	AURA
		GRAND MAL

By beginning data cleaning activities early in the study, data linkage problems can be resolved shortly after data is entered, and prior to analysis and data lock, improving data quality.

## IMPUTING MISSING DATA

Daily eDiary data collection often results in large amounts of missing data. In the MONEAD study, many study days had no response entered by a given subject. Separate from the daily eDiary report, the traditional EDC system captured subject responses to questions pertaining to joint review of eDiary data between the subject and clinical site staff. Specifically, subjects were asked at each clinic visit whether they had experienced a seizure since the last visit. This question was used to ensure that all seizures were captured in the eDiary, but it can also be used as a source of data for imputation. If the subject reported no seizures since the last visit in the traditional EDC system, the missing eDiary daily seizure data points can be imputed as “no seizure” for that time period. The LAG function allows for easy identification of the previous visit date:

```
PROC SORT data=clinic_data;
  BY patid visit_date;

DATA to_impute;
  SET clinic_data;
  BY patid visit_date;
  imputeStart=LAG(visit_date);
  imputeEnd= visit_date-1;
  IF seizure_since_last_visit='N';
RUN;
```

**Figure 3.** Example of the use of the LAG function to determine the previous visit date

	patid	visno	visit_date	seizure_since_last_visit		
1	ABC1	01	02JAN2014			
2	ABC1	02	10APR2014	N		
3	ABC1	03	03JUL2014	N		
4	ABC1	04	07OCT2014	Y		
5	ABC1	05	05JAN2015	N		
6	ABC2	01	01JUL2014			
7	ABC2	02	31OCT2014	N		
8	ABC2	03	01FEB2015	Y		

↓

	patid	visno	visit_date	seizure_since_last_visit	imputeStart	imputeEnd
1	ABC1	02	10APR2014	N	02JAN2014	09APR2014
2	ABC1	03	03JUL2014	N	10APR2014	02JUL2014
3	ABC1	05	05JAN2015	N	07OCT2014	04JAN2015
4	ABC2	02	31OCT2014	N	01JUL2014	30OCT2014

The eDiary data is collected daily, while the clinic visits occur approximately every 90 days. The clinic data must be reshaped into daily data in order to combine the two sources. Once the date of the previous visit is identified, a %DO iterative statement allows the user to create one record per day for each subject visit, starting with the last visit date and ending the day before the current visit:

```

PROC SQL NOPRINT;
    SELECT count(*) INTO :numRec FROM to_impute;
QUIT;

DATA impute_all;
    SET _NULL_;
RUN;

%MACRO imputing;
    %DO i=1 %TO &numRec;
        DATA new_records;
            SET to_impute;
            IF _N_=&i;
            DO date=imputeStart TO imputeEnd;
                source='Clinic Data';
                seizureOccurred='N';
                OUTPUT;
            END;
        RUN;

        DATA impute_all;
            SET impute_all new_records;
        RUN;
    %END;
%MEND imputing;
%imputing;

```

**Figure 4.** Partial data set created by the IMPUTING macro. Records are created for subject ABC1 from 02Jan2014 through 02Jul2014, and 07Oct2014 through 04Jan2015. Records are created for subject ABC2 from 01Jul2014 through 30Oct2014.

	patid	date	source	seizureOccurred
1	ABC1	02JAN2014	Clinic Data	N
2	ABC1	03JAN2014	Clinic Data	N
3	ABC1	04JAN2014	Clinic Data	N
4	ABC1	05JAN2014	Clinic Data	N
.				
.				
.				
181	ABC1	01JUL2014	Clinic Data	N
182	ABC1	02JUL2014	Clinic Data	N
183	ABC1	07OCT2014	Clinic Data	N
184	ABC1	08OCT2014	Clinic Data	N
.				
.				
.				
271	ABC1	03JAN2015	Clinic Data	N
272	ABC1	04JAN2015	Clinic Data	N
273	ABC2	01JUL2014	Clinic Data	N
274	ABC2	02JUL2014	Clinic Data	N
.				
.				
.				
393	ABC2	29OCT2014	Clinic Data	N
394	ABC2	30OCT2014	Clinic Data	N

This new data set can be combined with the data in the eDiary to impute seizure data for days where the subject did not enter any seizure information. The approach outlined here also constitutes a quality check of the eDiary data, allowing for identification of subjects whose eDiary data conflicts with the data given at the clinic.

