Supplemental Files

- Supplemental Methods
- Supplemental Table 1
- Supplemental Table 2
- Supplemental Table 3
- Supplemental Table 4
- Supplemental Table 5
- Supplemental Table 6
- Supplemental Discussion
- Sample programing code

Supplemental Methods

We here discuss the rationale for the three assumptions—consistency, positivity and conditional exchange ability, which were collectively referred to as identifiability conditions.

Consistency

The consistency assumption holds when the target intervention is well-defined. Our motivated intervention was HbA_{1c}-guided therapy, so we assumed this strategy was well-defined as long as the patients received HbA_{1c} testing regularly.

Positivity

The positivity assumption means that there is a positive probability—not zero or one—at each follow-up month to be assigned to SGLT2-inhibitor treatment based on the past observed covariate history. Our study population included the prevalent users of antidiabetic drugs without advanced kidney disease (i.e., no contraindication). Thus, we assumed that it was implausible that a patient who would never use SGLT2 inhibitor(s) or a patient who would absolutely use the one(s) was not involved in our cohort.

In a post-hoc sensitivity analysis, the different inclusion criteria were applied for the potential concern for the (near-)violation of this assumption.

Conditional exchangeability

Conditional exchangeability means that no unmeasured confounding variable exists. The duration of diabetes for each patient and the preference for SGLT2 inhibitors may be unmeasured confounding variables not available in our data set. However, we hypothesized that the lack of such information could be mitigated by including age, baseline eGFR, the trajectory of HbA_{1c}, and treatment history with antidiabetic drugs as measured covariates.

References

1. Westreich D. Epidemiology by Design: A Causal Approach to the Health Sciences. Oxford, UK. Oxford University Press; 2020.

Supplemental Table 1: Renal risk in different treatment strategy (only persons with eGFR ≥45 mL/min/1.73 m^2)

Intervention	Crude risk (95% CI)	Risk difference (95% CI)	Risk ratio (95% CI)
Natural Course	15.4 (15.0 to 16.0)%	(Reference)	(Reference)
A_{1c}≥7.0%, within 3 mo	11.9 (9.7 to 14.6)%	-3.5 (- 5.5 to -0.89)%	0.77 (0.64 to 0.94)
A_{1c}≥7.0%, within 6 mo	12.4 (10.3 to 14.8)%	-3.1 (-4.9 to -0.68)%	0.80 (0.68 to 0.96)
A_{1c}≥7.0%, within 9 mo	12.8 (10.9 to 15.0)%	-2.7 (- 4.3 to -0.50)%	0.83 (0.72 to 0.97)
A_{1c}≥7.0%, within 12 mo	13.2 (11.4 to 15.2)%	-2.3 (-3.8 to -0.28)%	0.85 (0.75 to 0.98)
A_{1c}≥6.5%, within 3 mo	11.3 (8.8 to 14.3) %	-4.1 (- 6.4 to -1.1)%	0.73 (0.58 to 0.93)
A_{1c}≥6.5%, within 6 mo	11.8 (9.6 to 14.6)%	-3.6 (-5.8 to -0.90)%	0.77 (0.63 to 0.94)
A_{1c}≥6.5%, within 9 mo	12.3 (10.2 to 14.8)%	-3.2 (-5.0 to -0.67)%	0.80 (0.67 to 0.96)
A_{1c}≥6.5%, within 12 mo	12.7 (10.8 to 15.0)%	-2.7 (-4.4 to -0.45)%	0.82 (0.71 to 0.97)
A_{1c}≥7.5%, within 3 mo	12.7 (10.9 to 14.9)%	-2.7 (-4.4 to -0.60)%	0.82 (0.72 to 0.96)
A_{1c}≥7.5%, within 6 mo	13.1 (11.4 to 15.1)%	-2.4 (-3.8 to -0.41)%	0.85 (0.75 to 0.97)
A_{1c}≥7.5%, within 9 mo	13.4 (11.9 to 15.2)%	-2.0 (-3.3 to -0.25)%	0.87 (0.78 to 0.98)
A_{1c}≥7.5%, within 12 mo	13.7 (12.3 to 15.4)%	-1.7 (-2.9 to -0.10)%	0.89 (0.80 to 0.99)
Never Treated	15.5 (15.0 to 16.1)%	0.095 (-0.067 to 0.23)%	1.01 (1.00 to 1.02)

A 95% CI was calculated by bootstrapping, without p-value output. A 95% CI of a rate that does not cross 0 or a ratio that does not cross 1 was regarded as statistically significance.

eGFR: estimated glomerular filtration rate, A_{1c}: hemoglobin A_{1c}, CI: confidence interval

