Paper AS22

Multi-dimension Comparison using SAS: in the Context of Counting Hemophilia Bleeding Episode

Qianyun (Chivia) Zou, Sanofi, Cambridge, USA
Wenruo Hu, Sanofi, Beijing, China

ABSTRACT

Hemophilia is a rare genetic bleeding disorder characterized by a deficiency in clotting factors, resulting in prolonged bleeding following injuries or surgical procedures. In clinical research, annualized bleeding rate (ABR) is commonly used primary endpoint to demonstrate clinical benefit in terms of efficacy, which is also acknowledged by health authorities. It is calculated as total number of bleeding episodes divided by duration of analysis period in years.

A bleeding episode is defined as any occurrence of hemorrhage that may require administration of BPA or factor. It differs from a bleeding event, which refers to individual occurrences recorded in raw data -- a single episode may comprise multiple events. Per the various combinations of causality, bleeding location, duration, and injection strategy, the counting of bleeding is a multi-dimension comparison. This complexity poses a significant challenge in hemophilia research.

Traditionally, episode identification has relied on hierarchical, record-by-record comparisons. While this method is straightforward, it can be inconsistent and may overlook rare scenarios, leading to potential bias. As hemophilia is a rare blood disorder, the sample size in clinical trials is relatively small. Even a few bleeding episodes may have major impact on the study result interpretation. In contrast, this paper introduces a matrix-based approach that offers a more systematic and robust framework for episode comparison. By organizing data into a structured matrix, this method simplifies the evaluation process, enhances reproducibility, and improves the clarity of clinical insights. The matrix approach not only streamlines data handling but also enhanced the accuracy of bleeding episodes to provide a solid interpretation of treatment outcomes in hemophilia studies.

INTRODUCTION

In the hemophilia research, the primary analysis focused on Annualized Bleeding Rate (ABR), defined as the number of bleeding episodes a patient experiences in a year. The definition of bleeding episode types is based on consensus opinion of International Society on Thrombosis and Hemostasis (ISTH)[1].

The start time of a bleeding episode will be considered the time at which symptoms

of bleeding episode first develop.

1. Bleeding or any symptoms of bleeding at the same location that occurs within 72 hours of the last injection used to treat a bleed or within 72 hours after the start of an untreated bleed at that location will be considered a part of the original bleeding event and will count as one bleeding episode towards the annualized bleeding rate (ABR).

2. Any bleeding symptoms at the same location that begin more than 72 hours from the last injection used to treat a bleed at that location or more than 72 hours after the start of an untreated bleed at that location will constitute a new bleeding event.

3. Any bleeding symptom with two consecutive injections started more than 72 hours apart will be split into two separated bleedings, the latter bleeding start date/time will be imputed with the second injection start date/time.

4. The first injection more than 72 hours after the bleeding start date/time will be considered as a separate bleeding episode with bleeding start date/time imputed with the injection start date/time. The original bleeding will be considered as an untreated bleeding episode.

Counting algorithm regarding types of bleeding episodes:

• Traumatic bleeds that occur in different locations simultaneously (same date/time) in the

same participant will be considered as one bleeding episode.

• Spontaneous bleeds that occur simultaneously (same date/time) in different locations in the same participant will be considered as separate bleeds.

The following content will present two different approaches of counting methods of bleeding episodes, and explain their pros and cons.

