

Online Appendix

A. Data generation

In our data generating process, for each individual, we used the following algorithm:

Step 1: Generate the counterfactual event time T_0 from a marginal Exponential distribution with event rate 0.01.

Step 2: Generate pre-baseline values of exposure and confounders, and baseline values of non-time-varying confounders:

$$A(0) = 0,$$

$$C_1(0), C_2(0) \sim N(0, 10) \text{ (independent)}$$

$$U_1(0) \sim \text{Bin}(0.4) \text{ OR}$$

$$U_1(0) \sim N(0, 10)$$

$$V_1 = -5 \times 1\{T_0 < 5\} + \epsilon, \epsilon \sim N(0, 10)$$

$$V_2 \sim \text{Bin}(p_V) \text{ where } p_V = 0.8 \text{ if } T_0 < 5, \text{ and } p_V = 0.4 \text{ if } T_0 \geq 5$$

Step 3: For each visit $k = 1, \dots, 10$, repeat the following steps:

3.1 Generate the confounders:

$$C_1(t_k) = -0.1 \times T_0 + A(t_{k-1}) + C_1(t_{k-1}) + \mathcal{N}(0, 2),$$

$$C_2(t_k) = -0.1 \times T_0 + 0.5A(t_{k-1}) + C_2(t_{k-1}) + \mathcal{N}(0, 2),$$

$$\text{logit}(P[U_1(t_k) = 1]) = 0.6475 + \gamma_1 \times T_0 + 0.4055A(t_{k-1}) + 0.6931U(t_{k-1})$$

(Binary U) OR

$$U_1(t_k) = -0.1 \times T_0 + 0.5A(t_{k-1}) + U_1(t_{k-1}) + \mathcal{N}(0, 2) \text{ (Normal } U).$$

3.2 Generate the exposure $A(t_k)$ as binomial with probability p_k , where:

$$p_k = \text{expit}(-1 + 0.405A(t_{k-1}) - 0.05C_1(t_k) - 0.05C_2(t_k) + \gamma_3U_1(t_k) - 0.15V_1 - 0.15V_2)$$

3.3 Simulate event status $Y(t_{k+1})$ at the end of interval $(k, k+1)$, and possibly event time T :

- * if $T_0 > \int_0^{t_{k+1}} \exp\{\beta A(t_j)\} dj$ then $Y(t_{k+1}) = 0$
- * if $T_0 \leq \int_0^{t_{k+1}} \exp\{\beta A(t_j)\} dj$ then $Y(t_{k+1}) = 1$ and
 $T = t_k + (T_0 - \int_0^{t_k} \exp\{\beta A(t_j)\} dj) \exp\{-\beta A(t_j)\}$.

Table A1: Parameters for simulations.

Scenario	$T_0 - U_1$ relationship	$Log(OR_{U1})$	n_{VS}	Event rate (VS)
1 Base	-0.1053	-1.0986	1000	0.1
2 U bias +	-0.1053	1.0986	1000	0.1
3 U stronger risk factor	-0.2231	-1.0986	1000	0.1
4 VS size 500	-0.1053	-1.0986	500	0.1
5 VS 1.5x event rate	-0.1053	-1.0986	1000	0.15
6 VS 2x event rate	-0.1053	-1.0986	1000	0.2
7 Informative censoring	-0.1053	-1.0986	1000	0.1
8 Continuous U	-0.1	-0.05	1000	0.1

B. Further results

Table A2: Simulation results: bias of the estimates for the log hazard ratio for exposure, for simulations reported in the main paper, and from each of the five methods shown in the main paper, as well as gold standard MSM. Results from models with un-truncated weights, and with weights truncated at 30 and at the 99.9th percentile, are shown.