Supplemental Table 2: **Change in eGFR in different treatment strategy**

Intervention	Change in eGFR (95% CI)
Natural Course	(Reference: 62.0 mL/min/1.73 m^2)
A_{1c}≥7.0%, within 3 mo	1.38 (-0.05 to 3.11) mL/min/1.73 m^2
A_{1c}≥7.0%, within 6 mo	1.27 (-0.12 to 2.90) mL/min/1.73 m^2
A_{1c}≥7.0%, within 9 mo	1.16 (-0.18 to 2.68) mL/min/1.73 m^2
A_{1c}≥7.0%, within 12 mo	1.07 (-0.22 to 2.46) mL/min/1.73 m^2
A_{1c}≥6.5%, within 3 mo	1.60 (0.02 to 3.62) mL/min/1.73 m^2
A_{1c}≥6.5%, within 6 mo	1.49 (-0.02 to 3.34) mL/min/1.73 m^2
A_{1c}≥6.5%, within 9 mo	1.38 (-0.11 to 3.15) mL/min/1.73 m^2
A_{1c}≥6.5%, within 12 mo	1.27 (-0.15 to 2.92) mL/min/1.73 m^2
A_{1c}≥7.5%, within 3 mo	1.06 (-0.15 to 2.42) mL/min/1.73 m^2
A_{1c}≥7.5%, within 6 mo	0.97 (-0.21 to 2.22) mL/min/1.73 m^2
A_{1c}≥7.5%, within 9 mo	0.88 (-0.24 to 2.04) mL/min/1.73 m^2
A_{1c}≥7.5%, within 12 mo	0.80 (-0.31 to 1.91) mL/min/1.73 m^2
Never Treated	-0.05 (-0.20 to 0.09) mL/min/1.73 m^2

A 95% CI was calculated by bootstrapping, without p-value output. A 95% CI that does not cross 0 was regarded as statistically significance.

eGFR: estimated glomerular filtration rate, A_{1c}: hemoglobin A_{1c}, CI: confidence interval

Supplemental Table 3: **Renal risk in different treatment strategy for persons with eGFR ≥20 mL/min/1.73 m^2**

Intervention	Crude risk (95% CI)	Risk difference (95% CI)	Risk ratio (95% CI)
Natural Course	21.9 (21.3 to 22.6)%	(Reference)	(Reference)
A_{1c}≥7.0%, within 3 mo	18.6 (16.0 to 20.8)%	- 3.4 (- 5.9 to -1.1)%	0.84 (0.73 to 0.95)
A_{1c}≥7.0%, within 6 mo	19.0 (16.8 to 21.0)%	-2.9 (-5.1 to -0.91)%	0.87 (0.76 to 0.96)
A_{1c}≥7.0%, within 9 mo	19.5 (17.5 to 21.3)%	-2.4 (-4.3 to -0.65)%	0.89 (0.80 to 0.97)
A_{1c}≥7.0%, within 12 mo	19.9 (18.1 to 21.6)%	-2.0 (-3.7 to -0.49)%	0.91 (0.83 to 0.98)
A_{1c}≥6.5%, within 3 mo	17.8 (14.9 to 20.6)%	-4.1 (-7.1 to -1.4)%	0.81 (0.68 to 0.93)
A_{1c}≥6.5%, within 6 mo	18.4 (15.9 to 20.7)%	-3.5 (-6.1 to -1.2)%	0.84 (0.72 to 0.95)
A_{1c}≥6.5%, within 9 mo	19.0 (17.3 to 21.3)%	-3.0 (-5.3 to -0.89)%	0.87 (0.76 to 0.86)
A_{1c}≥6.5%, within 12 mo	19.4 (17.3 to 21.1)%	-2.5 (-4.5 to -0.66)%	0.89 (0.79 to 0.97)
A_{1c}≥7.5%, within 3 mo	19.4 (17.3 to 21.1)%	-2.5 (-4.5 to -0.81)%	0.88 (0.79 to 0.96)
A_{1c}≥7.5%, within 6 mo	19.8 (18.0 to 21.4)%	-2.1 (-3.8 to -0.57)%	0.90 (0.85 to 0.98)
A_{1c}≥7.5%, within 9 mo	20.1 (18.6 to 21.6)%	-1.8 (-3.2 to -0.45)%	0.92 (0.85 to 0.98)
A_{1c}≥7.5%, within 12 mo	20.5 (19.1 to 21.9)%	-1.5 (-2.8 to -0.23)%	0.93 (0.87 to 0.99)
Never Treated	22.0 (21.4 to 22.7)%	0.08 (-0.03 to 0.19)%	1.00 (1.00 to 1.01)

