Prognostic models for short-term annual risk of severe complications and mortality in patients living with type 2 diabetes using a national medical claim database

Alexandre Vimont[1,2] MSc, Sophie Béliard[3] PhD, René Valéro[3] Pr, Henri Leleu[2] PhD, Isabelle Durand-Zaleski[1] Pr

1, Assistance Publique Hôpitaux de Paris, URC-ECO, CRESS-UMR1153, Paris, France, 2, Public Health Expertise (PHE), Paris, France
3, Aix Marseille University, APHM, INSERM, INRAE, C2VN, University Hospital La Conception, Department of Nutrition, Metabolic Diseases and Endocrinology, Marseille, France

BACKGROUND

Patient-level clinical risk assessment for diabetic patients may be complemented with models developed from broader population and exhaustive information.

The objective of this study was to develop, validate and compare models predicting short term acute severe complications and mortality in patients with diabetes mellitus (DM) from the French national medical claims database.

METHOD

Database: "*Echantillon Généraliste des Bénéficiaires*", (EGB) from the French national healthcare database (SNDS).
Population: Adult patients with type 2 diabetes (T2D) identified through their history of treatment or hospitalization
1. Ongoing long-term disease (LTD) with a diagnosis of DM
2. Hospitalization with a primary or secondary diagnosis of DM
3. Treatment deliveries at ≥3 different dates of oral antidiabetics or insulin
Patients with type 1 diabetes were identified and excluded based on onset age of the disease.
Outcome
Severe acute CV complications: admissions with primary diagnosis of heart failure (HF), peripheral arterial disease, myocardial infarction (MI), stroke, unstable angina (UA), transient ischemic attack (TIA) and CV-related death
Any severe acute complications: CV complications + metabolic disorder such as ketoacidosis coma, ketoacidosis, acidosis and hypoglycemia, acute renal insufficiency, amputation and sepsis of any type
All-cause mortality: in-hospital and out-hospital death
Explanatory variables:
Socio-demographic: Gender, age, CMU-c, deprivation index.
Medical: LTD (Presence of LTD, duration and type), Comorbidities, adapted diabetes complications severity index (aDSCI) and antidiabetic treatments
Prognostic models
Logistic regression (LR) model, random forest (RF) model and a neural network (NN)
Statistical analysis
Pooling of repeated observations (PRO) was used to predict annual onset of each outcome. Model performances were assessed using discrimination (C-statistics), balanced accuracy, sensibility and specificity calculated on a 100-fold cross-validation basis.
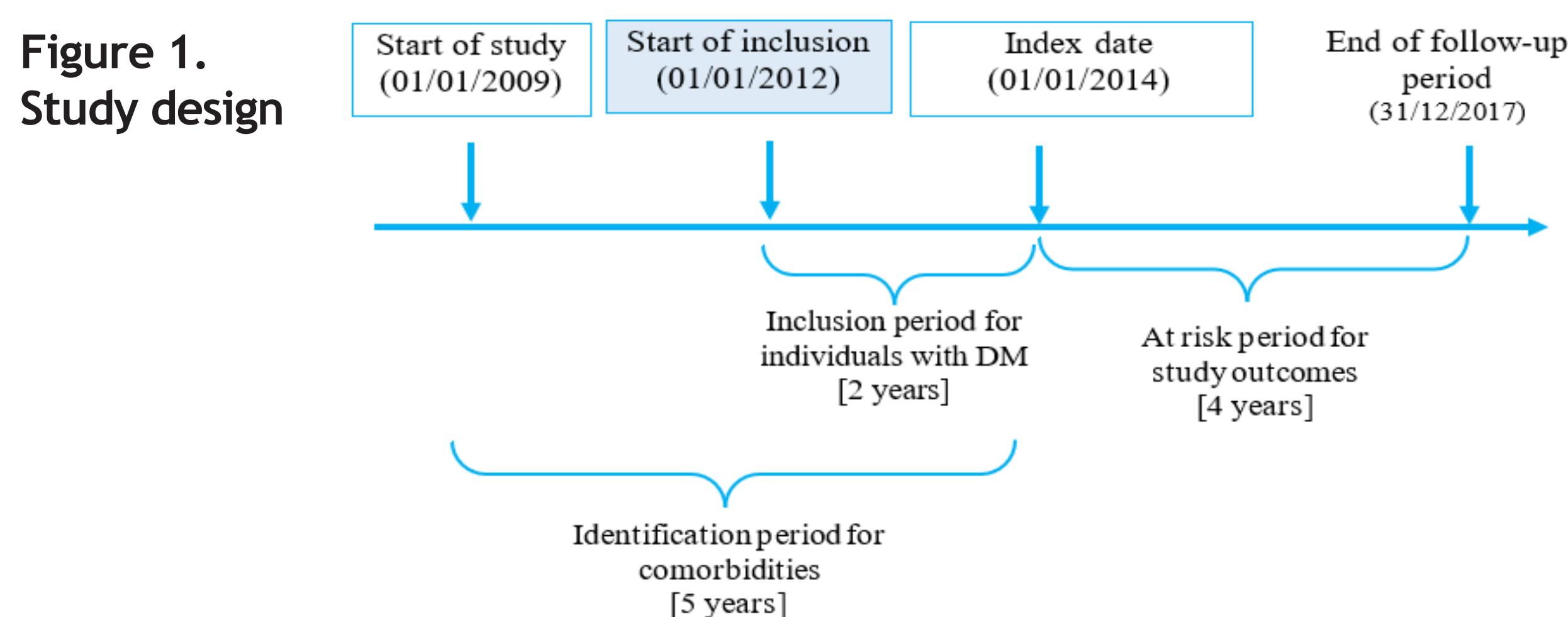

