Biomolecules and Biomedicine

ISSN: 2831-0896 (Print) | ISSN: 2831-090X (Online)
Journal Impact Factor® (2024): 2.2

CiteScore® (2024): 5.2
www.biomolbiomed.com | blog.bjbms.org

SUPPLEMENTAL DATA

Bariatric metabolic surgery and cancer risk: Target trial emulation using iterative time distribution matching

Jazeel Abdulmajeed[1,2], Zumin Shi[3], Manar E. Abdel-Rahman[4], Fakhar Shahid[5], Mohammed F. Alam[4], Mashael Al-Shafai[6,7], Muhammad E. H. Chowdhury[8], Abdullah Shaito[7], Adedayo A. Onitilo[9,10], Suhail A. Doi[1*]

[1]Department of Population Medicine, College of Medicine, QU Health, Qatar University, Doha, Qatar;
[2]Business and Health Intelligence Department, Primary Health Care Corporation, Doha, Qatar;
[3]Department of Nutrition Sciences, College of Health Sciences, QU Health, Qatar University, Doha, Qatar;
[4]Department of Public Health, College of Health Sciences, QU Health, Qatar University, Doha, Qatar;
[5]Department of Bariatric and Metabolic Surgery, Hamad Medical Corporation, Doha, Qatar;
[6]Department of Biomedical Sciences, College of Health Sciences, QU Health, Qatar University, Doha, Qatar;
[7]Biomedical Research Center (BRC), QU Health, Qatar University, Doha, Qatar;
[8]Department of Electrical Engineering, Qatar University, Doha, Qatar;
[9]Marshfield Clinic Health System, Inc. Marshfield Medical Center, Wisconsin, USA;

[10]Wisconsin NCI Community Oncology Research Program (WiNCORP), University of Wisconsin, Wisconsin, USA.

*Correspondence to Suhail A. Doi: sdoi@qu.edu.qa

Full article is available at the following link: Bariatric metabolic surgery and cancer risk: Target trial emulation using iterative time distribution matching

DAG code

```
dag {
bb="0,0,1,1"
"enrolment factors" [pos="0.130,0.549"]
"inclusion - BMS/non-BMS" [pos="0.383,0.631"]
"prognostic factors" [adjusted,pos="0.817,0.567"]
BMS [exposure,pos="0.169,0.400"]
C [pos="0.509,0.507"]
Ca [outcome,pos="0.644,0.393"]
mediators [pos="0.406,0.246"]
"enrolment factors" -> "inclusion - BMS/non-BMS"
"inclusion - BMS/non-BMS" -> BMS
"inclusion - BMS/non-BMS" -> C
"prognostic factors" -> "inclusion - BMS/non-BMS"
"prognostic factors" -> Ca
BMS -> Ca
BMS -> mediators
mediators -> Ca
}
```

Stata Code for ITDM analysis

Metadata: Explanation of variables

bariatric_sx_hx	- *Indicator for bariatric metabolic surgery history (1 = Yes, 0 = No)*
age_at_barsx	- *Age at the time of bariatric metabolic surgery (for BMS recipients)*
dox	- *Date of exit from study*
doe_30y	- *Date of origin at age 30*
ca_case	- *Indicator for cancer diagnosis (1 = Yes, 0 = No)*
dob	- *Date of birth*
gender	- *Gender (1= Female, 2= Male)*
nat_grp	- *Nationality group indicator (1 = Qatari, 0 = Non-Qatari)*
bmi40_bef40	- *Indicator for BMI >40 before age 40 (1 = Yes, 0 = No)*

dm_ind - *Indicator for diabetes diagnosis at origin (1 = Yes, 0 = No)*

smok_stat - *Smoking status indicator (0 = Non-smoker 1 = Current smoker 99 = Unknown)*

