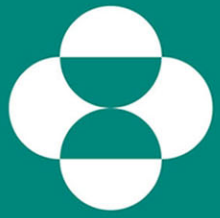
HOW TO DERIVE A LONGITUDINAL ADAM DATASET FOR INVERSE PROBABILITY OF CENSORING WEIGHTING ANALYSIS

Brussels PHUSE SDE 2018



MSD

INVENTING FOR LIFE

Yirong Cao, HTA Statistical Programming

Shahrul Mt-Isa, HTA Statistics

Background

- In many oncology studies, patients who are initially randomized to the control group are allowed to switch over to the treatment group when disease progresses.

- Simple adoption of the original overall survival time or censor for these patients could bias the study results such as underestimating/overestimating treatment effect.

- Inverse probability of censoring weighting analysis (IPCW hereafter) is one of methods to adjust the overall survival time and censoring using weights according to their probability to switch treatment.

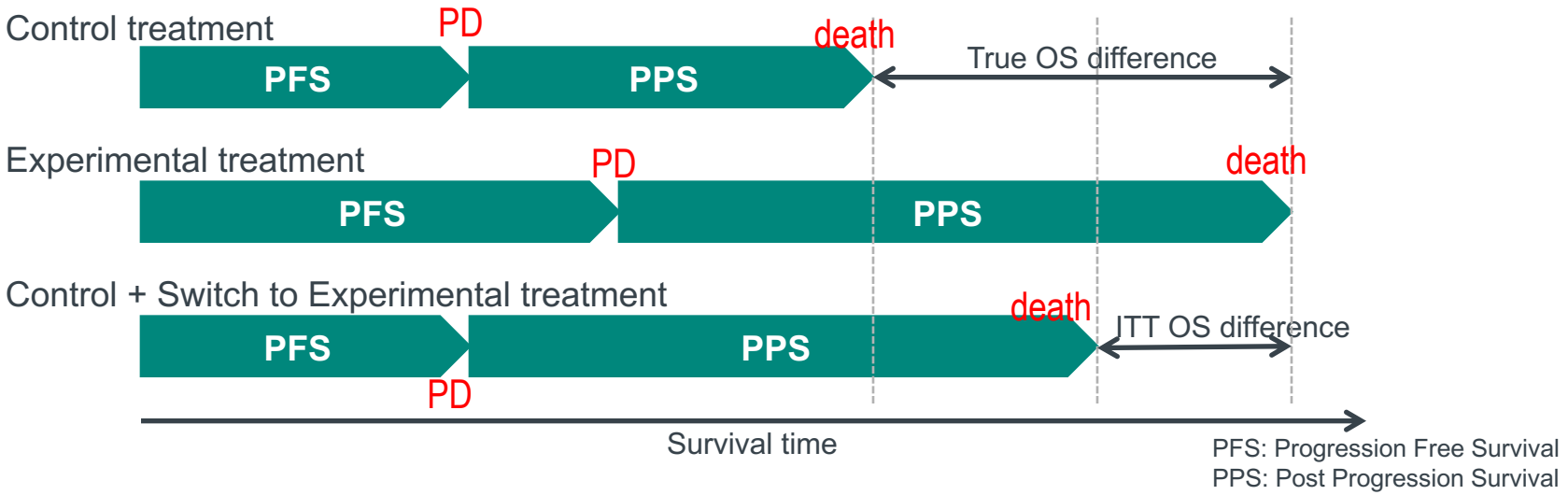
- IPCW analysis requires the longitudinal ADAM dataset to be in specific format. However, there is little literature describing the process of data creation in SAS® to support the implementation of the method.

MSD
INVENTING FOR LIFE

Problem Statement

Endpoint: overall survival (OS)

- Time from randomization to death (any cause)



➡ ITT OS difference underestimates true treatment effect
Treatment switch introduces bias

