SA-138

ESTICID: a SAS macro for Estimating Clinically Important Difference using an anchor-based approach in longitudinal data

Zongshi Qin, Yongpei Yu, Peking University

ABSTRACT

Clinical trial reports increasingly require an interpretable link between a numerical change in a patient-reported outcome measure (PROM) and the change that patients perceive as important. However, anchor-based estimation is often implemented inconsistently, particularly when repeated measurements are incomplete or anchors have more than two ordered levels. %ESTICID is a SAS macro that operationalizes an anchor-based mixed-effects modelling workflow for longitudinal PROM data. The macro accepts a single long-format data set, identifies baseline values, derives change scores, evaluates the association between the anchor and PROM change, and fits a linear mixed-effects model using PROC MIXED. It supports continuous and categorical anchors, user-defined reference categories, direction of improvement, and covariance structures. Outputs include adjusted estimates with 95% confidence intervals and forest plots that distinguish clinically important differences within groups from differences between anchor-defined groups. An illustrative multicenter randomized trial data set is used to demonstrate the workflow. %ESTICID is intended to make clinically important difference estimation more reproducible and easier to implement in SAS-based clinical research.

Keywords: *clinically important difference; patient-reported outcome; longitudinal data; anchor-based method; PROC MIXED; SAS macro*

INTRODUCTION

PROMs are essential for quantifying symptoms, functioning, and health-related quality of life. A statistically significant PROM change is not necessarily meaningful to patients. Clinically important difference (CID) estimation provides a patient-centred interpretive bridge, but a single universal threshold is often inadequate. CID can be considered both within groups—how much change is meaningful to patients—and between groups—whether a meaningful difference is attributable to an intervention beyond a reference group. Anchor-based approaches are clinically intuitive because they connect PROM change to an external indicator such as the Patient Global Impression of Change (PGI-C).

In practice, anchor-based analyses become difficult when data are longitudinal, unbalanced, and collected with multi-level anchors. Analysts must derive baseline change consistently, specify the anchor correctly, account for within-person correlation, and clearly report the estimand. %ESTICID was developed to provide a transparent SAS workflow for these recurring tasks.

METHODS

Analysis framework

For participant i at visit j, the macro models PROM change from baseline as a function of an anchor and a participant-specific random intercept:

$$\Delta PROM_{ij} = \beta_0 + f(anchor_{ij}) + u_i + \varepsilon_{ij}$$

The function f(anchor) is linear when the anchor is treated as continuous and uses indicator variables when it is categorical. The random intercept u_i accounts for correlation among repeated observations from the same participant. The macro excludes the baseline record from the outcome model after deriving change from baseline. Under the usual missing-at-random assumption, linear mixed-effects modelling allows all observed follow-up records to contribute without requiring a complete-case analysis.

Required data and processing workflow

The input data set is in long format, with one record per participant and assessment. It must contain a unique identifier, an ordered time or visit variable, a PROM, and an anchor. The earliest time record is

treated as baseline. Before interpretation, the analyst should assess whether the anchor is clinically relevant and sufficiently associated with PROM change; a weak or clearly non-linear relationship is a reason to reconsider the anchor or treat it categorically.

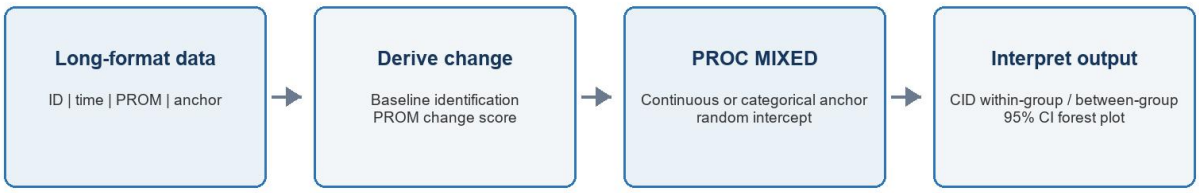

Figure 1. %ESTICID analysis workflow for longitudinal PROM data.

Macro interface

The principal call is intentionally compact. Parameters make the data structure and key clinical decisions explicit rather than hidden in ad hoc programming.

```
%ESTICID(
  data=supp1, id=ssid, time=time, anchor=pgi_c, prom=prom,
categoric_anchor=n, reference=0, covstruct=cs);
```

Table 1. Core %ESTICID macro parameters.

