

SUPPORTING INFORMATION

Methodological appendix

Markov model

Multistate chain Markov modelling is a mathematical process to simulate the progression of individuals through a series of states. The probability during a short time interval (δt) that a patient will move from state i to state j is $\lambda_{ij}\delta t$. We define λ_{ij} as the progression rate, the instantaneous risk of moving from state i to state j . These transition rates can depend on cofactors and are expressed by $\lambda_{ij}(z(t)) = \lambda_{ij}^{(0)} \exp(\beta_{ij}^T z(t))$, where $\lambda_{ij}^{(0)}$ represents the baseline transition rate from state i to state j ; β_{ij}^T represents the vector of the regression coefficient associated with the vector of covariates $z(t)$, affecting the transition i to j . The $e^{\beta_{ij}}$ are thus the vector of the hazard ratios of the covariates; we assumed that the covariates had a similar effect on all transitions, meaning that $e^{\beta_{ij}} = e^{\beta}$ for all i, j . Covariates may be time-dependent (time-inhomogeneous model).

In our representation of NAFLD disease, direct transitions can occur from normal liver (NL) to F0-steatosis, from one stage of fibrosis to the next, and from simple steatosis to NASH. The progression of steatosis, fibrosis or NASH are estimated as the result of the combination of progression and regression, as in other modelling studies of NAFLD.¹ We took into account four covariates, that are described below: sex, period of onset of being overweight, BMI category at 20 years old and the presence of diabetes. Finally, the “sex”, “age at the onset of overweight”, “BMI at 20 years-old” covariates are, by definition, constant over time, whereas the value of the covariate “presence of diabetes” is piecewise-constant, equal to 0 before the individual develops its diabetes, and equal to 1 from the time of the diagnosis of diabetes.

Estimation of the parameters

The parameters of the model, i.e. the baseline transition rates $\lambda_{ij}^{(0)}$ and the effects of the covariates e^{β} , are all back-calculated, by fitting the model to the training cohort using the maximum likelihood estimates method. To do so, each patient's disease progression was simulated in the model based on the individual duration of exposure and characteristics, assuming that the patient had a NL when he

initially became overweight, and evolved during an individual exposure time (defined below) toward the stage given by the biopsy. The resulting parameters are the set of parameters that maximize the likelihood on the training cohort, and are described in Table 2 (cf. main text).

Exposure time

The exposure time is the duration of exposure to being overweight/obesity ($\text{BMI} \geq 25$), when the patient is at-risk of developing NAFLD. It was calculated as the difference between the age at biopsy and the age of overweight onset. To do so, each of the four age categories of overweight onset has been assigned a given age: 6 for the category “before the age of 7”, 11 for “between age 7 and puberty”, 14 for “during puberty”, and 20 for “post puberty”. In the case of an age of onset of overweight “post puberty”, the previous procedure was refined using the known age for the onset of obesity, if available, assuming that the onset of overweight occurred 5 years before the onset of obesity.

For example, 40 years-old patients declaring age of onset of overweight “before 7 years old” are considered with $40-6=34$ years of exposure, whereas 40 years-old patients declaring age of onset of overweight “post puberty” and age of obesity at 30, are considered with $40-(30-5)=15$ years of exposure.

Model evaluation

Once the parameters are estimated, the resulting model is tested in both the training cohort and the testing cohort using a global error. To do so, each patient’s disease progression was simulated in the model based on the observed individual duration of exposure and characteristics, assuming that he had a NL initially. This leads to a predicted distribution of the patients at time of biopsy, among the 10 states, that allows to calculate a global error score between that predicted distribution, and the observed distribution. The global error score was calculated as the weighted mean of the relative differences between the number of individuals at each stage observed at the date of biopsy and the number predicted by the model throughout the period of exposure

The global error scores of the main analysis, detailed for each histological stage, are presented in Supporting Table 1.