Figure 1. Study design

Table 1. Incidence of included severe complications

Type of complication	Number (N, %)	Incidence per 1,000 pts-yr
CV complications	**2,606 (77%)**	**31.8**
Heart failure	924 (27.3%)	11.3
Peripheral arterial disease	429 (12.7%)	5.2
MI	253 (7.5%)	3.1
Stroke	349 (10.3%)	4.3
Unstable Angina	239 (7.1%)	2.9
TIA	226 (6.7%)	2.8
CV death	186 (5.5%)	2.3
Other complications	**780 (23%)**	**9.5**
Metabolic disorder	252 (7.4%)	3.1
Sepsis	234 (6.9%)	2.9
Renal disorder	241 (7.1%)	2.9
Amputation	53 (1.6%)	0.6
All complications	**3,386 (100%)**	**41.3**

RESULTS

A total of 22,708 patients with T2D were identified, with mean age of 68 years and average duration of T2D of 9.7 years. Age, aDSCI, disease duration, diabetes medications and chronic cardiovascular disease were the most important predictors for all outcomes. Discrimination with C-statistic ranged from 0.715 to 0.786 for severe CV complications, from 0.670 to 0.847 for other severe complications and from 0.814 to 0.860 for all-cause mortality, with RF having consistently the highest discrimination.
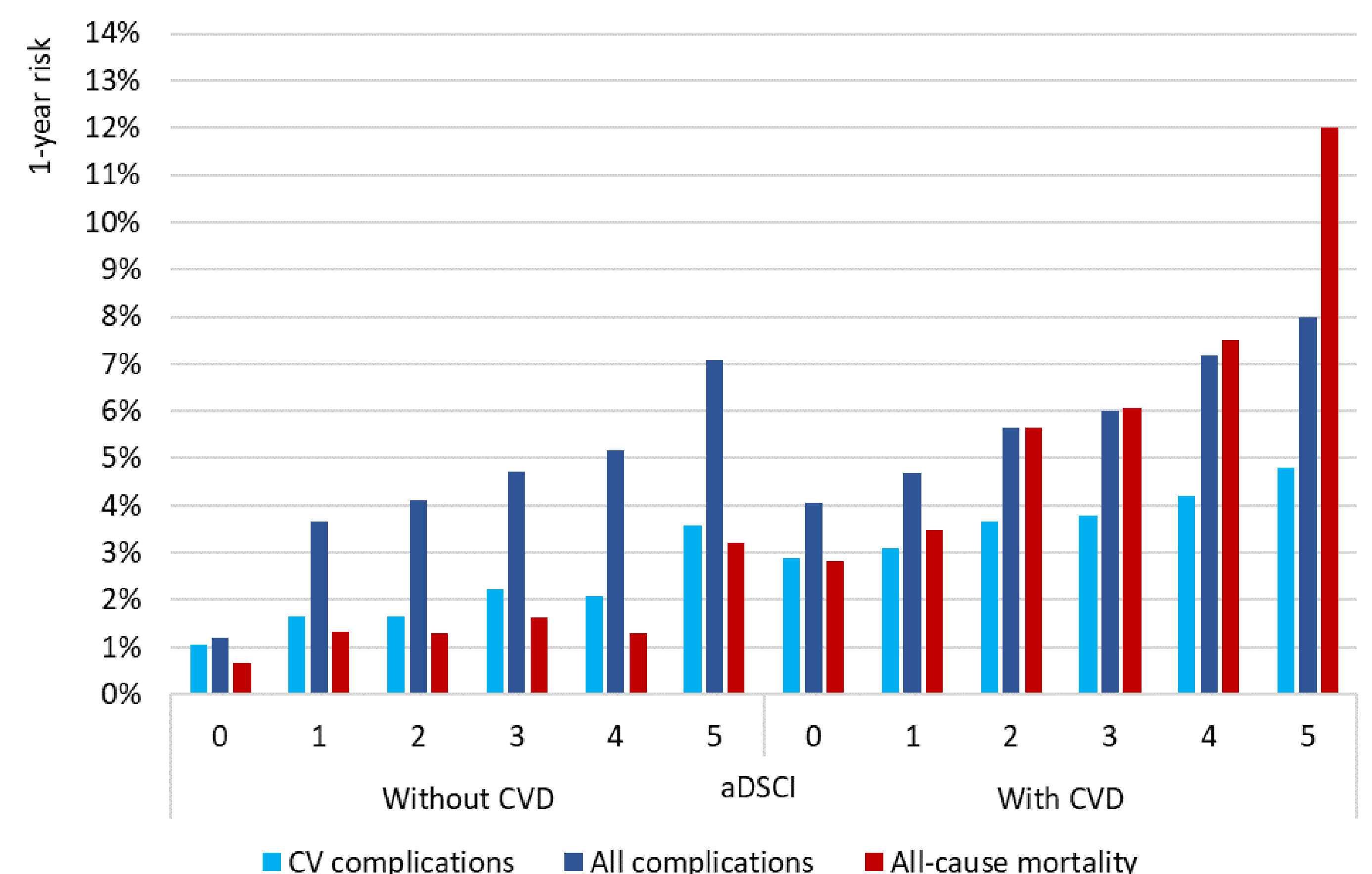
Table 2 - Comparative performance for severe complications risk prediction models

	Training sets C-statistic	Validation sets				
		C-statistic	95%CI	Se	Sp	Balance accuracy
CV complications						
LR	0.738	**0.715**	0.658-0.772	73%	60%	**0.665**