Parameter	Required input	Purpose
data=	Long-format SAS data set	One record per participant per assessment.
id=; time=	Subject identifier; ordered visit/time variable	Sorts data and identifies the earliest record as baseline.
prom=; anchor=	PROM and clinically interpretable anchor	Derives PROM change and links it to perceived change or status.
categoric_anchor=	Y or N	Selects continuous or categorical handling of the anchor.
reference=	Anchor level representing no change	Defines the comparison group for between-group CID estimation.
covstruct=	Covariance structure	Controls the repeated-measures covariance specification in PROC MIXED.

ILLUSTRATIVE APPLICATION

The workflow was illustrated with data from a multicenter randomized controlled trial in chronic prostatitis/chronic pelvic pain syndrome (n=414). NIH-CPSI was the PROM and PGI-C was the anchor. The data set included repeated assessments; PGI-C categories were handled as an ordered categorical anchor, with the no-change category prespecified as the reference. The macro derived each participant's change from baseline, estimated model-based PROM changes by PGI-C category, and displayed estimates with 95% confidence intervals.

This presentation separates two related outputs. The mean change within the minimally improved anchor group is a within-group CID candidate. The contrast between minimally improved and no-change categories is a between-group CID candidate. Reporting both makes the clinical question explicit: the former supports responder interpretation, whereas the latter is useful when interpreting comparative treatment effects. In simulation analyses based on the same longitudinal setting, estimates remained stable as follow-up observations were increasingly missing under a missing-at-random mechanism.

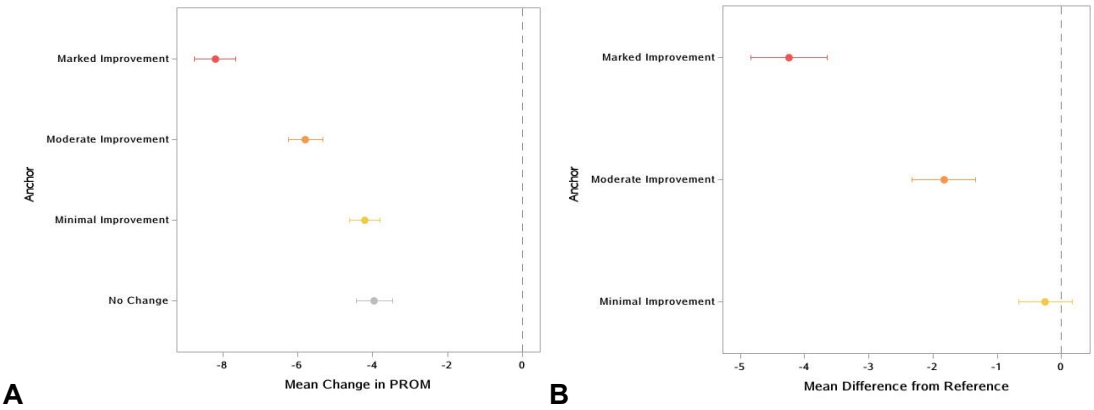

Figure 2. Estimated $CID_{within\text{-}group}$ (A) and $CID_{between\text{-}group}$ (B) for minimal, moderate, and marked categories using estimating clinicallyimportant differences.

PRACTICAL CONSIDERATIONS

The macro does not substitute for clinical judgment. An anchor should be interpretable, directionally coherent with the PROM, and selected before analysis. The reference category should correspond to the decision question—often no change for a between-group contrast. The direction-of-improvement parameter must be checked carefully because some scales improve as scores decrease. Covariance structure should be chosen in line with the visit schedule and sample size. Finally, estimated CIDs should be presented with confidence intervals and contextualized as a range of minimal, moderate, or marked importance rather than treated as an immutable cutoff.

CONCLUSION

%ESTICID offers SAS users a reproducible anchor-based approach to estimating and visualizing clinically important differences in longitudinal PROM data. By combining baseline derivation, mixed-effects modelling, anchor handling, and interpretable graphical outputs in a single macro, it supports transparent reporting of both within-group and between-group CID estimands.

REFERENCES

1. Wu J, Yao L, Liao K, Yorke J, Yu Y, Liu Z, Yao X, Qin Z. Estimating minimal, moderate, and marked clinically important differences for patient reported outcome measure in randomised controlled trial. Chin Med J (Engl). 2026;139(10):1538-1544.
2. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clinic Proceedings. 2002;77:371–383.
3. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Review of Pharmacoeconomics & Outcomes Research. 2011;11:171–184.
4. Twisk J, de Boer M, de Vente W, Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. Journal of Clinical Epidemiology. 2013;66:1022–1028.