Moreover, we used a 10-fold cross-validation approach to strengthen the results as much as possible. We randomly split the study cohort in 10 equal sized sub-cohorts. We then used one sub-cohort to test the model, fitted on the other 9 sub-cohorts. That process is repeated 10 times, with each of the sub-cohort used exactly once as the test cohort, thus providing 10 global error scores. The resulting mean global error score (\pm standard deviation, min-max) is 11.7% (\pm 3.0%, 7.9%-19.1%), that is close to the one presented in our main analysis. Moreover, we checked that the 10 sets of estimated parameters are very close to those of the baseline analysis.

Model prediction

The model can predict the current state of the liver state for any hypothetical patient without information on the patient's liver stage, by evaluating the probability of being in each disease stage, based on the patient's characteristics, and simulating disease progression from age at the beginning of overweight until the present age.

Moreover, it can also able predict the 5-year risk of disease progression, based on the current assessment of fibrosis. We chose to present the 5-year risk of the development of cirrhosis in two contexts: 1) partial information provided by non-invasive methods, i.e. only the stage of fibrosis without information on NASH status, such as a patient with fibrosis evaluated as <F2. In this case, the NASH status is unknown, and thus a weighted risk of cirrhosis is estimated taking into account the likelihood of having NASH at current evaluation or within 5 years.

Or 2) complete information provided by the biopsy i.e. the NASH status and the stage of fibrosis, for example a patient with a biopsy indicating F2-steatosis. In contexts 1) and 2), the model predicts the cumulative risk of cirrhosis at 5 years, by simulating disease progression from the fibrosis stage assessed at the current age.

Inclusion criteria

From 1994 to 2018, patients were prospectively included according to the following criteria: morbid or severe obesity (BMI > 35 kg/m²); at least one comorbidity factor (e.g., arterial hypertension or diabetes mellitus) for at least 5 years and resistance to medical treatment; absence of medical or psychological

contraindications for bariatric surgery; absence of current excessive drinking, as defined by average daily consumption of alcohol of 20 g/d for women and 30 g/d for men, and no history of excessive drinking for a period longer than 2 years at any time in the past 20 years; absence of long-term consumption of hepatotoxic drugs; negative screening for chronic liver diseases, including negative testing for hepatitis B surface antigen and hepatitis C virus antibodies and no evidence of genetic hemochromatosis; social health care insurance coverage; and age older than 18. Among the 2,735 included patients, some also accepted to be included in genetic studies, and signed up for the ABOS, "Atlas Biologique de l'Obésité Sévère".

The flow chart leading to the study cohort is presented in Figure 2.

Histological analysis

As the Brunt and NAS scores were developed and published in 1999 and in 2005, respectively, all liver biopsies performed before 2005 were reviewed and graded.² After 2005, all biopsies were graded prospectively with these 2 scores, and results were standardized in a report. Pathologists were blinded to clinical and biological data. Fibrosis was characterized by both the Metavir and Kleiner scores.^{3,4} In the baseline analysis, we used the Metavir score, defined by: F0, no fibrosis; F1, enlarged portal tract without septa; F2, enlarged portal tract with rare septa or bridging fibrosis; F3, bridging fibrosis without cirrhosis; F4, cirrhosis.³ The Kleiner score is defined as follow: F0, no fibrosis; F1, focal pericellular fibrosis in zone 3; F2, perivenular and pericellular fibrosis confined to zones 2 and 3, with or without portal or periportal fibrosis; F3, bridging or extensive fibrosis with architectural distortion and no clear-cut cirrhosis; and F4, cirrhosis.⁴

Sensitivity analysis

Six sensitivity analyses were performed to ensure the accuracy and gauge the uncertainty of the model. Firstly, we performed a probabilistic sensitivity analysis to provide a plausible interval to each prediction result. The plausible intervals correspond to the range of results (min-max) obtained from predictions with 1,000 independent transition matrix, sampled from the uniform distribution of the

baseline transition matrix uncertainty set. The uncertainty sets are defined as the intersection of the 95% confidence regions.⁵

Second, an alternative analysis was performed using the Kleiner scale instead of the Metavir scale to grade fibrosis. The fitted model is described in Supporting Table 2. The baseline transition rate from NL to F0 steatosis is similar to the transition rate of the Metavir model, as well as the baseline transition rate from non-NASH to NASH. The progression rates of fibrosis are unlike, as expected, due to the different definitions of the fibrosis stages in these two models (Supporting Table 2). The global error score of this model is 7.8% in the training cohort, and 9.1% in the test cohort (Supporting Table 3).