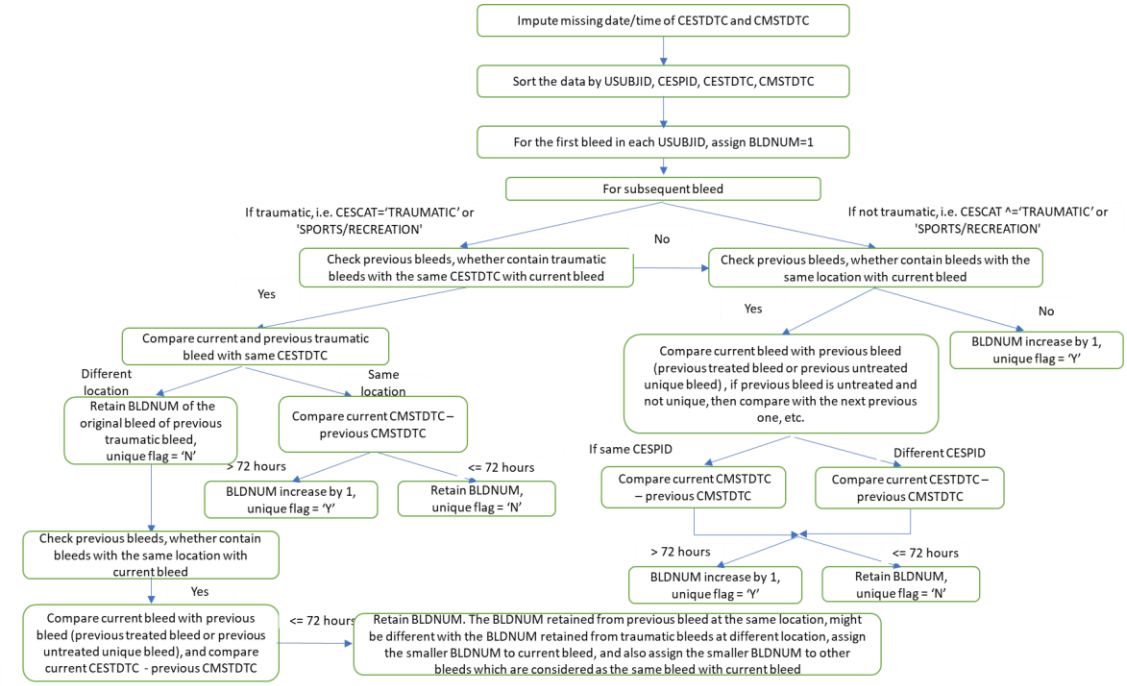
COUNTING METHODS

I. Traditional Approach

With the traditional approach, the algorithm will compare each pair of bleedings according to a hierarchical order. For example, it determines in the order of bleeding type -> bleeding location -> 72 hours rule. After consolidating current pair of bleeding, then compare with previous bleeding episodes.

This approach is easy to understand, but it has an outstanding disadvantage that the comparison chain depends on the sorting order and can only compare the adjacent records (Figure 1 is one example of comparison flow in the traditional approach). Besides, this chain could be very long, taking a lot of programming effort to cover all the possibilities, and not robust enough to handle complex cases.

Figure 1Traditional Approach Workflow



The traditional approach under the flow of figure 1, as it can only compare adjacent records with a fixed hierarchy, CESPID=2 and CESPID=3 was considered different bleed episodes (BLDNUM_T). On the other hand, when using the matrix approach, it compares regardless of the sorting order and hierarchy, and thus CESPID=2 and CESPID=3 was considered as the same bleed episode (BLDNUM_M).

Figure 2 Example Data Case

ASEQ	CESPID	CELOCATION	CESCAT	CESTDTM	CMSTDTM	BLDNUM_M	BLDNUM_T
1	1	A	SPONTANEOUS	2019-01-01T16:30	2019-01-01T17:30	1	1
2	2	B	TRAUMATIC	2019-01-03T00:00	2019-01-03T00:00	2	2
3	3	C	SPONTANEOUS	2019-01-03T00:00	2019-01-03T00:00	2	3
4	4	C	TRAUMATIC	2019-01-03T00:00	2019-01-04T00:00	2	2
5	4	C	TRAUMATIC	2019-01-03T00:00	2019-01-05T00:00	2	2
6	5	F	SPONTANEOUS	2019-01-06T06:00	2019-01-06T06:30	3	4

II. Matrix Approach

To avoid the complexity of the traditional approach, a matrix approach was introduced to compare each pair of bleeding episodes 2x2 without a hierarchical order. It resolved the deficit of the traditional approach which only allows adjacent records comparison with certain order and provides more robust and comprehensive comparison.

The implementation is as below steps:

Step 1: Merge CE and ADCM data together. Create imputed ASTDTM and CMSTDTM.

Step 2: Check "72h" rule within each bleeding event and split bleeding data into the smallest groups.

Step 3: Summarize the group of bleeding data to obtain the "start" and "end" datetime.

Step 4: Use the matrix algorithm to find links among treated bleeds and untreated bleeds separately.