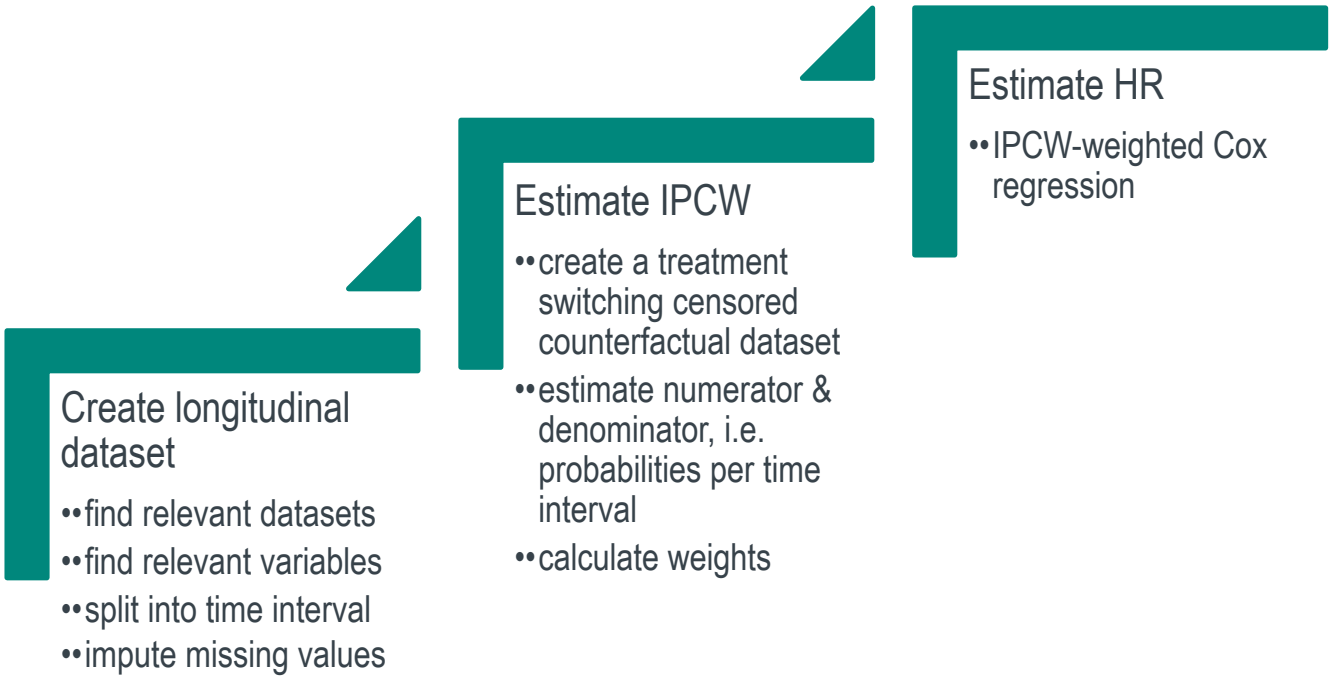
MSD INVENTING FOR LIFE

Overview of IPCW

- Method of handling treatment switchover by introducing **artificial censoring** and **weighting of remaining observations**
- Information after treatment switch is irrelevant
- Historical data ("confounders") characterise subjects
  - Baseline (for stabilisation of weights)
  - Time-dependent
- Confounders (covariates) predict future treatment and future outcome, conditional on past treatment history
- There *must* be "no unmeasured confounders"