Scenario	Trunc.	Gold Standard	Naïve	RC	RC + MR	MR-based MI	MR-based MI + A_{-1}
1 Base		-0.026	-0.240	-0.278	-0.146	-0.069	-0.064
	30	-0.046	-0.238	-0.277	-0.145	-0.070	-0.067
	99.9%	-0.055	-0.240	-0.278	-0.148	-0.073	-0.070
2 U bias +		-0.011	0.195	0.150	0.003	0.006	0.019
	30	-0.015	0.185	0.149	0.002	0.001	0.007
	99.9%	-0.017	0.175	0.141	-0.005	-0.008	-0.001
3 U stronger risk factor		-0.025	-0.388	-0.425	-0.141	-0.092	-0.087
	30	-0.053	-0.388	-0.424	-0.167	-0.094	-0.092
	99.9%	-0.065	-0.388	-0.422	-0.173	-0.097	-0.095
4 VS size 500		-0.032	-0.239	-0.279	-0.145	-0.066	-0.063
	30	-0.053	-0.242	-0.280	-0.146	-0.070	-0.067
	99.9%	-0.062	-0.244	-0.281	-0.150	-0.074	-0.071
5 VS 1.5x event rate		-0.045	-0.242	-0.282	-0.184	-0.118	-0.111
	30	-0.055	-0.243	-0.282	-0.185	-0.117	-0.115
	99.9%	-0.062	-0.244	-0.283	-0.188	-0.121	-0.118
6 VS 2x event rate		-0.038	-0.245	-0.282	-0.205	-0.145	-0.141
	30	-0.056	-0.245	-0.282	-0.205	-0.146	-0.143
	99.9%	-0.066	-0.247	-0.283	-0.208	-0.149	-0.147
7 Informative censoring		-0.013	-0.243	-0.283	-0.139	-0.060	-0.056
	30	-0.035	-0.243	-0.283	-0.140	-0.061	-0.059
	99.9%	-0.042	-0.244	-0.283	-0.143	-0.063	-0.061
8 Continuous U		-0.036	-0.156	-0.193	-0.098	-0.039	-0.051
	30	-0.049	-0.166	-0.195	-0.101	-0.059	-0.065
	99.9%	-0.055	-0.169	-0.200	-0.107	-0.063	-0.069

Table A3: Simulation results: relative standard deviation (SD) and relative root mean squared error (RMSE), relative to the naïve un-truncated MSM, for simulations reported in the main paper (*:The naïve estimate, with un-truncated weights, is used as the denominator for, respectively, the SD's and the RMSE's for all other estimators. E.g. 0.65 for SD of un-truncated RC estimator in the 1st row indicates that this SD was 35% lower than for the un-truncated naïve estimator.)

Scenario		Trunc.	SD relative to naïve untruncated					RMSE relative to naïve untruncated						
			Gold Standard	Naïve	RC	RC + MR	MR-based MI	MR-based MI + A_{-1}	Gold Standard	Naïve	RC	RC + MR	MR-based MI	MR-based MI + A_{-1}
1	Base	-	1.23	1.00	0.65	0.69	1.00	1.01	0.57	1.00	1.07	0.63	0.53	0.52
		30	0.80	0.73	0.61	0.65	0.75	0.76	0.41	0.94	1.06	0.62	0.43	0.43
		99.9%	0.74	0.67	0.57	0.62	0.71	0.72	0.39	0.94	1.06	0.62	0.43	0.42
2	U bias +	-	1.18	1.00	0.81	0.89	1.06	1.09	0.50	1.00	0.78	0.37	0.45	0.47
		30	0.96	0.85	0.79	0.87	0.91	0.92	0.41	0.93	0.77	0.37	0.38	0.39
		99.9%	0.89	0.79	0.75	0.83	0.87	0.87	0.38	0.88	0.73	0.35	0.37	0.37
3	U stronger risk factor	-	1.21	1.00	0.66	1.88	1.02	1.01	0.39	1.00	1.06	0.69	0.39	0.39
		30	0.79	0.68	0.59	0.72	0.74	0.74	0.28	0.97	1.05	0.47	0.33	0.33
		99.9%	0.73	0.63	0.55	0.65	0.70	0.71	0.28	0.97	1.04	0.47	0.33	0.32
4	VS size 500	-	1.34	1.00	0.66	0.76	1.05	1.03	0.63	1.00	1.08	0.64	0.55	0.53
		30	0.81	0.73	0.62	0.72	0.81	0.81	0.42	0.96	1.08	0.64	0.46	0.45
		99.9%	0.75	0.68	0.59	0.69	0.78	0.78	0.42	0.96	1.08	0.64	0.45	0.45
5	VS 1.5x event rate	-	1.11	1.00	0.66	0.68	0.97	0.99	0.54	1.00	1.07	0.74	0.62	0.61
		30	0.80	0.72	0.61	0.64	0.72	0.72	0.42	0.95	1.07	0.74	0.54	0.54
		99.9%	0.74	0.66	0.57	0.59	0.68	0.68	0.41	0.94	1.07	0.74	0.54	0.54
6	VS 2x event rate	-	1.15	1.00	0.59	0.61	0.98	0.97	0.58	1.00	1.04	0.79	0.71	0.69
		30	0.71	0.65	0.56	0.57	0.65	0.65	0.40	0.93	1.04	0.78	0.61	0.60
		99.9%	0.65	0.60	0.51	0.52	0.61	0.62	0.40	0.93	1.04	0.78	0.61	0.60
7	Informative censoring	-	1.20	1.00	0.64	0.69	0.98	0.99	0.62	1.00	1.05	0.60	0.54	0.54
		30	0.74	0.67	0.58	0.62	0.70	0.70	0.40	0.93	1.04	0.59	0.42	0.41
		99.9%	0.70	0.63	0.54	0.58	0.67	0.68	0.39	0.92	1.04	0.59	0.41	0.41
8	Continuous U	-	0.90	1.00	0.57	0.60	0.92	0.93	0.68	1.00	0.94	0.61	0.70	0.72
		30	0.54	0.53	0.45	0.50	0.54	0.55	0.45	0.82	0.91	0.57	0.48	0.49
		99.9%	0.50	0.49	0.41	0.45	0.52	0.52	0.44	0.82	0.92	0.57	0.47	0.49

