PharmaSUG 2024 - Paper ST- 199

Demystifying Incidence Rates: A Step-by-Step Guide to Adverse Event Analysis for Novice Programmers

Yuting Peng, Ruohan Wang, Amgen Inc

ABSTRACT

This paper delves into the intricacies of Adverse Event (AE) reporting within the realm of clinical trials, focusing on three distinct methods of treatment-emergent AE incidence rate computation: Incidence Rate (IR), Event Incidence Rate Adjusted by Patient-Years (EIR), and Exposure-adjusted Incidence Rate (EAIR). While Incidence Rate (IR) and Event Incidence Rate (EIR) have traditionally found favor in clinical trial assessments, this study is prompted by a recent information request from the FDA during a submission of a Supplemental Biologics License Application (sBLA), advocating the adoption of Exposure-adjusted Incidence Rate (EAIR) for summarizing AE data. This paper aims to provide a comprehensive understanding of each method by delineating their definitions, comparing their nuances, and contrasting their applications. Furthermore, the article extends its contribution by elucidating the structuring of Analysis Data Model (ADaM) datasets and offering corresponding SAS® code implementations using simulated data, fostering practical insights for researchers navigating AE reporting complexities. Through this exploration, the paper seeks to enhance comprehension, guide methodological choices, and contribute to the evolving landscape of clinical trial data analysis and regulatory compliance.

INTRODUCTION

Incidence rates serve as a critical metric for evaluating the frequency of adverse events within specific populations or among individuals subjected to interventions, including investigational products. The U.S. Food and Drug Administration (FDA), as a pivotal regulatory authority, places significant emphasis on the meticulous reporting of incidence rates. This emphasis arises from the FDA's central role in appraising the safety profiles of drugs, extending beyond procedural requisites to integrate seamlessly into a holistic framework. This framework spans safety assessments, risk-benefit analyses, labeling considerations, risk communications, and regulatory decision-making. In this context, the precise calculation of incidence rates becomes not only a procedural necessity but a paramount factor influencing data quality and subsequent regulatory determinations that profoundly shape drug approval and patient safety. This paper comprehensively explores three distinct methods of computing incidence rates—Incidence Rate (IR), Event Incidence Rate Adjusted by Patient-Years (EIR), and Exposure-adjusted Incidence Rate (EAIR)—shedding light on their implications for effective safety reporting in clinical trials.

These measurements assume a crucial role in comprehending, monitoring, and addressing the occurrence of diseases and health events in clinical trials for several reasons:

Measure of Disease Occurrence:

> Incidence rates quantify the occurrence of new cases of a particular disease or health event within a defined population over a specified period. This measure is fundamental for understanding the burden of diseases in communities.

Risk Assessment in Clinical Trials:

> In clinical research, incidence rates are used to assess the safety and efficacy of interventions. They provide information on the frequency of adverse events or the occurrence of specific outcomes within treatment groups.

Exposure Adjustment:

> Exposure-adjusted incidence rates take into account variations in the duration of exposure to a risk factor, treatment, or condition. This adjustment is important for providing a more accurate estimate of the true risk.

Communication of Health Risks:

> Incidence rates provide a clear and quantifiable way to communicate health risks to the public, healthcare professionals, and policymakers. This facilitates informed decision-making and risk communication.

This paper incorporates a case study that relies on simulated data. All instances within the study are based on identical data assumptions, assuming the availability of ADaM datasets like ADSL and ADAE, which are expected to be prepared and stored in the SAS® library titled "ADAM".