MSD
INVENTING FOR LIFE

Procedures of Implementation



MSD INVENTING FOR LIFE

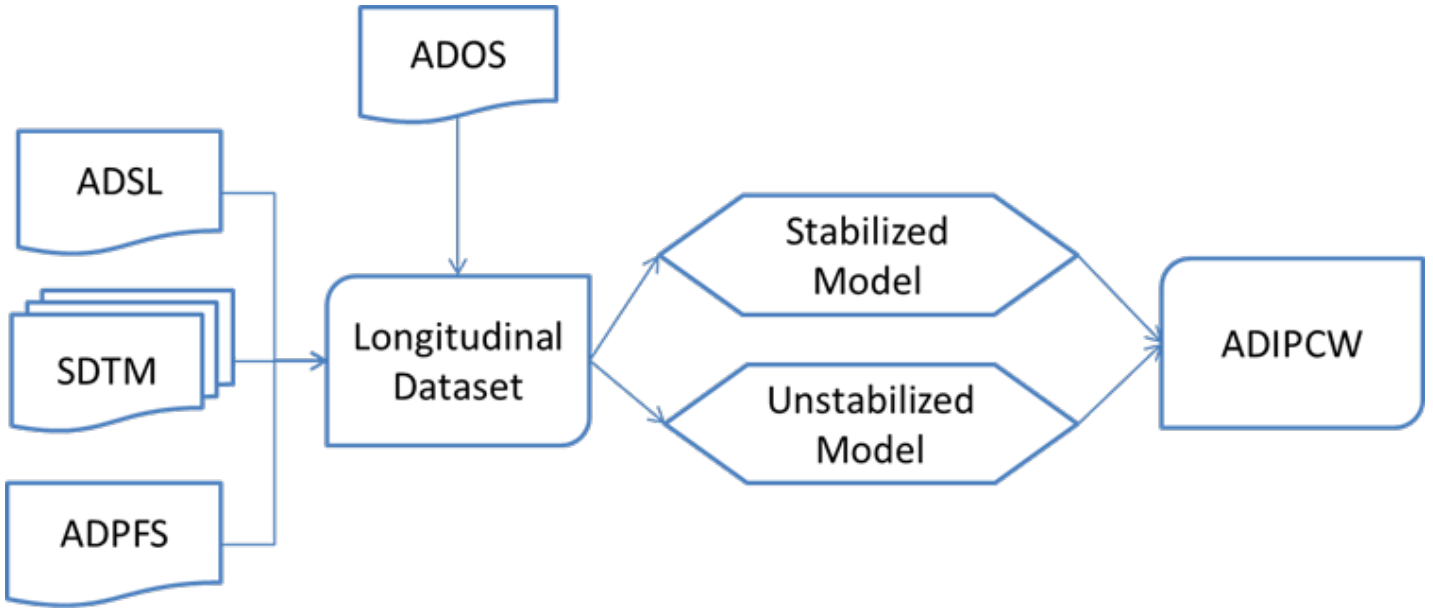
Objective

- This presentation provides a thorough discussion on the following areas:
  - How to accurately populate different types of variables under a longitudinal data structure
  - How to perform logistic models to obtain the IPC weights which feed into time dependent Cox model
  - How to produce the time dependent Cox model

MSD INVENTING FOR LIFE

ADAM Dataset

Here is the flow chart to create the ADAM dataset:



Step 1: Longitudinal Data Structure

- Control arms contain the records from day 1 to the OS day which can be death or censored OS day.

- Treatment contains the single record at subject level because they will not be included in the calculation of weights and are only used for the calculation of hazard ratio.

SAS Code

```
data str;
    set ados;
    do i=1 to aval_os;
       anwk=i;
       output;
    end;
run;
```

Screenshot (an example)

	usubjid	trt01p	aval_os	anwk
1	Subject 001	Control	6	1
2	Subject 001	Control	6	2
3	Subject 001	Control	6	3
4	Subject 001	Control	6	4
5	Subject 001	Control	6	5
6	Subject 001	Control	6	6
7	Subject 002	Control	12	1
8	Subject 002	Control	12	2
9	Subject 002	Control	12	3
10	Subject 002	Control	12	4
11	Subject 002	Control	12	5
12	Subject 002	Control	12	6
13	Subject 002	Control	12	7
14	Subject 002	Control	12	8
15	Subject 002	Control	12	9
16	Subject 002	Control	12	10
17	Subject 002	Control	12	11
18	Subject 002	Control	12	12

Step 2: Rules for Collecting Variables

- The variable indicating death (1 or 0) is only populated at the last record which is the day of death or censored day and the rest of records are 0.
- Baseline variables and demographic variables collected from ADSL will retain the same values for the same subject from day 1 to the last day.
- Time dependent variables such as tumor size will retain the same values until the next values are available (LOCF).
- Special variables are treated in different ways: disease progression censor is populated after PFS censor day , otherwise, it is equal to 1.
- The switchover is flagged at day level (cens).