C. Cohort inclusion criteria for MarketScan example

- At least one specific code for type II diabetes mellitus in outpatient/inpatient settings: ICD-9 code 250.X0/250.X2, between one year before to one month after cohort entry
- No code for type I diabetes (ICD-9 codes 250.X1/250.X3) or for unspecified diabetes code (ICD-9 code 250.0) in the one-year look-back period
- No record of pregnancy (ICD-9: 650) or gestational diabetes (ICD-9: 648.8) within 9 months prior or 6 months after the diagnosis date.
- No diagnosis of cystic fibrosis (ICD-9 277.0) any time prior to the diabetic diagnosis.

D. Exposure definition for MarketScan example

Exposure is defined as shown in Figure 1. If a prescription is given at visit j , we consider the following cases:

1. If a prescription at visit(j) lapses before visit($j + 1$), we consider exposed until the minimum of:
 - (a) visit(j) + days' supply + population median of prescription lengths/2 = 15, or
 - (b) visit($j + 1$).

Note that the case in Figure 1 (a) involves adding a record in which exposure lapses, which is *not* at a visit to a healthcare professional. In this case, the weight for MSM will be set at 1, since it does not represent an active physician decision.

2. If a prescription at visit(j) lapses after visit($j + 1$), we consider exposed until the earlier of:
 - (a) $\max(k)$ such that visit(k) < visit(j) + days' supply, or
 - (b) $\min(k)$ where there is a new prescription at visit(k). This implies that a new prescription will overwrite an old prescription.