This paper focuses specifically on treatment-emergent adverse events within the safety analysis set. Other summaries related to adverse events, such as those concerning adverse events of special interest, can adopt similar approaches. In this synthetic data case study, treatment-emergent adverse events are defined as those with an onset date occurring on or after the first dose of the study drug and up to 30 days after the last dose of the study drug. While the definition of treatment-emergent adverse events may vary across studies, this is the criterion applied in our analysis. Now, let us delve into the individual concepts step by step:

INCIDENCE RATE (%)

CONCEPT AND FORMULA

The Incidence Rate (IR), sometimes referred to as crude percentage, is a widely used metric for summarizing safety data. IR is calculated by dividing the number of subjects experiencing a specific event by the total number of subjects exposed to the drug, regardless of follow-up duration (Piacentino et al, 2024). It is suitable for uniform treatment durations, short-term exposures, or acute events closely following drug administration. However, for varying exposure durations or long-term follow-ups, IR is inadequate as it neglects the consideration of drug use duration (He et al, 2015). The mathematical formula for IR calculation naturally stems from its definition:

*(NUMBER of SUBJECTS with EVENTS of INTEREST / TOTAL NUMBER of SUBJECTS) * 100%.*

SAS EXAMPLE

We can use the safety population flag (SAFFL) and treatment emergent adverse events flag (TRTEMFL) as crucial filters to extract the targeted analysis records. The NODUPKEY option plays a key role in counting each unique subject under each preferred term (AEDECOD), categorized by actual treatment group (TRT01AN) and system organ class (AEBODSYS). Following the filtering of relevant records, we employ PROC FREQ to obtain the subject count, and the provided code successfully accomplishes this task.:

```
PROC SORT DATA=ADAM.ADAE(WHERE=(SAFFL='Y' AND TRTEMFL='Y'))
          OUT=UNIQUE_AE NODUPKEY;
    BY TRT01AN AEBODSYS AEDECOD USUBJID;
RUN;

PROC FREQ DATA=UNIQUE_AE NOPRINT;
    TABLE TRT01AN*AEBODSYS*AEDECOD / OUT=COUNT(DROP=PERCENT);
RUN;
```

To determine the denominator for calculating the incidence rate, which involves tallying the total number of subjects in each treatment group within the safety analysis set from the ADSL dataset, we again use PROC FREQ. The relevant code snippet is provided below:

```
PROC FREQ DATA=ADAM.ADSL(WHERE=(SAFFL='Y')) NOPRINT;
    TABLE TRT01AN / OUT=BIGN(DROP=PERCENT);
RUN;
```

After obtaining counts for both the numerator and denominator, the datasets are merged. The incidence rate is then calculated using the prescribed formula. To enhance the clarity and organization of the data layout, a PROC TRANSPOSE step is employed:

```
DATA INC_RATE;
    MERGE COUNT(IN=A RENAME=(COUNT=NUM))
          BIGN(IN=B);
    BY TRT01AN;
    IF A;
    INC=ROUND(NUM/COUNT*100,.01);
PROC SORT;
    BY AEBODSYS AEDECOD;
RUN;

PROC TRANSPOSE DATA=INC_RATE OUT=INC_RATE_TR(DROP=_NAME_) PREFIX=INC;
    BY AEBODSYS AEDECOD;
    ID TRT01AN;
    VAR INC;
RUN;
```

Table 1 below illustrates an example of the calculated subject count and Incidence Rate:

TABLE 1. SUBJECT COUNT AND INCIDENCE RATE

System Organ Class	Preferred Term	Subject Count (Incidence Rate %) Treatment Group 1 N=6	Subject Count (Incidence Rate %) Treatment Group 2 N=5
General disorders and administration site conditions	Fatigue	1 (16.67%)	1 (20%)
Musculoskeletal and connective tissue disorders	Arthralgia	1 (16.67%)	2 (40%)
	Pain in extremity	1 (16.67%)	1 (20%)
Nervous system disorders	Headache	1 (16.67%)	0
Skin and subcutaneous tissue disorders	Pruritus	1 (16.67%)	1 (20%)

EVENT INCIDENCE RATE ADJUSTDED BY PATIENT-YEARS

CONCEPT AND FORMULA

Event Incidence Rate Adjusted by Patient-years (EIR) represents the frequency at which specific adverse events occur within a designated population over a defined time frame, relative to the duration of participants' exposure to potential risks (Scosyrev & Pethe, 2022). It is computed by dividing the total number of events by the cumulative patient-years, where patient-years reflect the duration for which individuals contribute data. Person-years are a valuable metric as they account for varying follow-up durations across groups, ensuring a more accurate assessment of risk (Andrade, 2019). EIR is particularly insightful when expecting recurring adverse event(s) throughout the study. In such cases, the event rate offers a more comprehensive perspective on the adverse events profile compared to relying solely on the Incidence Rate.