For the following the number indicates which iteration

agebar1 to agebar10

Assigned age at time-zero

leadtime1 to leadtime10

Time difference between age 30y and time-zero

dbs1 to dbs10

Date of Time-zero

For the following the number in the variable indicates a row maximum from that number of rows

dbs_m1 to dbs_m10

time-zero at chosen number of iterations

agebar_m1 to agebar_m10

age at time-zero at chosen number of iterations

fu1 to fu10

Follow-up time from time-zero to exit (in years) at chosen number of iterations

Stata code

```
clear all

****ITDM iterations to generate random T0 for BMS non-recipients

****ITDM iteration 1
**Generate random age at bariatric metabolic surgery for all participants based on age at bariatric metabolic surgery of BMS recipients
local numseed = round(runiform(1,100))
sort bariatric_sx_hx
mata: mata clear
```

```
putmata age_at_barsx
putmata bariatric_sx_hx
mata:
  rseed(`numseed’)
  yrows = rows(age_at_barsx)
  grpn = _mm_panels(bariatric_sx_hx)
  grpn
  index = mm_sample((0\yrows),grpn)
  agebar1 = age_at_barsx[index[,1],1]
end
getmata agebar1
**Retain original age at bariatric metabolic surgery for BMS recipients
replace agebar1 = age_at_barsx if bariatric_sx_hx==1
**Generate lead time (immortal time)
gen leadtime1 = (agebar1-30)*365.25
**Generate assigned date of bariatric metabolic surgery (T0) for all participants
gen dbs1 = doe_30y+leadtime1
**Generate follow up time from assigned T0 to date of exit for all participants
gen fu1 = (dox-dbs1)/365.25
**Censor invalid entries with negative follow-up times
replace dbs1=. if fu1<0
**Select the maximum value from all available T0 (one in 1st iteration, two in 2nd
iteration etc..)
egen dbs_m1 = rowmax(dbs1)
**generate age at selected T0
gen agebar_m1 = (dbs_m1-dob)/365.25


****ITDM iteration 2
local numseed = round(runiform(1,100))
sort bariatric_sx_hx
mata: mata clear
putmata age_at_barsx
putmata bariatric_sx_hx
```

```
mata:
  rseed(`numseed’)
  yrows = rows(age_at_barsx)
  grpn = _mm_panels(bariatric_sx_hx)
  grpn
  index = mm_sample((0\yrows),grpn)
  agebar2 = age_at_barsx[index[,1],1]
end
getmata agebar2
replace agebar2 = age_at_barsx if bariatric_sx_hx==1
gen leadtime2 = (agebar2-30)*365.25
gen dbs2 = doe_30y+leadtime2
gen fu2 = (dox-dbs2)/365.25
replace dbs2=. if fu2<0
egen dbs_m2 = rowmax(dbs1 dbs2 )
gen agebar_m2 = (dbs_m2-dob)/365.25


****ITDM iteration 3
local numseed = round(runiform(1,100))
sort bariatric_sx_hx
mata: mata clear
putmata age_at_barsx
putmata bariatric_sx_hx
mata:
  rseed(`numseed’)
  yrows = rows(age_at_barsx)
  grpn = _mm_panels(bariatric_sx_hx)
  grpn
  index = mm_sample((0\yrows),grpn)
  agebar3 = age_at_barsx[index[,1],1]
end
getmata agebar3
replace agebar3 = age_at_barsx if bariatric_sx_hx==1
```

```
gen leadtime3 = (agebar3-30)*365.25
gen dbs3 = doe_30y+leadtime3
gen fu3 = (dox-dbs3)/365.25