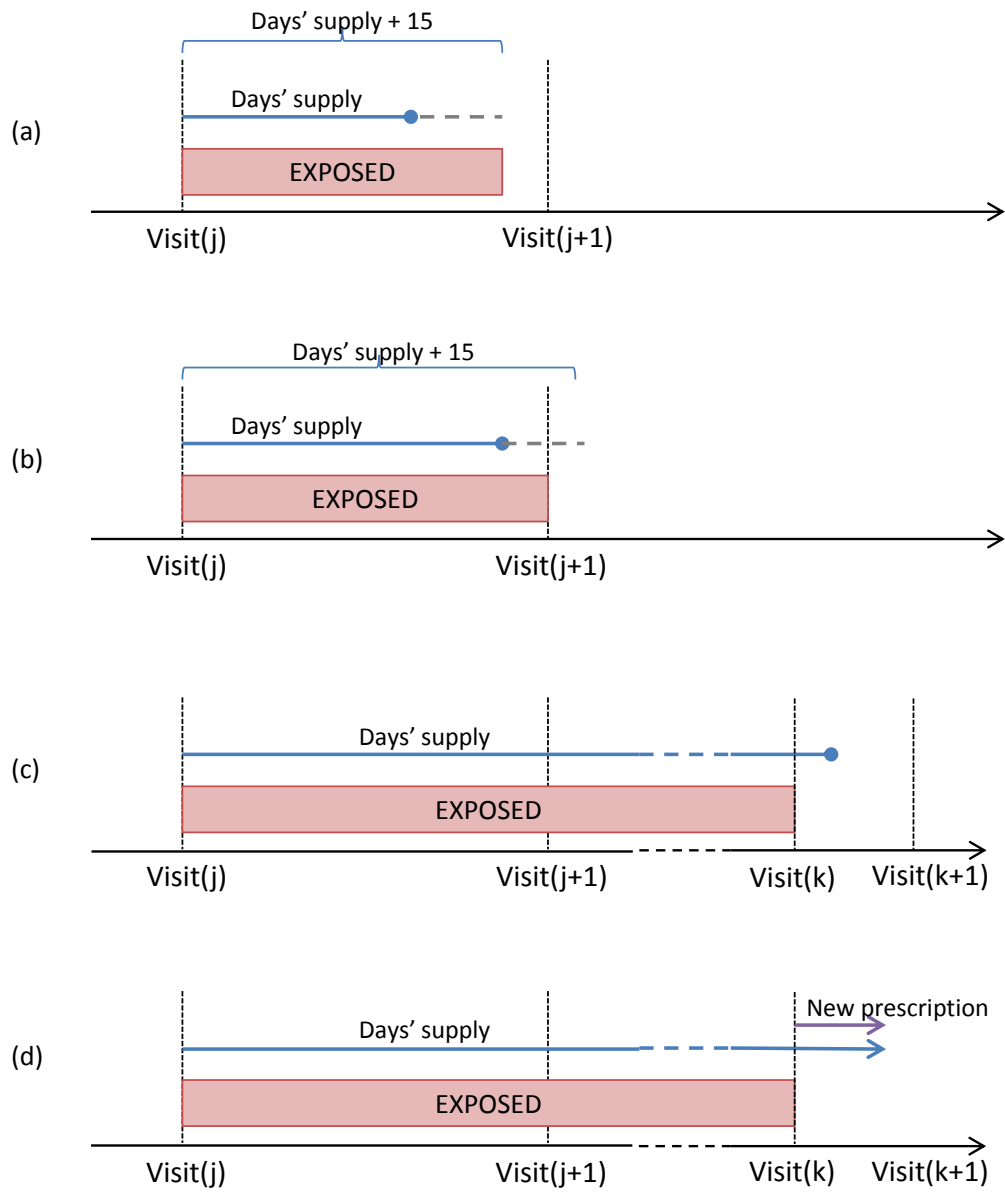


Figure 1: Exposure definition for applied example

E. HbA1c - Exposure relationship for MarketScan example

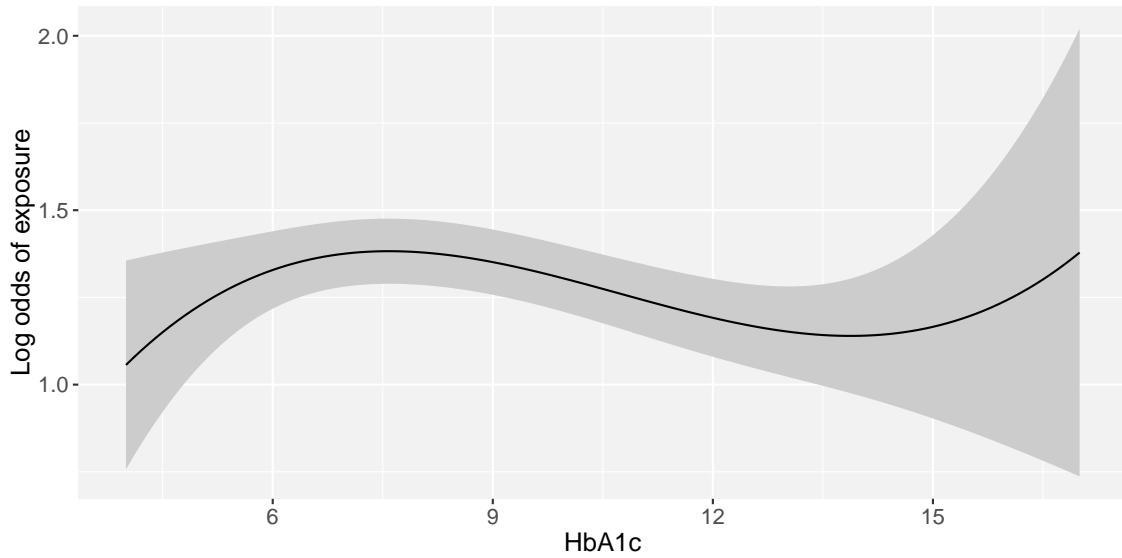


Figure 2: Estimated relationship of HbA1c to odds of being exposed to DPP-4 inhibitors, in the ‘gold standard’ propensity score model for calculation of weight. Other variables in the propensity score model were set at either zero (binary) or their mean (continuous).