To calculate EIR, we first define patient-years (exposure time). For subject-level patient-years calculation, the following formula can be used:

(LAST DOSE DATE – FIRST DOSE DATE + 1) / 365.25.

Then, to derive the Events Incidence Rate Adjusted by Patient-years, follow the below formula:

NUMBER of EVENTS / TOTAL PATIENT-YEARS,

Here, the total patient-years is simply the sum of each subject's patient-years.

SAS EXAMPLE

Since only last dose date of study drug (TRTEDT) and first dose date of study drug (TRTSDT) are needed for subject-level patient-years, it can be calculated and stored in ADSL dataset. The following code illustrate how it can be done:

```
DATA ADSL;
    SET ADAM.ADSL;
    IF NMISS(TRTEDT,TRTSDT)=0
          THEN TRTDURY=(TRTEDT-TRTSDT+1)/365.25;
    ELSE IF TRTSDT>.
          THEN TRTDURY=(INPUT("&SNAPDT.",DATE9.)- TRTSDT+1)/365.25;
RUN;
```

In the code above, TRTDURY is the newly created variable for patient year. For ongoing subjects, one practical suggestion is to temporarily consider using the latest data snap date, which can be stored in a global macro variable such as &SNAPDT, as their end of the study date for programming purposes.

To calculate the numerator for Event Incidence Rate Adjusted by Patient-years (EIR), begin by filtering out treatment emergent adverse events (TRTEMFL) associated with subjects in the safety population (SAFFL). Next, sort the ADAE dataset based on System Organ Class (AEBODSYS) and Preferred Term (AEDECOD). Then, utilize PROC FREQ to tally event counts within each treatment group (TRT01AN). The provided code serves as an illustrative example of the specified task:

```
*SORT AND FILTER THE ADAE;
PROC SORT DATA=ADAE(WHERE=(SAFFL='Y' AND TRTEMFL='Y')) OUT=AE;
    BY AEBODSYS AEDECOD;
RUN;

*COUNT EVENTS PER PREFERRED TERM;
PROC FREQ DATA=AE NOPRINT;
    BY AEBODSYS AEDECOD;
    TABLE TRT01AN / OUT=EVENT(DROP=PERCENT);
RUN;

PROC SORT DATA=EVENT;
    BY TRT01AN AEBODSYS AEDECOD;
RUN;
```

To get the denominator, simply sum over the subject-level patient-years that was previously created in ADSL using the code below:

```
PROC SQL NOPRINT;
    CREATE TABLE PY AS
          SELECT TRT01AN,SUM(TRTDURY) AS DENOM
                FROM ADSL
                      GROUP BY TRT01AN;
QUIT;
```

Now that both number of events and total patient-years are calculated, we will merge the two datasets PY and EVENT and use the formula above to get the EIR using the code below:

```
DATA EIR;
    MERGE EVENT(IN=A)
          PY(IN=B);
    BY TRT01AN;
    EIR=COUNT/DENOM*100;
PROC SORT;
    BY AEBODSYS AEDECOD;
RUN;
```

PROC TRANSPOSE can also be used to have a better layout of the resulting data.

```
PROC TRANSPOSE DATA=EIR OUT=EIR_TR(DROP=_NAME_) PREFIX=EIR;
    BY AEBODSYS AEDECOD;
    ID TRT01AN;
    VAR EIR;
RUN;
```

Presented here in Table 2 is the output illustrating the calculated event count and Event Incidence Rate Adjusted by Patient-years (EIR), generated using the provided code above:

TABLE 2. EVENT COUNT AND EVENT INCIDENCE RATE ADJUSTED BY PATIENT-YEARS

System Organ Class	Preferred Term	Event Count (Event Incidence Rate Adjusted by Patient-years) Treatment Group 1 N=6	Event Count (Event Incidence Rate Adjusted by Patient-years) Treatment Group 2 N=5
General disorders and administration site conditions	Fatigue	1 (3.51)	1 (0.76)
Musculoskeletal and connective tissue disorders	Arthralgia	1 (3.51)	2 (1.53)
	Pain in extremity	1 (3.51)	1 (0.76)
Nervous system disorders	Headache	1 (3.51)	0
Skin and subcutaneous tissue disorders	Pruritus	1 (3.51)	1 (0.76)

EXPOSURE-ADJUSTED INCIDENCE RATE

CONCEPT AND FORMULA

To account for potential variations in drug exposure duration, the Exposure-adjusted Incidence Rate (EAIR) should be considered. It is calculated by dividing the number of subjects experiencing an event by the total exposure time of all subjects at risk. For individuals without events, exposure time is from the first drug intake to the last follow-up, while for those with at least one event, it is from the first drug exposure to the event. EAIR represents the average events per unit time of exposure or follow-up, assuming a constant event risk. It is suitable for estimating occurrence rates under the assumption of constant hazard rates. However, if events are delayed or event risks vary over time, EAIR may not be an appropriate measure (He et al, 2015). The formula for the EAIR is:

NUMBER of SUBJECTS with EVENTS / TOTAL PATIENT-YEARS

In contrast to the subject-level patient-years introduced earlier used in EIR calculation, the patient-years (exposure time) are now adjusted based on the specific preferred term being considered. Each subject's patient-years are computed based on the date of their first adverse event (if any) or the end of their study

date. In essence, the determination of each subject's patient-years hinges on whether the subject experienced the adverse event related to the specified preferred term.

For subject with at least one adverse event of interest, the formula is:

(DATE of FIRST EVENT OCCURRED – FIRST DOSE DATE + 1) / 365.25,

for those without, the formula is:

(LAST DATE in STUDY – FIRST DOSE DATE + 1) / 365.25.

Similar to EIR calculation, the total patient-years is then determined by summing all subjects' patient-years.

SAS EXAMPLE

To address variability across preferred terms, one strategy involves generating a placeholder dataset DUMMY. Within this dataset, each unique combination of system organ class (AEBODSYS) and preferred term (AEDECOD) from ADAE is selected based on the criteria of being treatment emergent (TRTEMFL) and originating from subjects within the safety analysis set (SAFFL). This unique combination is then assigned to every subject in the safety population, akin to a Cartesian join, irrespective of whether they experienced a specific adverse event or not. The following code demonstrates how to execute this task effectively:

```
PROC SQL NOPRINT;
    CREATE TABLE DUMMY AS
          SELECT DISTINCT A.USUBJID, B.AEBODSYS, B.AEDECOD FROM
                ADSL A, ADAE B WHERE A.SAFFL='Y' AND B.TRTEMFL='Y';
QUIT;
```

For subjects who experience an adverse event, only the first event that occurred will be considered in the calculation of patient-years. Consequently, we can filter out the first event for each subject under each unique preferred term using the code below:

```
PROC SORT DATA=ADAE(WHERE=(SAFFL='Y' AND TRTEMFL='Y')) OUT=SORTED_AE;
    BY USUBJID AEBODSYS AEDECOD ASTDT;
RUN;

DATA FIRST_AE;
    SET SORTED_AE;
    BY USUBJID AEBODSYS AEDECOD;
    IF FIRST.AEDECOD THEN OUTPUT;
RUN;
```