replace dbs3=. if fu3<0
egen dbs_m3 = rowmax(dbs1 dbs2 dbs3 )
gen agebar_m3 = (dbs_m3-dob)/365.25


****ITDM iteration 4
local numseed = round(runiform(1,100))
sort bariatric_sx_hx
mata: mata clear
putmata age_at_barsx
putmata bariatric_sx_hx
mata:
  rseed(`numseed’)
  yrows = rows(age_at_barsx)
  grpn = _mm_panels(bariatric_sx_hx)
  grpn
  index = mm_sample((0\yrows),grpn)
  agebar4 = age_at_barsx[index[,1],1]
end
getmata agebar4
replace agebar4 = age_at_barsx if bariatric_sx_hx==1
gen leadtime4 = (agebar4-30)*365.25
gen dbs4 = doe_30y+leadtime4
gen fu4 = (dox-dbs4)/365.25
replace dbs4=. if fu4<0
egen dbs_m4 = rowmax(dbs1 dbs2 dbs3 dbs4 )
gen agebar_m4 = (dbs_m4-dob)/365.25


****ITDM iteration 5
local numseed = round(runiform(1,100))
```

```
sort bariatric_sx_hx
mata: mata clear
putmata age_at_barsx
putmata bariatric_sx_hx
mata:
  rseed(`numseed’)
  yrows = rows(age_at_barsx)
  grpn = _mm_panels(bariatric_sx_hx)
  grpn
  index = mm_sample((0\yrows),grpn)
  agebar5 = age_at_barsx[index[,1],1]
end
getmata agebar5
replace agebar5 = age_at_barsx if bariatric_sx_hx==1
gen leadtime5 = (agebar5-30)*365.25
gen dbs5 = doe_30y+leadtime5
gen fu5 = (dox-dbs5)/365.25
replace dbs5=. if fu5<0
egen dbs_m5 = rowmax(dbs1 dbs2 dbs3 dbs4 dbs5 )
gen agebar_m5 = (dbs_m5-dob)/365.25


****ITDM iteration 6
local numseed = round(runiform(1,100))
sort bariatric_sx_hx
mata: mata clear
putmata age_at_barsx
putmata bariatric_sx_hx
mata:
  rseed(`numseed’)
  yrows = rows(age_at_barsx)
  grpn = _mm_panels(bariatric_sx_hx)
  grpn
  index = mm_sample((0\yrows),grpn)
```

```
  agebar6 = age_at_barsx[index[,1],1]
end
getmata agebar6
replace agebar6 = age_at_barsx if bariatric_sx_hx==1
gen leadtime6 = (agebar6-30)*365.25
gen dbs6 = doe_30y+leadtime6
gen fu6 = (dox-dbs6)/365.25
replace dbs6=. if fu6<0
egen dbs_m6 = rowmax(dbs1 dbs2 dbs3 dbs4 dbs5 dbs6)
gen agebar_m6 = (dbs_m6-dob)/365.25


****ITDM iteration 7
local numseed = round(runiform(1,100))
sort bariatric_sx_hx
mata: mata clear
putmata age_at_barsx
putmata bariatric_sx_hx
mata:
  rseed(`numseed')
  yrows = rows(age_at_barsx)
  grpn = _mm_panels(bariatric_sx_hx)
  grpn
  index = mm_sample((0\yrows),grpn)
  agebar7 = age_at_barsx[index[,1],1]
end
getmata agebar7
replace agebar7 = age_at_barsx if bariatric_sx_hx==1
gen leadtime7 = (agebar7-30)*365.25
gen dbs7 = doe_30y+leadtime7
gen fu7 = (dox-dbs7)/365.25
replace dbs7=. if fu7<0
egen dbs_m7 = rowmax(dbs1 dbs2 dbs3 dbs4 dbs5 dbs6 dbs7)
gen agebar_m7 = (dbs_m7-dob)/365.25
```

```
****ITDM iteration 8
local numseed = round(runiform(1,100))
sort bariatric_sx_hx
mata: mata clear
putmata age_at_barsx
putmata bariatric_sx_hx
mata:
  rseed(`numseed’)
  yrows = rows(age_at_barsx)
  grpn = _mm_panels(bariatric_sx_hx)
  grpn
  index = mm_sample((0\yrows),grpn)
  agebar8 = age_at_barsx[index[,1],1]
end
getmata agebar8
replace agebar8 = age_at_barsx if bariatric_sx_hx==1