A 95% CI was calculated by bootstrapping, without p-value output. A 95% CI of a rate that does not cross 0 or a ratio that does not cross 1 was regarded as statistically significance.

eGFR: estimated glomerular filtration rate, A_{1c}: hemoglobin A_{1c}, CI: confidence interval

Supplemental Table 4: Results from model further adjusted by antihypertensive medication and ACEi/ARB use

Intervention	Crude risk (95% CI)	Risk difference (95% CI)	Risk ratio (95% CI)
Natural Course	19.2 (18.7 to 20.1) %	(Reference)	(Reference)
A_{1c}≥7.0%, within 3 mo	16.1 (14.5 to 18.5) %	-3.1 (- 5.1 to -0.99) %	0.84 (0.75 to 0.95)
A_{1c}≥7.0%, within 6 mo	16.6 (15.2 to 18.5) %	-2.7 (-4.4 to -0.84) %	0.86 (0.78 to 0.96)
A_{1c}≥7.0%, within 9 mo	16.9 (15.7 to 18.6) %	-2.3 (- 3.8 to -0.74) %	0.88 (0.81 to 0.96)
A_{1c}≥7.0%, within 12 mo	17.3 (16.2 to 18.9) %	-2.0 (-3.3 to 0.54) %	0.90 (0.84 to 0.97)
A_{1c}≥6.5%, within 3 mo	15.6 (13.6 to 18.2) %	-3.7 (- 6.0 to -1.2) %	0.81 (0.70 to 0.93)
A_{1c}≥6.5%, within 6 mo	16.0 (14.4 to 18.4) %	-3.2 (-5.2 to -1.1) %	0.83 (0.74 to 0.95)
A_{1c}≥6.5%, within 9 mo	16.5 (15.1 to 18.4) %	-2.8 (-4.5 to -0.88) %	0.86 (0.77 to 0.95)
A_{1c}≥6.5%, within 12 mo	16.8 (15.2 to 18.5) %	-2.4 (-3.9 to -0.76 %	0.88 (0.80 to 0.96)
A_{1c}≥7.5%, within 3 mo	16.9 (15.6 to 18.6) %	-2.3 (-3.9 to -0.75) %	0.88 (0.81 to 0.96)
A_{1c}≥7.5%, within 6 mo	17.2 (16.1 to 18.7) %	-2.0 (-3.4 to -0.65) %	0.89 (0.83 to 0.97)
A_{1c}≥7.5%, within 9 mo	17.5 (16.5 to 18.9) %	-1.7 (-2.9 to -0.49) %	0.91 (0.85 to 0.97)
A_{1c}≥7.5%, within 12 mo	17.8 (16.8 to 19.2) %	-1.5 (-2.5 to -0.30) %	0.92 (0.87 to 0.98)
Never Treated	19.3 (18.8 to 20.2) %	0.055 (-0.030 to 0.18) %	1.01 (1.00 to 1.01)

A 95% CI was calculated by bootstrapping, without p-value output. A 95% CI of a rate that does not cross 0 or a ratio that does not cross 1 was regarded as statistically significance.

ACEi: angiotensin converting enzyme inhibitor, ARB: angiotensin II receptor blocker, eGFR: estimated glomerular filtration rate, A_{1c}: hemoglobin A_{1c}, CI: confidence interval

Supplemental Table 5: Dipeptidyl peptidase 4 inhibitor[1]

Intervention	Risk difference (95% confidence interval)
Natural Course	(Reference)
A_{1c}≥7.0%, within 3 mo	1.17 (-0.67 to 3.36)%
A_{1c}≥7.0%, within 6 mo	1.69 (0.14 to 3.53)%
A_{1c}≥7.0%, within 9 mo	2.12 (0.64 to 3.63)%
A_{1c}≥7.0%, within 12 mo	2.43 (1.06 to 3.87)%
A_{1c}≥6.5%, within 3 mo	0.84 (-1.53 to 3.52)%
A_{1c}≥6.5%, within 6 mo	1.50 (-0.41 to 3.75)%
A_{1c}≥6.5%, within 9 mo	2.04 (0.31 to 4.01)%
A_{1c}≥6.5%, within 12 mo	2.45 (0.78 to 4.11)%
A_{1c}≥7.5%, within 3 mo	1.29 (0.04 to 2.90)%
A_{1c}≥7.5%, within 6 mo	1.67 (0.47 to 2.90)%
A_{1c}≥7.5%, within 9 mo	1.95 (0.84 to 3.08)%
A_{1c}≥7.5%, within 12 mo	2.14 (1.11 to 3.17)%

1: Data from 24,245 persons were used for the computation.