## USING POP-UP WINDOW DATA

One of the advantages of an electronic diary over a paper diary is the ability to prompt subjects to enter data at specific time points. The MONEAD eDiary application includes a pop-up window asking the subject if they had a seizure the previous day. This pop-up appears regardless of whether the subject opens the MONEAD application on their device. While subjects are instructed to open the application and enter detailed seizure information, subjects often fail to comply with these additional data entry steps. In these cases, data from pop-up window becomes a useful data point to determine whether a seizure occurred. Because this pop-up record refers to “yesterday”, additional programmatic steps must be taken to modify the seizure record on the preceding day based off the response to the pop-up. To ensure traceability of records, we re-populate the record ID field with a value that reflects the date of the original record and note the source of the data:

```
DATA yesterday;
  SET ediary;
  WHERE seizure_yesterday IN ('Y','N');

  recordID = 'P' || STRIP(PUT(date, YMMDD6.));
  source = 'Pop-up window';

  date = date - 1;
  seizureOccurred = seizure_yesterday;
RUN;
```

**Figure 5.** Example of pop-up window data being modified for use in analysis

	patid	date	seizure_yesterday
1	ABC1	18AUG2014	Y
2	ABC1	19AUG2014	N
3	ABC1	20AUG2014	N

↓

	patid	date	recordID	source	seizureOccurred
1	ABC1	17AUG2014	P140818	Pop-up window	Y
2	ABC1	18AUG2014	P140819	Pop-up window	N
3	ABC1	19AUG2014	P140820	Pop-up window	N

This data set can be combined with the daily eDiary entries to fill in missing data points. A limitation of the pop-up data is that it does not supply additional detailed seizure data such as type, number or duration of seizures. However, it can still provide some information that would otherwise be missing. To make full use of the benefits of an eDiary, it is important to understand how the software works and what information is being populated in the output data sets.

## CONCLUSION

Electronic diary data is a valuable supplement to visit-based clinical data collection. When an eDiary is used in a study, the statistical programming team should familiarize themselves with how the software

works and begin planning how to combine the data with the rest of the study data. By starting early in the process, this allows sufficient time to resolve discrepancies in the data and resolve any questions about the data structure. eDiary data requires extra effort to incorporate into analysis, but the clinical data collected makes the effort well worth it.

## REFERENCES

1. Food and Drug Administration. 2009. "Guidance for Industry: Patient-Reported Outcome Measure: Use in Medical Product Development to Support Labeling Claims." Accessed March 1, 2018.  
<https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>
2. Paty J. and Stokes, T. 2002. "Electronic Diaries, Part 1: What is a Subject Diary, and How Do Regulations Apply?" *Applied Clinical Trials*.
3. Deshpande, P, et al. 2011. "Patient-reported outcomes: A new era in clinical research." *Perspectives in Clinical Research*, 2(4): 137-144.
4. Stone, AA, et al. 2003. "Patient compliance with paper and electronic diaries." *Controlled Clinical Trials*, 24(2): 182-199.
5. Oystein, D and Hagen, KB. 2007. "Despite technical problems personal digital assistants outperform pen and paper when collected patient diary data." *Journal of Clinical Epidemiology*, 60(1): 8-17.
6. Lauritsen, K, et al. 2004. "Symptom recording in a randomised clinical trial: paper diaries vs. electronic or telephone data capture." *Controlled Clinical Trials*, 25(6): 585-597.
7. Food and Drug Administration. 2013. "Guidance for Industry: Electronic Source Data in Clinical Investigations." Accessed December 12, 2017.  
<https://www.fda.gov/downloads/drugs/guidances/ucm328691.pdf>

## ACKNOWLEDGMENTS

The authors would like to thank the clinical principal investigators, Kimford J. Meador and Page B. Pennell, and all members of the MONEAD Study Group, as well as the women and families who have given their time to participate. The MONEAD study is supported by the National Institute of Neurological Disorders and Stroke (NINDS) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), grant numbers U01-NS038455 and U01-NS050659.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Julia Skinner  
Emmes Corporation  
401 N Washington St, Suite 700  
Rockville, MD 20850  
301-251-1161  
[jskinner@emmes.com](mailto:jskinner@emmes.com)

Ryan May  
Emmes Corporation  
401 N Washington St, Suite 700  
Rockville, MD 20850  
301-251-1161  
[rmay@emmes.com](mailto:rmay@emmes.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.