gen leadtime8 = (agebar8-30)*365.25
gen dbs8 = doe_30y+leadtime8
gen fu8 = (dox-dbs8)/365.25
replace dbs8=. if fu8<0
egen dbs_m8 = rowmax(dbs1 dbs2 dbs3 dbs4 dbs5 dbs6 dbs7 dbs8)
gen agebar_m8 = (dbs_m8-dob)/365.25


****ITDM iteration 9
local numseed = round(runiform(1,100))
sort bariatric_sx_hx
mata: mata clear
putmata age_at_barsx
putmata bariatric_sx_hx
mata:
  rseed(`numseed’)
```

```
  yrows = rows(age_at_barsx)
  grpn = _mm_panels(bariatric_sx_hx)
  grpn
  index = mm_sample((0\yrows),grpn)
  agebar9 = age_at_barsx[index[,1],1]
end
getmata agebar9
replace agebar9 = age_at_barsx if bariatric_sx_hx==1
gen leadtime9 = (agebar9-30)*365.25
gen dbs9 = doe_30y+leadtime9
gen fu9 = (dox-dbs9)/365.25
replace dbs9=. if fu9<0
egen dbs_m9 = rowmax(dbs1 dbs2 dbs3 dbs4 dbs5 dbs6 dbs7 dbs8 dbs9)
gen agebar_m9 = (dbs_m9-dob)/365.25


****ITDM iteration 10
local numseed = round(runiform(1,100))
sort bariatric_sx_hx
mata: mata clear
putmata age_at_barsx
putmata bariatric_sx_hx
mata:
  rseed(`numseed')
  yrows = rows(age_at_barsx)
  grpn = _mm_panels(bariatric_sx_hx)
  grpn
  index = mm_sample((0\yrows),grpn)
  agebar10 = age_at_barsx[index[,1],1]
end
getmata agebar10
replace agebar10 = age_at_barsx if bariatric_sx_hx==1
gen leadtime10 = (agebar10-30)*365.25
gen dbs10 = doe_30y+leadtime10
```

```
gen fu10 = (dox-dbs10)/365.25
replace dbs10=. if fu10<0
egen dbs_m10 = rowmax(dbs1 dbs2 dbs3 dbs4 dbs5 dbs6 dbs7 dbs8 dbs9 dbs10)
gen agebar_m10 = (dbs_m10-dob)/365.25

** Checking distributions (eg. below: T0 from 6 iterations)
tabstat agebar_m6, by( bariatric_sx_hx ) stat(median iqr n)
distplot agebar_m6,  over( bariatric_sx_hx )

***ITDM transition plot
gen iteration = .
gen ksd = .
forvalues k = 1(1)10 {
	ksmirnov agebar_m`k', by(bariatric_sx_hx)
	replace ksd = r(D) in `k'
	replace iteration = `k' in `k'
}

twoway (connect ksd iteration, mlabel(ksd) mlabformat(%4.3f) mlabpos(12)) ,
title("ITDM Transition Plot") ytitle("Kolmogorov–Smirnov D Value")
xtitle("Iterations")



**  ITDM with 6 iterations provided the best alignment of the immortal time
distribution between groups

***** Survival analysis setup and Cox regression analysis



** ITDM analysis (stset using T0 assigned from 6 iterations)
stset dox, fail(ca_case==1) origin(time doe_30y) enter(time dbs_m6) scale(365.25)
id(id)
stcox i.bariatric_sx_hx  i.nat_grp i.bmi40_bef40 i.dm_ind  agebar_m6 , strata(gender
smok_stat)
```

```
** PTDM analysis (stset using T0 assigned from 1 iteration)
stset dox, fail(ca_case==1) origin(time doe_30y) enter(time dbs_m1) scale(365.25) id(id)
stcox i.bariatric_sx_hx  i.nat_grp i.bmi40_bef40 i.dm_ind  agebar_m1, strata(gender smok_stat)

** naive analysis (stset using date of entry at 30y)
stset dox, fail(ca_case==1) origin(time doe_30y) scale(365.25) id(id)
stcox i.bariatric_sx_hx i.nat_grp i.bmi40_bef40 i.dm_ind, strata(gender smok_stat)
```