Supplemental Table 6: Glucagon-like peptide-1[1]

Intervention	Risk difference (95% confidence interval)
Natural Course	(Reference)
A_{1c}≥7.0%, within 3 mo	-0.11 (-3.01 to 3.86) %
A_{1c}≥7.0%, within 6 mo	0.008 (-2.62 to 3.38) %
A_{1c}≥7.0%, within 9 mo	0.10 (-2.30 to 3.00)%
A_{1c}≥7.0%, within 12 mo	0.18 (-2.07 to 2.92) %
A_{1c}≥6.5%, within 3 mo	-0.23 (-3.63 to 4.57) %
A_{1c}≥6.5%, within 6 mo	-0.08 (-3.18 to 4.07) %
A_{1c}≥6.5%, within 9 mo	0.02 (-2.77 to 3.51) %
A_{1c}≥6.5%, within 12 mo	0.13 (-2.45 to 3.18)%
A_{1c}≥7.5%, within 3 mo	-0.007 (-2.25 to 2.95) %
A_{1c}≥7.5%, within 6 mo	0.84 (-1.98 to 2.56) %
A_{1c}≥7.5%, within 9 mo	0.15 (-1.78 to 2.42) %
A_{1c}≥7.5%, within 12 mo	0.21 (-1.60 to 2.19)%

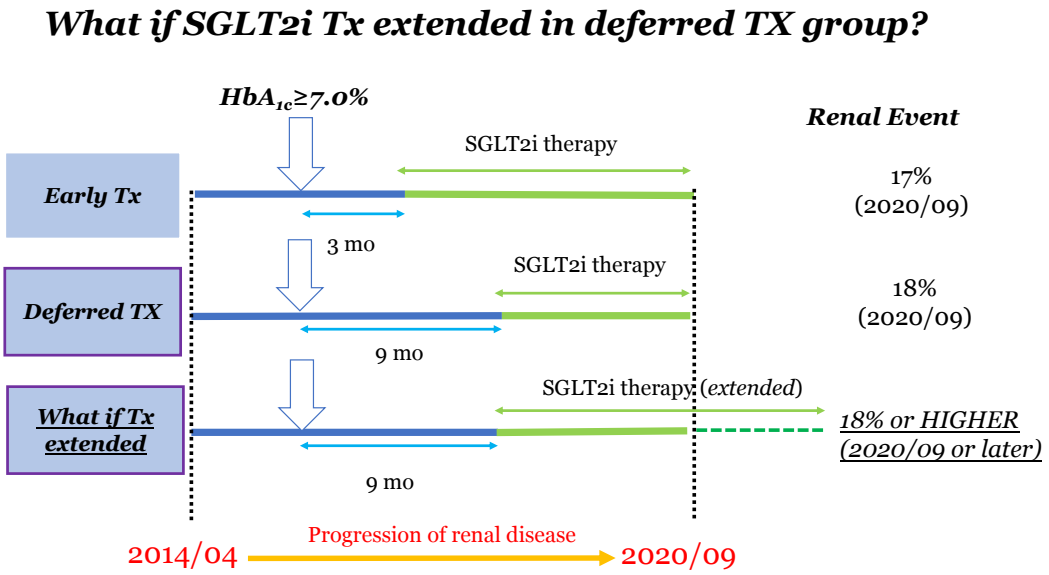
1: Data from 1261 persons were used for the computation.

Supplemental Discussion

Our study favored the early introduction of SGLT2 inhibitors. This result could be influenced the duration of exposure to SGLT2 inhibitors as the end of follow up was fixed in September, 2020 for all persons. In the illustrative example below, the early treatment population initiating SGLT2 inhibitors therapy had 6-month longer duration under drug exposure (the upper vs the middle). Thus, it naturally comes to the mind whether or not the lower incidence of the composite renal event in earlier initiation population was resulted from the different SGLT2 inhibitors exposure time.

To explain this influence, imagine the situation where had the duration of SGLT2 inhibitors be extended in deferred treatment population (the middle and the lower). In the lower hypothetical group, the renal event would have occurred in 18% (as identical to the middle) in September, 2020. Given the irreversible nature of eGFR trajectory (ie, declining over time), the incidence of the renal outcome in the lower group could have been expected as 18%—or even higher—even if the SGLT2 therapy extended.

That is, we assume that the renal benefit in earlier treatment was largely explained by the early introduction of SGLT2 inhibitors, rather than the longer duration of drug exposure within the fixed observation period.