CONTACT INFORMATION

Zongshi Qin, MD, PhD

Peking University

qinzs_pucri@hsc.pku.edu.cn

Yongpei Yu, PhD
Peking University
yuyp_pucri@bjmu.edu.cn

APPENDIX SAS MACRO CODE

```
/*****************************************************************
Macro Name          : %CIDER
Purpose             : Estimate and visualize Clinically Important
                      Difference (CID) using an anchor-based approach
                      for longitudinal patient-reported outcome data.
Description         : This macro fits mixed-effects models to estimate
                      changes in PROMs (Patient-Reported Outcome Measures)
                      according to an anchor (e.g., PGI-C, CGI-C) over time.
                      Both continuous and categorical anchor types are supported.
                      It handles missing baseline anchors by imputing 0,
                      and generates correlation statistics, scatterplots,
                      and forest plots of adjusted PROM changes across anchor levels.
Parameters:
 data             = Input dataset name (must contain ID, time, PROM, and anchor)
 id               = Subject or participant ID variable
 time             = Time variable for repeated measurements
 anchor           = Anchor variable (e.g., PGI, CGI, clinician-reported impression)
 prom             = PROM variable to evaluate (e.g., fatigue score, anxiety index)
 categoric_anchor = Y : anchor is categorical (model will use LSMeans)
                    N : anchor is continuous (model will use linear estimates)
   reference      = Select the reference anchor (e.g., reference=4)
 covstruct        = Covariance structure used in the repeated statement
                    (default=UN). Options include:
                    - UN     : Unstructured
                    - CS     : Compound symmetry
                    - AR(1)  : First-order autoregressive
                    - VC     : Variance components
```

```
                    - TOEP   : Toeplitz
                    - FA(n)  : Factor analytic (specify order)
 Output             :
  - Correlation analysis between anchor and PROM
  - Scatter plot with regression line
  - Mixed-effects model estimates of adjusted PROM changes by anchor
  - Forest plot of estimated PROM change (with 95% CI)

 Notes:
  - The macro calculates delta values from baseline.
  - If anchor at baseline is missing, it is imputed as zero.
  - Baseline rows are excluded from modeling.
*********************************************************************/



%macro CIDER(data=, id=, time=, anchor=, prom=, categoric_anchor=, reference=, covstruct=,);

  %let prom_var=Chg_PROM;
  %let anchor_var=Chg_anchor;

  %let anchor_label=Improvment;
  %let anchor_class=;
  %if %upcase(&categoric_anchor)=Y %then %do;
     %let anchor_class=&anchor_var;
     %if %eval(&reference.=) %then %do;
        %let reference=4;
                    %put Reference was not assigned for categorical anchor change, use the
default value 4.;
     %end;
  %end;

    %if %superq(covstruct)= %then %let covstruct=UN;
  %put Var: &prom_var.;
```

```
proc sql noprint;
    select min(&anchor.), max(&anchor.) into :min_anchor, :max_anchor from &data.;
quit;

proc sort data=&data out=_sorted;
    by &id &time;
run;

data _changes;
    set _sorted;
    by &id;
    retain base_prom base_anchor;

    if first.&id then do;
        base_prom   = ifn(missing(&prom.), 0, &prom.);
        base_anchor = ifn(missing(&anchor.), 0, &anchor.);
    end;
    else do;
        Chg_PROM   = &prom. - base_prom;
        Chg_anchor = &anchor. - base_anchor;
        output;
    end;
run;

proc corr data=_changes nosimple;
    var &anchor_var. &prom_var.;
run;

proc sgplot data=_changes;
    reg x=&anchor_var. y=&prom_var. / cli lineattrs=(color=red thickness=2);
    scatter x=&anchor_var. y=&prom_var. / markerattrs=(symbol=circlefilled color=blue size=8);
    xaxis label="Anchor";
    yaxis label="Change in PROM (PROM-C)";
```

```
run;

%if %upcase(&categoric_anchor) ne Y %then %do;
   ods output estimates=est_pgi;
   proc mixed data=_changes;
      class &id. &time. &anchor_class.;
      model &prom_var. = &anchor_var. / solution ddfm=kr;
      repeated &time / subject=&id. type=&covstruct. r rcorr;
      estimate "1" intercept 0 Chg_anchor 1 / cl;
      estimate "2" intercept 0 Chg_anchor 2 / cl;