Table S1. Table outlining the protocol for emulation of a target trial

Component	Target trial	Emulated trial using real-world data
Design	Open-label two-parallel arm superiority randomised trial.	
Aim	Estimate the effect of receiving bariatric surgery after age 30y on overall cancer incidence	Same
Eligibility	No cancer diagnosis or bariatric surgery prior to age 30y with a BMI > 30 at age 30y	No cancer diagnosis or bariatric surgery prior to age 30y with a BMI > 30 noted between age 30y-40y. In addition individuals are required to have a measurement of minimal adjustment set determined using a DAG
Exclusions	Patients with a cancer in the year after bariatric surgery	Same

Treatment strategies	1. Bariatric surgery after age 30y 2. No bariatric surgery after age 30y	Same
Treatment assignment	Eligible patients are randomly assigned to either strategy	Patients are non-randomly assigned to a treatment strategy. Randomisation is emulated via adjustment using a DAG
Treatment implementation	None	Any time after eligibility
Outcome	Incident cancer after bariatric surgery	Same
Type of outcome	Cancer incidence	Same
Start of follow up	For each individual follow-up starts at the time of assignment to a strategy (and all eligibility criteria are met)	The first time when all eligibility criteria are met.
End of follow-up	The occurrence cancer, loss to follow-up or administrative censoring (Dec 2024),	The occurrence of cancer, loss to follow up, administrative censoring

	whichever comes first.	
Causal contrast	Initiating at assignment versus not initiating at assignment	Initiating versus not initiating at comparable times after start of follow-up
Estimands	Ratio of cancer hazard (hazard ratio) between arms	

Abbreviations: BMI: Body mass index; DAG: Directed acyclic graph.

Table S2. Distribution of cancer cases in original cohort vs analysis cohort*

Cancer Type	**BMS recipients**			**BMS non-recipients**			**Total**
	In original cohort	Before ITDM process	After ITDM process (analysis cohort)	In original cohort	Before ITDM process	After ITDM process (analysis cohort)	
Any cancer	41	21 (51%)	21 (51%)	2071	1,716 (83%)	1,562 (75%)	2112
Thyroid	14	7 (50%)	7 (50%)	447	368 (82%)	326 (73%)	461
Breast	5	1 (20%)	1 (20%)	387	352 (91%)	331 (86%)	392
Hematologic	7	2 (29%)	2 (29%)	380	273 (72%)	237 (62%)	387

Reproductive	5	5 (100%)	5 (100%)	187	159 (85%)	144 (77%)	192
Gastrointestinal	2	1 (50%)	1 (50%)	139	119 (86%)	112 (81%)	141
Unknown Primary	2	2 (100%)	2 (100%)	123	100 (81%)	92 (75%)	125
Urinary	4	2 (50%)	2 (50%)	98	87 (89%)	82 (84%)	102
Bone and Soft Tissue	0	0	0	95	74 (78%)	68 (72%)	95
Central Nervous System	1	1 (100%)	1 (100%)	59	49 (83%)	42 (71%)	60
Skin	0	0	0	56	50 (89%)	47 (84%)	56
Respiratory	0	0	0	23	22 (96%)	22 (96%)	23
Other	1	0 (0%)	0 (0%)	77	63 (82%)	59 (77%)	78

% represents the proportion of cases at each stage relative to the original cohort. *It is important to note that the decrease in cancer cases among BMS recipients occurred prior to the age of 30 and not during the ITDM procedure. Abbreviations: BMS: Bariatric metabolic surgery; ITDM: Iterative time distribution matching.

Table S3. Sensitivity analysis removing two random cancer cases (5 iterations)

Iteration	HR	LCI	UCI
1	0.44	0.28	0.71
2	0.44	0.28	0.71
3	0.47	0.30	0.73
4	0.44	0.28	0.71
5	0.44	0.28	0.70

Hazard ratios (HRs) and 95% confidence intervals (CIs) derived from the Cox proportional hazards model. Abbreviations: HR: Hazard ratio; LCI: Lower confidence interval bound; UCI: Upper confidence interval bound; CI: Confidence interval.