MSD
INVENTING FOR LIFE

Screenshot for Baseline and Time Dependent Variables

usubjid	aval_os	switchwk	trt01p	anwk	age	tmr_base	tmr	pfscnsr	death	cens	start	stop
Subject 001	6	3	Control	1	21	50	55	1	0	0	0	1
Subject 001	6	3	Control	2	21	50	60	0	0	0	1	2
Subject 001	6	3	Control	3	21	50	65	0	0	1	2	3
Subject 001	6	3	Control	4	21	50	70	0	0	1	3	4
Subject 001	6	3	Control	5	21	50	75	0	0	1	4	5
Subject 001	6	3	Control	6	21	50	80	0	1	1	5	6
Subject 002	12	6	Control	1	22	150	155	1	0	0	0	1
Subject 002	12	6	Control	2	22	150	160	1	0	0	1	2
Subject 002	12	6	Control	3	22	150	165	1	0	0	2	3
Subject 002	12	6	Control	4	22	150	170	1	0	0	3	4
Subject 002	12	6	Control	5	22	150	175	0	0	0	4	5
Subject 002	12	6	Control	6	22	150	180	0	0	1	5	6
Subject 002	12	6	Control	7	22	150	185	0	0	1	6	7
Subject 002	12	6	Control	8	22	150	190	0	0	1	7	8
Subject 002	12	6	Control	9	22	150	195	0	0	1	8	9
Subject 002	12	6	Control	10	22	150	200	0	0	1	9	10
Subject 002	12	6	Control	11	22	150	205	0	0	1	10	11
Subject 002	12	6	Control	12	22	150	210	0	0	1	11	12
Subject 003	18	9	Control	1	33	250	255	1	0	0	0	1
Subject 003	18	9	Control	2	33	250	260	1	0	0	1	2
Subject 003	18	9	Control	3	33	250	265	1	0	0	2	3
Subject 003	18	9	Control	4	33	250	270	1	0	0	3	4
Subject 003	18	9	Control	5	33	250	275	1	0	0	4	5
Subject 003	18	9	Control	6	33	250	280	1	0	0	5	6
Subject 003	18	9	Control	7	33	250	285	1	0	0	6	7
Subject 003	18	9	Control	8	33	250	290	0	0	0	7	8
Subject 003	18	9	Control	9	33	250	295	0	0	1	8	9
Subject 003	18	9	Control	10	33	250	300	0	0	1	9	10
Subject 003	18	9	Control	11	33	250	305	0	0	1	10	11
Subject 003	18	9	Control	12	33	250	310	0	0	1	11	12
Subject 003	18	9	Control	13	33	250	315	0	0	1	12	13
Subject 003	18	9	Control	14	33	250	320	0	0	1	13	14

MSD
INVENTING FOR LIFE

Step 3: Calculating Weights - Creating Dependent Variable

- Match switchover dummy variable against the independent variables at the prior day, i.e. y(i) ~ x(i-1).
- If the analysis time (anwk) is later than switch time (switchwk), then censt is set to be missing. Therefore, those records are not weighted.

usubjid	aval_os	switchwk	trt01p	anwk	age	tmr_base	tmr	death	pfscnsr	cens	start	stop	censt
Subject 001	6	3	Control	1	21	50	55	0	1	0	0	1	0
Subject 001	6	3	Control	2	21	50	60	0	0	0	1	2	1
Subject 001	6	3	Control	3	21	50	65	0	0	1	2	3	.
Subject 001	6	3	Control	4	21	50	70	0	0	1	3	4	.
Subject 001	6	3	Control	5	21	50	75	0	0	1	4	5	.
Subject 001	6	3	Control	6	21	50	80	1	0	1	5	6	.
Subject 002	12	6	Control	1	22	150	155	0	1	0	0	1	0
Subject 002	12	6	Control	2	22	150	160	0	1	0	1	2	0
Subject 002	12	6	Control	3	22	150	165	0	1	0	2	3	0
Subject 002	12	6	Control	4	22	150	170	0	1	0	3	4	0
Subject 002	12	6	Control	5	22	150	175	0	0	0	4	5	1
Subject 002	12	6	Control	6	22	150	180	0	0	1	5	6	.
Subject 002	12	6	Control	7	22	150	185	0	0	1	6	7	.
Subject 002	12	6	Control	8	22	150	190	0	0	1	7	8	.
Subject 002	12	6	Control	9	22	150	195	0	0	1	8	9	.
Subject 002	12	6	Control	10	22	150	200	0	0	1	9	10	.
Subject 002	12	6	Control	11	22	150	205	0	0	1	10	11	.
Subject 002	12	6	Control	12	22	150	210	0	0	1	11	12	.