RF	0.944	**0.786**	0.747-0.825	71%	72%	**0.717**
NN	0.739	**0.738**	0.686-0.79	70%	67%	**0.685**
Other complications						
LR	0.738	**0.706**	0.653-0.759	59%	70%	**0.645**
RF	0.981	**0.847**	0.800-0.894	80%	81%	**0.801**
NN	0.732	**0.670**	0.626-0.714	74%	52%	**0.630**
All-cause mortality						
LR	0.823	**0.814**	0.769-0.859	76%	73%	**0.745**
RF	0.941	**0.860**	0.821-0.899	80%	79%	**0.794**
NN	0.845	**0.841**	0.791-0.891	82%	71%	**0.765**

Logistic regression (LR) model, random forest (RF) model and neural network (NN)

Our study demonstrated the feasibility of prognostic models based on claims information and the possibility to derive relevant risk factors such as diabetes duration (available with LTD duration), history of diabetes-related complications (aDSCI) and antidiabetic treatments which are not systematically included in routine clinical models.

Figure 2 - Annual median predicted risks by RF stratified by aDSCI and prior CVD

CONCLUSION

Models based on national medical claims data could reliably predict severe complications and mortality in patients with T2D, without requiring medical records or biological measures.

These models could be relevant for all health professionals to identify high-risk patients and optimize their monitoring, and more generally for payers to implement preventive measures.