      estimate "3" intercept 0 Chg_anchor 3 / cl;
   run;
   ods output close;

   data _graph_data_a;
      set est_pgi;
      flag = label*1;
   run;

   data _graph_data_b;
            length cid_class $20;
                  set est_pgi  end=end_flag;
      high=80;
                  low=0;
                  x=estimate;
                  if label="1" then cid_class="Minimal";
                  else if label="2" then cid_class="Moderate";
                  else if label="3" then cid_class="Substantial";
                  output;
                  if end_flag=1 then do;
        high=100;
                    low=0;
                    x_ref=0;
```

```
                    x=.;
                    output;
                   end;
run;

        proc sql noprint;
    select (int(max(abs(Estimate*1.2))/10)+1)*10 into :x_max from _graph_data_b;
    select -1*(int(max(abs(Estimate*1.2))/10)+1)*10 into :x_min from _graph_data_b;
    select Estimate
       into :max_abs_sign
       from _graph_data_b
       having abs(Estimate) = max(abs(Estimate));
          quit;
          %put &x_min. &x_max. &max_abs_sign.;
          %if %sysevalf(&max_abs_sign.<=0) %then %do;
    %let x_max=0;
          %end;
          %if %sysevalf(&max_abs_sign.>0) %then %do;
    %let x_min=0;
          %end;

          data _anno_set;
             length function $8 label $100 X1SPACE $30 Y1SPACE $30 anchor $10;
             set _graph_data_b(where=(x ne .));
                   id="itemtext";
             function = "text";
    label = trim(left(put(x, 8.2)));
                   width = 20;
    widthunit = "percent";
             x1 = x;
             y1 = 82.5;
             X1SPACE="datavalue";
             Y1SPACE="datavalue";
```

```
        anchor = 'center';
textsize = 9.5;
                output;


                id="itemtext";
        function = "text";
label = "CID";
                width = 20;
widthunit = "percent";
        x1 = x;
        y1 = 85.5;
        X1SPACE="datavalue";
        Y1SPACE="datavalue";
        anchor = 'center';
textsize = 9.5;
                output;


                id="itemtext";
        function = "text";
label = cid_class;
                width = 20;
widthunit = "percent";
        x1 = x;
        y1 = 88.5;
        X1SPACE="datavalue";
        Y1SPACE="datavalue";
        anchor = 'center';
textsize = 9.5;
                output;


                id="itemtext";
        function = "text";
label = "Reference";
```

```
                        width = 20;
        widthunit = "percent";
                x1 = 0;
                y1 = 102.5;
                X1SPACE="datavalue";
                Y1SPACE="datavalue";
                anchor = 'center';
        textsize = 9.5;
                        output;

                        keep
           id function label width widthunit x1 y1 X1SPACE Y1SPACE anchor textsize
                        ;
             run;


  %end;


  %if %upcase(&categoric_anchor) = Y %then %do;
     ods output diffs=diff_pgi;
     proc mixed data=_changes;
        class &id. &time. &anchor_class.(ref="&reference.");
        model &prom_var. = &anchor_var. / solution ddfm=kr;
        repeated &time / subject=&id. type=cs r rcorr;
        lsmeans &anchor_var. / cl diff;
     run;
     ods output close;


     data _graph_data_a;
        set diff_pgi(where=(_&anchor_var.=&reference. and (&anchor_var.
in(%eval(&reference.+1) %eval(&reference.+2) %eval(&reference.+3))))) end=end_flag;
                        if _&anchor_var.=&reference. and &anchor_var.
in(%eval(&reference.+1) %eval(&reference.+2) %eval(&reference.+3)) then do;
                    if &anchor_var. = %eval(&reference.+1) then flag=1;
                            else if &anchor_var. = %eval(&reference.+2) then flag=2;
```

```
                        else if &anchor_var. = %eval(&reference.+3) then flag=3;
                    end;
    run;


    data _graph_data_b;
            length cid_class $20;
      set diff_pgi(where=(_&anchor_var.=&reference. and (&anchor_var.
in(%eval(&reference.+1) %eval(&reference.+2) %eval(&reference.+3))))) end=end_flag;
                    if _&anchor_var.=&reference. and &anchor_var.
in(%eval(&reference.+1) %eval(&reference.+2) %eval(&reference.+3)) then do;
        high=80;
                      low=0;
                      x=estimate;