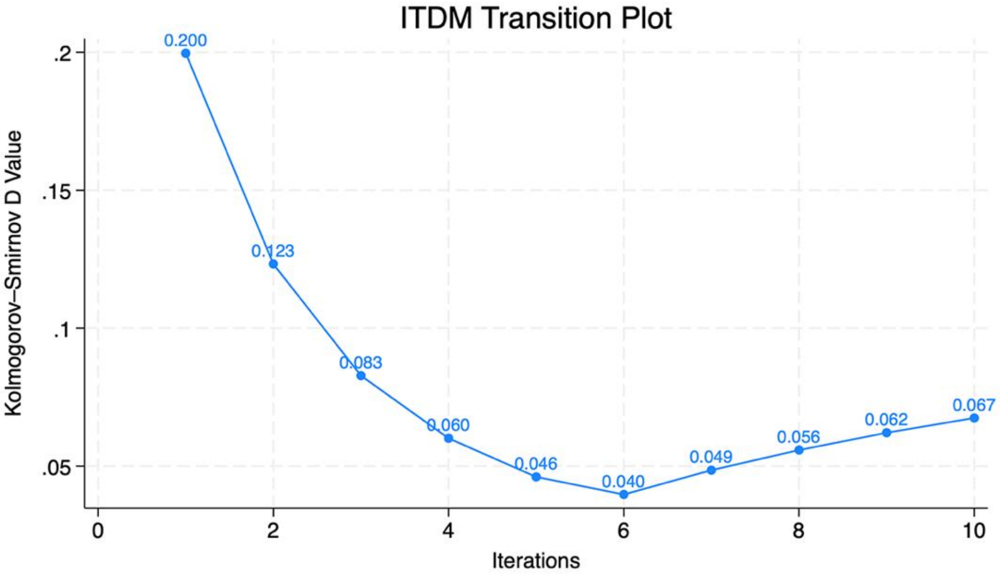

Figure S1. ITDM transition plot showing optimal iteration selection by minimum Kolmogorov–Smirnov *D* value. The curve displays the Kolmogorov–Smirnov D statistic comparing the distribution of assigned immortal time between BMS recipients and non-recipients across 10 ITDM iterations. Lower D indicates better alignment; the minimum occurs at iteration 6 (D ≈ 0.040), which was selected to define the final T0 (mT0) for the non-surgery group. Abbreviations: BMS: Bariatric metabolic surgery; ITDM: Iterative time distribution matching.