Now we have fixed number of treated bleeding episodes.

Step 5: Repeat the matrix algorithm to find links among all bleeds (between treated bleeds and untreated bleeds).

Matrix algorithm (SAS code in Appendix A)

1) Create a matrix saving all the pairwise links (directly or indirectly) between either of two bleeds.

Now we have a matrix, and each element of the matrix represents connection of 2 bleeds (matrix a).

2) Then, to calculate power of the matrix to find out if there is any further connection. If there are n bleeds for a subject, calculate (n-1) times power of matrix.

2nd power of matrix => find out connection up to 3 bleeds (matrix a2_2).

3rd power of matrix => find out connection up to 4 bleeds (matrix a2_3).

a				
1	0	0	0	0
0	1	0	1	0
0	0	1	1	0
0	1	1	1	0
0	0	0	0	1

a2_2				
1	0	0	0	0
0	2	1	2	0
0	1	2	2	0
0	2	2	3	0
0	0	0	0	1

a2_3				
1	0	0	0	0
0	4	3	5	0
0	3	4	5	0
0	5	5	7	0
0	0	0	0	1

a3				
1	0	0	0	0
0	9	8	12	0
0	8	9	12	0
0	12	12	17	0
0	0	0	0	1

cespid	_BLDNUM	bdtrtfl	celoc	onsetdtm	maxmedtm	BLDNUM
1	1	Y	A	2019-01-01T16:30:00	2019-01-01T17:30:...	1
2	2	Y	B	2019-01-03T00:00:00	2019-01-03T00:00:...	2
3	3	Y	C	2019-01-03T00:00:00	2019-01-03T00:00:...	2
4	4	Y	C	2019-01-03T00:00:00	2019-01-05T00:00:...	2
5	5	Y	F	2019-01-06T06:00:00	2019-01-06T06:30:...	3

3) Eventually, by calculating power of the matrix, we can search connections among all possible combination (matrix a3). All >0 elements are connected to each other.

4) Extract elements that >0 for each row to obtain BLDNUM.

Step 6: Generate final sequential bleeding episode numbers BLDNUM.

CONCLUSION

The matrix approach presented in this paper offers a significant advancement in the classification and counting of bleeding episodes in hemophilia studies. By eliminating the need for hierarchical comparisons and instead leveraging a pairwise matrix-based algorithm, this method simplifies the logic, reduces programming complexity, and enhances scalability. It ensures accurate identification of linked bleeding events across multiple dimensions—such as timing, location, and treatment status—while maintaining

consistency with clinical definitions. This approach not only streamlines the analytical process but also improves efficiency in statistical programming for clinical trials.

REFERENCES

[1] Iorio A, Stonebraker JS, Chambost H, Makris M, Coffin D, Herr C, et al; Data and Demographics Committee of the World Federation of Hemophilia. Establishing the Prevalence and Prevalence at Birth of Hemophilia in Males: A Meta-analytic Approach Using National Registries. Ann Intern Med. 2019 Oct;171(8):540-6