SGLT2i: Sodium–glucose cotransporter 2 inhibitors, Tx: treatment/therapy

Sample programing code

```
library(gfoRmula) #version 0.3.2
library(parallel) #for speed up

##############################################################################################
A: time-varying treatment (eg, SGLT2 inhibitors use)
L1, L2: time-varying covariates (can be increased in number as needed). In this
example, L1 is continuous (eg, HbA1c) and L2 is binary.
L3: baseline covariate (can be increased in number as needed)
Y: binary outcome
t0: time index.
id: unique identifier for each individua
##############################################################################################

id <-'id'
time_points <- max(my_data$t0) + 1
time_name <-'t0'
covnames <- c('L1','L2','A')
outcome_name <-'Y'
outcome_type<- 'survival'
compevent_name <-'D'
covtypes <- c('normal', 'binary', 'binary')
histories <- c(lagged, lagavg)
histvars <- list(c('A','L1', 'L2'), c('A','L1', 'L2'))
covparams <- list(covmodels = c(L1 ~ lag1_A + lag_cumavg1_A + lag1_L1  +
                                  lag_cumavg1_L1 + lag1_L2  + lag_cumavg1_L2 + L3 + t0,
                            L2 ~ lag1_A + lag_cumavg1_A + lag1_L1 + lag_cumavg1_L1 +
                              lag1_L2 + lag_cumavg1_L2 + L3 + t0,
                            A ~ lag1_A + lag_cumavg1_A + lag1_L1 + lag_cumavg1_L1 +
                              lag1_L2 + lag_cumavg1_L2 + L3 + t0))


ymodel <- Y ~ A + L1 + L2 + lag1_A + lag_cumavg1_A + lag1_L1 + lag_cumavg1_L1 +
  lag1_L2 + lag_cumavg1_L2 + L3 + t0

dyn_int <- function(newdf, pool, intvar, intvals, time_name, t){
  threshold <- intvals[[1]]
  m <- intvals[[2]]
  if (t == 0){
    newdf[L1 >= threshold, `:=` (cond_met_ever = 1, cond_tracker = t)]
    newdf[L1 < threshold, cond_met_ever := 0]
  } else {
    newdf[cond_met_ever == 0 & L1 >= threshold, `:=`
         (cond_met_ever = 1, cond_tracker = t)]
  }

  if (t > 0){
    newdf[pool[get(time_name) == (t - 1), get(intvar) == 1], (intvar) := 1]
  }

  newdf[cond_met_ever == 0, (intvar) := 0]

  if (t >= m){ newdf[cond_tracker <= (t - m), (intvar) := 1]
  }
}

histories <- c(lagged, lagavg)
intvars<- list('A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A')
int_descript<- c('Thres_7.0_3', 'Thres_7.0_6', 'Thres_7.0_9', 'Thres_7.0_12',
                 'Thres_6.5_3', 'Thres_6.5_6', 'Thres_6.5_9', 'Thres_6.5_12',
                 'Thres_7.5_3', 'Thres_7.5_6', 'Thres_7.5_9', 'Thres_7.5_12',
                 'Never Treat')
interventions<- list(list(c(dyn_int, 7.0, 3)), list(c(dyn_int, 7.0, 6)),
list(c(dyn_int, 7.0, 9)), list(c(dyn_int, 7.0, 12)),
```

```
                    list(c(dyn_int, 6.5, 3)), list(c(dyn_int, 6.5, 6)), list(c(dyn_int,
6.5, 9)), list(c(dyn_int, 6.5, 12)),
                    list(c(dyn_int, 7.5, 3)), list(c(dyn_int, 7.5, 6)), list(c(dyn_int,
7.5, 9)), list(c(dyn_int, 7.5, 12)),
                    list(c(static, rep(0, time_points))))

nsimul <- 10000

gform<- gformula(obs_data = my_data, id = id,
                 time_points = time_points,
                 time_name = time_name, covnames = covnames,
                 outcome_name = outcome_name,
                 outcome_type = outcome_type,
                 compevent_name = compevent_name,
                 covtypes = covtypes,
                 covparams = covparams, ymodel = ymodel,
                 intvars = intvars,
                 interventions = interventions,
                 int_descript = int_descript,
                 histories = histories, histvars = histvars,
                 basecovs = c('age', 'sex', 'Grade'), nsimul = nsimul,
                 parallel = TRUE, ncores = parallel::detectCores()-1,
                 ref_int = 0,
                 nsamples = 200,
                 seed = 611)

print(gform)
```

Reference

1. McGrath S, Lin V, Zhang Z, Petito LC, Logan RW, Hernán MA, Young JG. gfoRmula: An R Package for Estimating the Effects of Sustained Treatment Strategies via the Parametric g-formula. *Patterns (N Y)*;1:100008.