To ensure non-null values for the event start date in ADAE (ASTDT), we assume that if a complete date is not provided, at least the year when the event started is available for all adverse events in the raw data. Then, the previously created FIRST_AE dataset will be merged with the DUMMY dataset, and all together with ADSL. This merging is done to calculate patient-years for each subject under each unique preferred term. The approach is as follows: if a subject has experienced an adverse event of a specific preferred term, the ASTDT will not be missing, and the patient-years for that record will be calculated using ASTDT and TRTSDT. Conversely, if the subject has not experienced the adverse event, ASTDT will be missing, and the patient-years will be calculated using EOSDT and TRTSDT. Again, for ongoing subject who do not have last dose date yet, we can use the latest data snapshot date as a temporary substitute for their last study date:

```
PROC SQL NOPRINT;
    CREATE TABLE ADAE AS
        SELECT A.*, B.TRT01AN, B.TRTSDT,
            IFN(^MISSING(B.EOSDT),B.EOSDT,INPUT("&SNAPDT",DATE9.)) AS EOSDT
            FORMAT=DATE9. "END OF STUDY DATE",
             C.ASTDT,
             CASE WHEN MISSING(C.ASTDT)
```

```
                THEN (CALCULATED EOSDT - B.TRTSDT+1)/365.25
                ELSE (C.ASTDT-B.TRTSDT+1)/365.25
          END AS SUBJECT_YEAR
          FROM DUMMY A LEFT JOIN ADSL B ON A.USUBJID=B.USUBJID
                       LEFT JOIN FIRST_AE C ON A.USUBJID=C.USUBJID AND
                                               A.AEBODSYS=C.AEBODSYS AND
                                               A.AEDECOD=C.AEDECOD;
QUIT;
```

To obtain the numerator for the Exposure-adjusted Incidence Rate (EAIR), we will utilize PROC FREQ to count occurrences within each preferred term categorized by treatment group. Make sure that only records with non-missing ASTDT are inputted into PROC FREQ when utilizing the ADAE dataset obtained from the previous step, as these records represent the actual adverse event occurrences. The other records in ADAE are merely placeholders. The code is as follows:

```
PROC FREQ DATA=ADAE(WHERE=(^MISSING(ASTDT))) NOPRINT;
    TABLE TRT01AN*AEBODSYS*AEDECOD / OUT=NUM(DROP=PERCENT);
RUN;
```

To determine the denominator for EAIR, the patient-years calculated in the ADAE dataset need to be aggregated by treatment group, System Organ Class (SOC), and Preferred Term (PT). The following code accomplishes this task:

```
PROC SQL NOPRINT;
    CREATE TABLE DENOM AS
          SELECT TRT01AN,AEBODSYS,AEDECOD,SUM(SUBJECT_YEAR) AS DENOM
                FROM ADAE
                     GROUP BY TRT01AN,AEBODSYS,AEDECOD;
QUIT;
```

The following code merges the two datasets and obtain the Exposure-adjusted Incidence Rate (EAIR) using the formula introduced at the beginning of this section, you can add an additional step to transpose the dataset for a more organized layout.

```
DATA EAIR;
    MERGE NUM(IN=A)
          DENOM(IN=B);
    BY TRT01AN AEBODSYS AEDECOD;
    EAIR=COUNT/DENOM*100;
PROC SORT;
    BY AEBODSYS AEDECOD;
RUN;

PROC TRANSPOSE DATA=EAIR OUT=EAIR_TR(DROP=_NAME_) PREFIX=EAIR;
    BY AEBODSYS AEDECOD;
    ID TRT01AN;
    VAR EAIR;
RUN;
```

As an illustration, the calculated incidence rate can be presented in a table format as demonstrated below in Table 3:

TABLE 3. SUBJECT COUNT AND EXPOSURE-ADJUSTED INCIDENCE RATE

System Organ Class	Preferred Term	Subject Count (Exposure-adjusted Incidence Rate) Treatment Group 1 N=6	Subject Count (Exposure-adjusted Incidence Rate) Treatment Group 2 N=5
General disorders and administration site conditions	Fatigue	1 (0.7)	1 (0.67)
Musculoskeletal and connective tissue disorders	Arthralgia	1 (0.58)	2 (1.86)
	Pain in extremity	1 (0.61)	1 (0.65)
Nervous system disorders	Headache	1 (0.65)	0
Skin and subcutaneous tissue disorders	Pruritus	1 (0.62)	1 (0.63)