Step 4: Modelling for the Probability of Switchover

- Stabilized model is constructed with switchover as the dependent variable, and IPCW analysis day and other baseline and demographic variables served as covariates.

```
proc logistic data= inputdata;
    model censt (event='0')= age tmr_base … anwk;
    output out=logit_num pred=pl_num;
run;
```

- Unstabilized model is built using the variables from stabilized model but also includes time dependent covariates corresponding to the baseline variables.

```
proc logistic data= inputdata;
    model censt (event='0')=age tmr_base … anwk pfscnsr tmr …;
    output out=logit_denom pred=pl_denom;
run;
```

MSD INVENTING FOR LIFE

Product Limit Weights and Adjustment

- The final weight is equal to the product limit of the ratio of the probabilities obtained from stabilized model and unstabilized model.

$$w_{ij} = \frac{\prod_{k=0}^{j} P(C(k)_i = 0 | C(k-1)_i = 0, X_i)}{\prod_{k=0}^{j} P(C(k)_i = 0 | C(k-1)_i = 0, X_i, Z(k)_i)}$$

- The weights are applied for the records before the switchover day, and after crossover, the weights are set up to be 0.

Screenshot for Final ADAM Datasets

usubjid	aval_os	switchwk	trt01p	anwk	age	tmr_base	tmr	death	pfscnsr	cens	start	stop	censt	wgt
Subject 001	6	3	Control	1	21	50	55	0	1	0	0	1	0	0.99988
Subject 001	6	3	Control	2	21	50	60	0	0	0	1	2	1	0.99951
Subject 001	6	3	Control	3	21	50	65	0	0	1	2	3	.	0
Subject 001	6	3	Control	4	21	50	70	0	0	1	3	4	.	0
Subject 001	6	3	Control	5	21	50	75	0	0	1	4	5	.	0
Subject 001	6	3	Control	6	21	50	80	1	0	1	5	6	.	0
Subject 002	12	6	Control	1	22	150	155	0	1	0	0	1	0	0.99988
Subject 002	12	6	Control	2	22	150	160	0	1	0	1	2	0	0.99951
Subject 002	12	6	Control	3	22	150	165	0	1	0	2	3	0	0.99889
Subject 002	12	6	Control	4	22	150	170	0	1	0	3	4	0	0.99804
Subject 002	12	6	Control	5	22	150	175	0	0	0	4	5	1	0.99693
Subject 002	12	6	Control	6	22	150	180	0	0	1	5	6	.	0
Subject 002	12	6	Control	7	22	150	185	0	0	1	6	7	.	0
Subject 002	12	6	Control	8	22	150	190	0	0	1	7	8	.	0
Subject 002	12	6	Control	9	22	150	195	0	0	1	8	9	.	0
Subject 002	12	6	Control	10	22	150	200	0	0	1	9	10	.	0
Subject 002	12	6	Control	11	22	150	205	0	0	1	10	11	.	0
Subject 002	12	6	Control	12	22	150	210	0	0	1	11	12	.	0
Subject 003	18	9	Control	1	33	250	255	0	1	0	0	1	0	0.99988
Subject 003	18	9	Control	2	33	250	260	0	1	0	1	2	0	0.99951
Subject 003	18	9	Control	3	33	250	265	0	1	0	2	3	0	0.99889
Subject 003	18	9	Control	4	33	250	270	0	1	0	3	4	0	0.99804
Subject 003	18	9	Control	5	33	250	275	0	1	0	4	5	0	0.99693
Subject 003	18	9	Control	6	33	250	280	0	1	0	5	6	0	0.99559
Subject 003	18	9	Control	7	33	250	285	0	1	0	6	7	0	0.994
Subject 003	18	9	Control	8	33	250	290	0	0	0	7	8	1	0.99217