APPENDIX A

Matrix algorithm SAS Code

```
proc iml;
                /* 1. read in temporary dataset with unique bleeding groups */
                use _temp ;
                read all var _all_ where(usubjid="&subj." & onsetdtm ^= . & maxmedtm ^= . );

                n=countn(id);

                /* 2. create shell of n*n matrix filled with 0 */
                a=j(n,n,0);

                /* 3. update each element of the n*n matrix as 1 if the corresponding 2 bleeds can be
linked */
                do i = 1 to n;
                        do j=1 to n;
                                if i=j then a[i,j]=1;
                                else if ( i^=j & (celocation[i,1]=celocation[j,1]) & (onsetdtm[i,1] <=
onsetdtm[j,1]) & ((maxmedtm[i,1] + 72*3600) >= onsetdtm[j,1] ) ) then a[i,j]=1;
                                else if ( i^=j & _cescat[i,1] = _cescat[j,1] & _cescat[i,1] =
'TRAUMATIC/SPORTS/RECREATION' & celocation[i,1]^=celocation[j,1] & onsetdtm[i,1] = onsetdtm[j,1] )
then a[i,j]=1;
                        end;
                end;

                /* 4. calculate n-1 power of the n*n matric to find all possible links and convert to diagonal
symmetrical matrix */
    a2 = 0.5*a + 0.5*t(a);
    a3 = a2 ** (n-1);

                /* 5. Initialize temporary bld matrix - ID + n*1 matrix with inital number as 0 */
                bld0=id||j(n,1,0);

                /* 6. obtain non-zero element from a3 matrix and assign unique number per set of linked
bleeds */

                        /* 6.1 create sequential order for locate element in the matrix - ord should be 1:1
match with original ID*/
```

```
                ord = t(1:n);

                do i = 1 to n;

                /* 6-2 check each row of a3 matrix and obtain >0 IDs as grp */
                        grp=ord[loc(a3[,i]>0),];

                /* 6-3 construct matrix contains all bleed ID that can link to ith bleed, assign i as
group number */
                        bld=grp[,1] || j(countn(grp),1,i);

                /* 6-4 substitute initial bld matrix with the updated bld number, for those checked
once, no need to update again */
                        if bld0[grp,2]=0 then bld0[grp,2]=bld[,2];
                end;
        create bld from bld0 [colname={"ID" "BLD_A"}];
        append from bld0;

        close _temp ;

        do;
                if &obsTRT. = 0 then goto TheEnd ;

        /* 1. read in temporary dataset with unique bleeding groups */
                use _temp_trt ;
                read all var _all_ where(usubjid="&subj." & onsetdtm ^= . & maxmedtm ^= .);
                t=countn(trt);

        /* 2. create shell of n*n matrix filled with 0 */
                b=j(t,t,0);

        /* 3. update each element of the n*n matrix as 1 if the corresponding 2 bleeds can be
linked */
                do i = 1 to t;
                        do j=1 to t;
                                if i=j then b[i,j]=1;
                                else if ( i^=j & (celocation[i,1]=celocation[j,1]) & (onsetdtm[i,1]
<= onsetdtm[j,1]) & ((maxmedtm[i,1] + 72*3600) >= onsetdtm[j,1] ) ) then b[i,j]=1;
                                else if ( i^=j & _cescat[i,1] = _cescat[j,1] & _cescat[i,1] =
'TRAUMATIC/SPORTS/RECREATION' & celocation[i,1]^=celocation[j,1] & onsetdtm[i,1] = onsetdtm[j,1] )
then b[i,j]=1;
                        end;
                end;

        /* 4. calculate n-1 power of the n*n matric to find all possible links and convert to diagonal
symmetrical matrix */
                b2 = 0.5*b + 0.5*t(b);
                b3 = b2 ** (t-1);

        /* 5. Initialize temporary bld matrix - ID + n*1 matrix with inital number as 0 */
                Tbld0=id||j(t,1,0);
```

```
                /* 6. obtain non-zero element from a3 matrix and assign unique number per set of linked
bleeds */

                        /* 6.1 create sequential order for locate element in the matrix - ord should be 1:1
match with original ID*/
                        ord = t(1:t);

                        /* obtain non-zero element from b3 matrix and assign unique number per set of
linked bleeds */
                        do i = 1 to t;

                                /* 6-2 check each row of b3 matrix and obtain >0 IDs as grp */
                                        tgrp=ord[loc(b3[,i]>0),];

                                /* 6-3 construct matrix contains all bleed ID that can link to ith bleed,
assign i as group number */
                                        tbld=tgrp[,1] || j(countn(tgrp),1,i);

                                /* 6-4 substitute initial bld matrix with the updated bld number, for those
checked once, no need to update again */
                                        if tbld0[tgrp,2]=0 then tbld0[tgrp,2]=tbld[,2];
                        end;

                        create Tbld from Tbld0 [colname={"ID" "BLD_T"}];
                        append from Tbld0;

                        close _temp_trt ;

                        TheEnd:

                end;
        quit;
```

CONTACT INFORMATION

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

Author Name: Qianyun (Chivia) Zou

Company: Sanofi

Address: 450 Water St, Cambridge, MA 02141

Email: chivia.zou@sanofi.com

Website: www.sanofi.com

Author Name: Wenruo Hu

Company: Sanofi

Email: wenruo.Hu@sanofi.com

Website: www.sanofi.com