                      if &anchor_var. = %eval(&reference.+1) then cid_class="Minimal";
                      else if &anchor_var. = %eval(&reference.+2) then cid_class="Moderate";
                      else if &anchor_var. = %eval(&reference.+3) then cid_class="Substantial";
                      output;
                    end;
                    if end_flag=1 then do;
        high=100;
                      low=0;
                      x_ref=0;
                      x=.;
                      output;
                    end;
    run;


          proc sql noprint;
      select (int(max(abs(Estimate*1.2))/10)+1)*10 into :x_max from _graph_data_b;
      select -1*(int(max(abs(Estimate*1.2))/10)+1)*10 into :x_min from _graph_data_b;
      select Estimate
        into :max_abs_sign
        from _graph_data_b
        having abs(Estimate) = max(abs(Estimate));
```

```
        quit;
        %put &x_min. &x_max. &max_abs_sign.;
        %if %sysevalf(&max_abs_sign.<=0) %then %do;
%let x_max=0;
        %end;
        %if %sysevalf(&max_abs_sign.>0) %then %do;
%let x_min=0;
        %end;



        data _anno_set;
           length function $8 label $100 X1SPACE $30 Y1SPACE $30 anchor $10;
           set _graph_data_b(where=(x ne .));
                id="itemtext";
           function = "text";
label = trim(left(put(x, 8.2)));
                width = 20;
widthunit = "percent";
           x1 = x;
           y1 = 82.5;
           X1SPACE="datavalue";
           Y1SPACE="datavalue";
           anchor = 'center';
textsize = 9.5;
                output;


                id="itemtext";
           function = "text";
label = "CID";
                width = 20;
widthunit = "percent";
           x1 = x;
           y1 = 85.5;
```

```
        X1SPACE="datavalue";
        Y1SPACE="datavalue";
        anchor = 'center';
textsize = 9.5;
            output;

            id="itemtext";
        function = "text";
label = cid_class;
            width = 20;
widthunit = "percent";
        x1 = x;
        y1 = 88.5;
        X1SPACE="datavalue";
        Y1SPACE="datavalue";
        anchor = 'center';
textsize = 9.5;
            output;

            id="itemtext";
        function = "text";
label = "Reference";
            width = 20;
widthunit = "percent";
        x1 = 0;
        y1 = 102.5;
        X1SPACE="datavalue";
        Y1SPACE="datavalue";
        anchor = 'center';
textsize = 9.5;
            output;

            keep
```

```
            id function label width widthunit x1 y1 X1SPACE Y1SPACE anchor textsize
                        ;
              run;


   %end;


   proc template;
      define statgraph hbar_ci;
         begingraph;
            layout overlay / yaxisopts=(type=discrete label="&anchor_label." offsetmin=0.15
offsetmax=0.15  DISCRETEOPTS=(TICKVALUELIST=("1" "2" "3") TICKDISPLAYLIST=("Minimal
improved" "Moderate improved" "Much improved")))
                           xaxisopts=(label="Mean change") border=false;
               scatterplot y=flag x=Estimate /
                  markerattrs=(symbol=circlefilled color=black size=8)
                  xerrorlower=Lower xerrorupper=Upper
                  errorbarattrs=(color=black)
                  errorbarcapscale=0.75;
            endlayout;
         endgraph;
      end;
   run;
   proc sgrender data=_graph_data_a template=hbar_ci;
   run;



   proc template;
      define statgraph high_low;
         begingraph/border=false;
            layout overlay / yaxisopts=(offsetmin=0 offsetmax=0.05 DISPLAY=none)
                           xaxisopts=(label="Mean change"  LINEAROPTS=(VIEWMIN=&x_min.
VIEWMAX=&x_max.)  offsetmin=0.075 offsetmax=0.075)
                           border=false walldisplay=none;
                                    highlowplot x=x low=Low high=High / lineattrs=(thickness=1
color=gray);
```

```
                                        highlowplot x=x_ref low=Low high=High /
lineattrs=(thickness=1 color=black);
                    annotate/id="itemtext";
               endlayout;
           endgraph;
       end;
   run;
   proc sgrender data=_graph_data_b template=high_low sganno=_anno_set;
   run;


     proc datasets library=work nolist;
      delete
          Diff_pgi Est_pgi _sorted _changes _anno_set _graph_data_a _graph_data_b
                ;
     quit;
%mend CIDER;
```