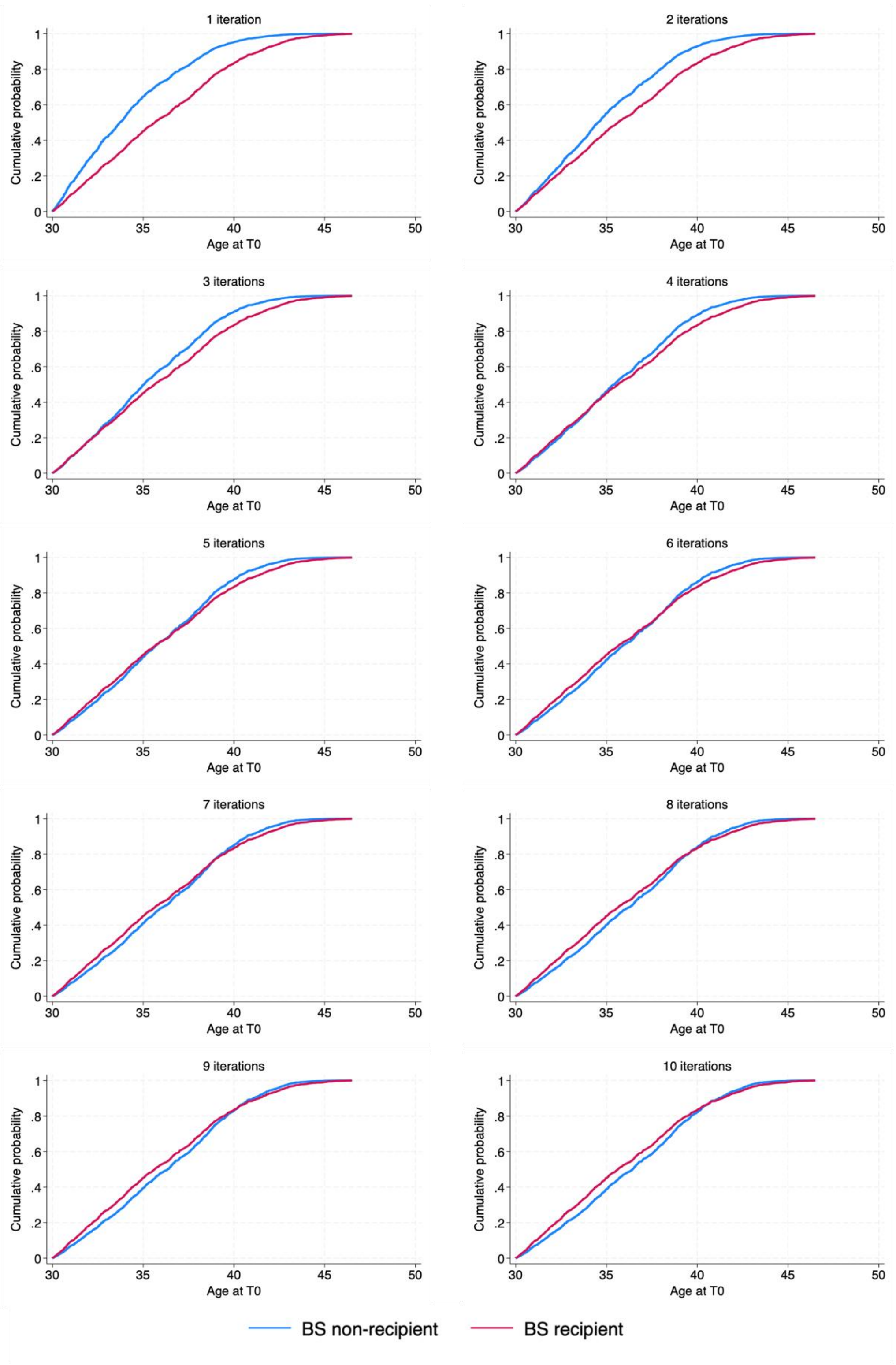

Figure S2. Cumulative distribution plots (Stata distplots) confirming ITDM alignment. Plots confirm that the ITDM method—following the PTDM process—aligns the time distribution before analysis and minimizes drop-outs due to invalid T_0

assignment. Abbreviations: ITDM: Iterative time distribution matching; PTDM: Prescription time distribution matching.

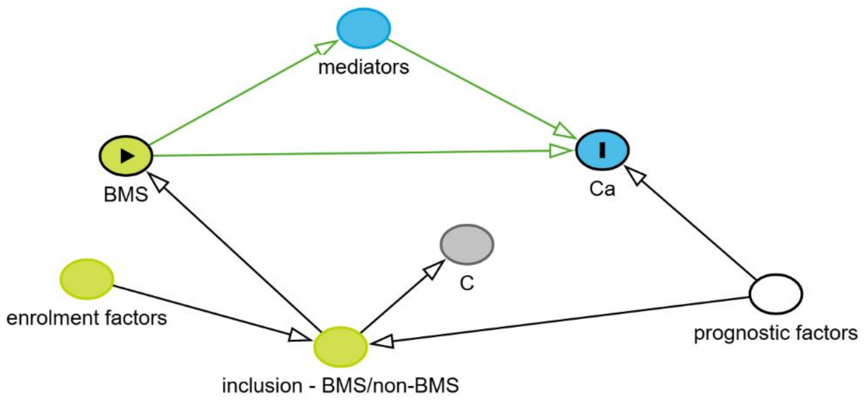

Figure S3. Directed acyclic graph (DAG). The prognostic factors include gender, Qatari nationality, type 2 diabetes mellitus at baseline, class III obesity (BMI ≥ 40 kg/m² or not) at baseline, age at BMS entry, and smoking status. Censoring, denoted as C, is deemed informative if prognostic factors are not adjusted. The mediators in the DAG consist of weight loss trajectory and metabolic remission, which do not require adjustment to ascertain the total causal effect.