Time Dependent Cox Model

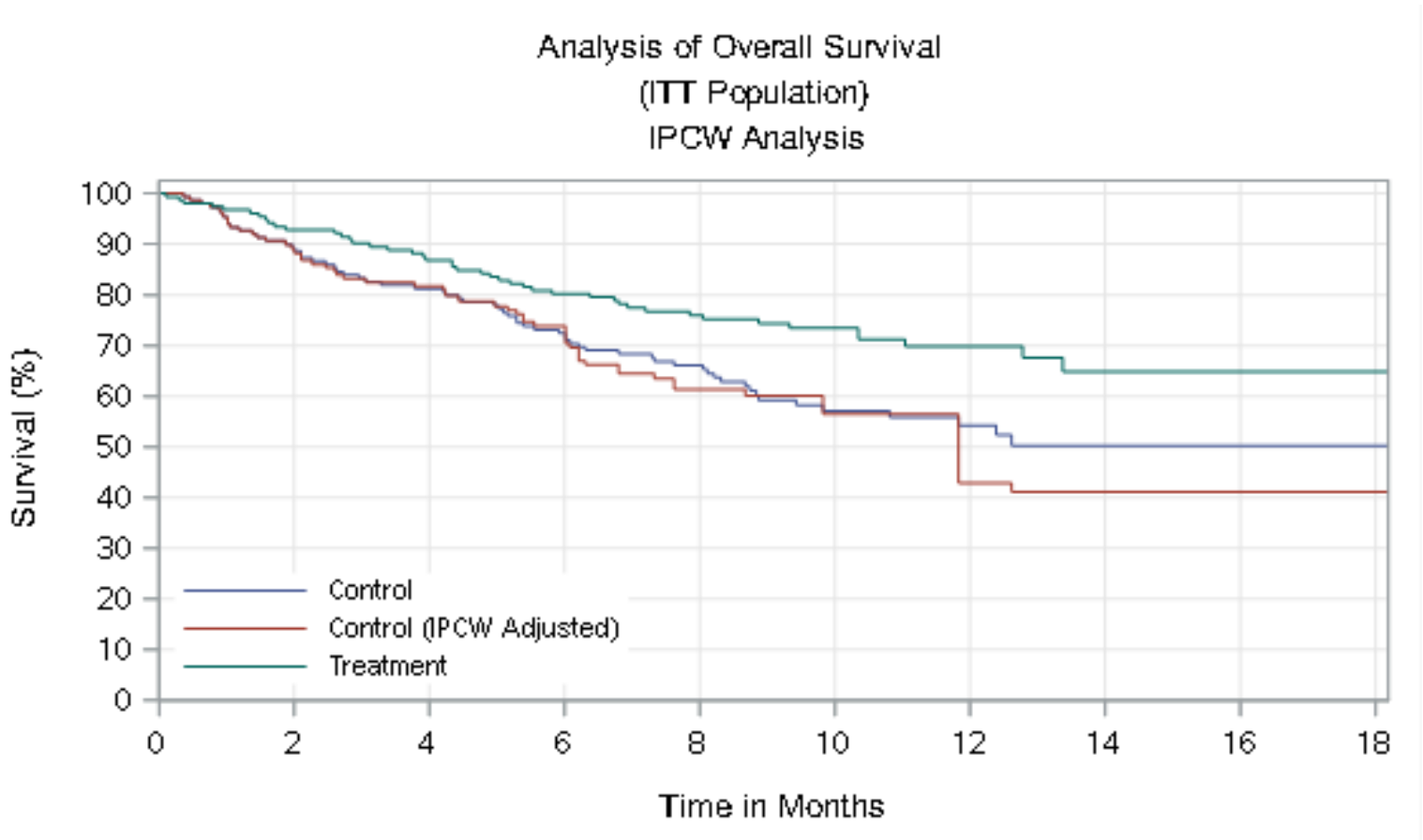
- The hazard ratio is obtained from a time dependent Cox model with weights.

```
ods output ParameterEstimates=ph_orig0;

proc phreg data=adipcw covs(aggregate);
     class trt01pn (ref='Control');
     model (start, stop)*death(0)=trt01p /ties=efron rl;
     weight wgt;
     id usubjid;
run;
```

MSD INVENTING FOR LIFE

Kaplan-Meier Plots



MSD
INVENTING FOR LIFE

Conclusions

- This presentation illustrates the steps to derive an ADAM dataset using the IPCW method:
    - Setting up a longitudinal data structure
    - Deriving baseline variables and time dependent variables from ADSL and SDTM datasets
    - Constructing stabilized and un-stabilized models
    - Calculating weights based on the probabilities from the above models
- It also discusses how to produce the time dependent Cox model.

MSD INVENTING FOR LIFE

Future Research

- Last observation carried forward (LOCF) is used in this presentation. Future research can employ other methods.

- This study uses day interval, and future studies may introduce other time unit such as week or month. However, this may introduce some uncertainty when subject's data do not cover the unit interval.

MSD
INVENTING FOR LIFE

THANK YOU!

Your comments and questions are valued and encouraged. Please contact:

Yirong Cao
MSD
MRL London Hub
Two Pancras Square
London, N1C 4AG
United Kingdom
yirong.cao@merck.com

Shahrul Mt-Isa
MSD
MRL London Hub
Two Pancras Square
London, N1C 4AG
United Kingdom
shahrul.mt-isa@merck.com

MSD