CONCLUSION

TABLE 4. SUBJECT COUNT/EVENT COUNT/IR/EIR/EAIR

		Treatment Group 1 N=6					Treatment Group 2 N=5				
System Organ Class	Preferred Term	Subject	Event	IR	EIR	EAIR	Subject	Event	IR	EIR	EAIR
General disorders and administration site conditions	Fatigue	1	1	16.67%	3.51	0.7	1	1	20%	0.76	0.67
Musculoskeletal and connective tissue disorders	Arthralgia	1	1	16.67%	3.51	0.58	2	2	40%	1.53	1.86
	Pain in extremity	1	1	16.67%	3.51	0.61	1	1	20%	0.76	0.65
Nervous system disorders	Headache	1	1	16.67%	3.51	0.65	0	0	0	0	0
Skin and subcutaneous tissue disorders	Pruritus	1	1	16.67%	3.51	0.62	1	1	20%	0.76	0.63

Common Theme:

Both EAIR and EIR emphasize the importance of adjusting incidence rates for differences in exposure or follow-up time among individuals in a study population.

Use in Clinical Research:

These adjusted rates are particularly relevant in clinical trials, where participants may have varying lengths of exposure to a treatment or intervention.

Comparative Analysis:

> Adjusted rates, such as EAIR or EIR, are useful for making fair comparisons between different treatment groups or populations, especially when the duration of exposure varies. may be more appropriate when anticipating the occurrence of certain events multiple times throughout the study. EAIR is particularly valuable when anticipating variations in event risks or potential delays.

Precision in Risk Assessment:

> Adjusted rates provide a more accurate estimate of the true risk associated with a specific event by considering the time each individual is at risk.

In summary, while the Incidence Rate provides a fundamental measure of event occurrence, exposure-adjusted rates such as EAIR or EIR enhance this metric by considering variations in exposure duration among individuals in a study population. Notably, EAIR incorporates time directly into the rate calculation, offering a nuanced approach. When applying this perspective to the numbers for the preferred term *Fatigue* in Table 4, Treatment Group 2 exhibits a higher Incidence Rate compared to Treatment Group 1. However, upon considering when the event occurred, the EAIR is actually lower for Treatment Group 2. Thus, although the Incidence Rate might suggest greater severity in Group 2, the EAIR reveals that Group 1 is, in fact, more at risk. These adjustments are vital for a more precise representation of data, playing a crucial role in obtaining accurate and meaningful risk assessments in both clinical and epidemiological studies.

While the FDA does not mandate EAIR use over EIR, justification for displaying only EIR or both may be necessary in study submissions, given the FDA's trendy inclination toward EAIR. The selection between EAIR and EIR also can depend on specific study endpoints and sponsors' preferences.

REFERENCES

Andrade C. Need for and Practical Interpretations of the Person-Year Construct in Neuropsychiatric Research. *Indian J Psychol Med*. 2019 Nov 11;41(6):600-601. doi: 10.4103/IJPSYM.IJPSYM_389_19.

He, X., Chen, L., Lei, L., Xia, H. A., & Lee, M. L. T. 2015. "A Simple Method for Estimating Confidence Intervals for Exposure Adjusted Incidence Rate and Its Applications to Clinical Trials." *J Biom Biostat*, 6: 238. doi:10.4172/2155-6180.1000238.

Piacentino, D., Ogirala, A., Lew, R. *et al*. A Novel Method for Deriving Adverse Event Prevalence in Randomized Controlled Trials: Potential for Improved Understanding of Benefit-Risk Ratio and Application to Drug Labels. *Adv Ther* 41, 152–169 (2024). doi:10.1007/s12325-023-02695-8.

Scosyrev E, Pethe A. Confidence intervals for exposure-adjusted rate differences in randomized trials. *Pharm Stat*. 2022 Jan;21(1):103-121. doi:10.1002/pst.2155.

CONTACT INFORMATION

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

Ruohan Wang
Amgen, Inc.
rwang11@amgen.com

Yuting Peng
Amgen, Inc.
ypeng05@amgen.com