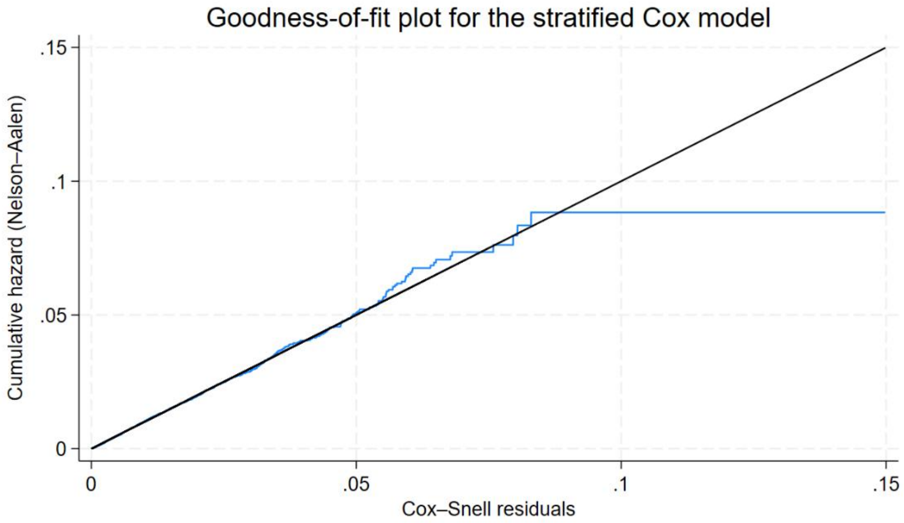

Figure S4. Goodness-of-fit plot for the stratified Cox proportional hazards model. The plot shows observed versus expected cumulative hazard functions across different risk groups. The close alignment of the curves suggests an adequate fit of the model.

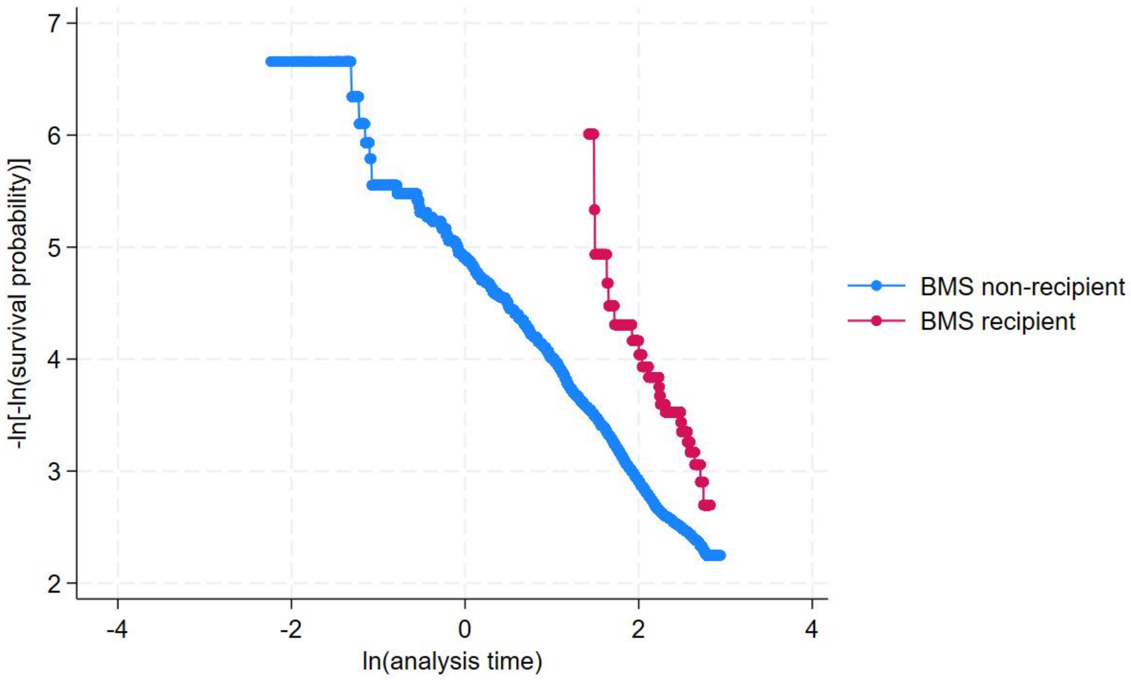

Figure S5. Log–log survival curves (stphplot) for covariates in the stratified Cox proportional hazards model. The survival curves exhibit comparable patterns throughout the majority of the follow-up period, with divergence occurring only at the extremes. This observation supports the approximate adherence to the proportional hazards assumption for the primary variables of interest.