Third, we evaluated an alternative model, fitted on the non-imputed cohort only (n=1,390 individuals with complete data): similar parameters were obtained (Supporting Table 4). The global error score of that model on the non-imputed cohort is 6.6%.

Fourth, an alternative analysis was performed using simplified age of onset of overweight, to test the impact of the recall bias. It makes the assumption that patients may have difficulty remembering precisely their age of onset of overweight. A new model was fitted considering this variable as binary: before/after puberty. To do so, we merged the answers “before 7 years old” and “between 7 years old and puberty” in a category “before puberty” (n=727, 40.4%), and the answers “at puberty” and “post puberty” in a category “at/after puberty” (n=1074, 59.6%). The resulting model has a global error score of 5.3% on the training cohort, 11.3% on the test cohort, and its parameters are presented on Supporting Table 5. The results are similar to the ones of the main analysis.

Fifth, as our analysis was based on data from patients referred for bariatric surgery, so all living, with no information on mortality in this population before the management of severe obesity, the baseline analysis did not take into account mortality. An alternative model was fitted to the data, using predetermined probability of death and hazard ratios of the sex and diabetes impacting the mortality, extracted from a similar study of control patients matched to bariatric surgery patients.⁶ The mortality transition rate is supposed constant over all states, except in the cirrhosis state for which we assumed a mortality ten times higher taking into account the risk of decompensated cirrhosis or HCC. The fitted

model is described in Supporting Table 6; its progression rates and hazard ratios are similar to those of the main model.

Finally, an alternative analysis was performed using a detailed covariate for diabetes status. This allowed to explore the impact of controlled diabetes versus uncontrolled diabetes on disease evolution. The analysis was ran using a three-category diabetes-related covariate: “no diabetes”, “controlled diabetes” and “uncontrolled diabetes”. Uncontrolled diabetes was defined as a fasting plasma glucose two times higher than the upper limit (7 mmol/L). The number of patients with uncontrolled diabetes is 285 (15.8%). The results of this model are presented in Supporting Table 7. The estimated parameters are very close to those of the baseline analysis. Only the diabetes-related parameters vary, as expected. Presence of controlled diabetes increases the progression rate of the disease by 183% (HR = 1.83, 95%CI: 1.32-2.53) whereas presence of uncontrolled diabetes increases it by 208% (HR = 2.08, 95%CI: 1.73-2.49), versus non-diabetic patients. The global error scores of this alternative model are similar to those of the baseline model: 4.2% on the training cohort and 9.8% on the testing cohort (compared to 4.5% and 8.8%, respectively).

REFERENCES

1. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. *Hepatology*. 2018;67(1):123-133.
2. Lassailly G, Caiazzo R, Buob D, et al. Bariatric Surgery Reduces Features of Nonalcoholic Steatohepatitis in Morbidly Obese Patients. *Gastroenterology*. 2015;149(2):379-388; quiz e315-376.
3. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology*. 1996;24(2):289-293.
4. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. 2005;41(6):1313-1321.
5. Zhang Y, Wu H, Denton BT, Wilson JR, Lobo JM. Probabilistic sensitivity analysis on Markov models with uncertain transition probabilities: an application in evaluating treatment decisions for type 2 diabetes. *Health Care Manag Sci*. 2019;22(1):34-52.
6. Sjostrom L, Narbro K, Sjostrom CD, et al. Effects of bariatric surgery on mortality in Swedish obese subjects. *N Engl J Med*. 2007;357(8):741-752.

Supporting Table 1 – Results of the baseline model’s adequacy and error rates in the training and testing cohorts

	Training cohort			Testing cohort		
	Observed number of patients	Predicted number of patients	Error score	Observed number of patients	Predicted number of patients	Error score
Normal liver	186	207	11.3%	88	100	13.6%
F0 with steatosis	676	662	2.1%	348	328	5.7%
F1 with steatosis	187	180	3.7%	82	88	7.3%
F2 with steatosis	22	23	4.5%	7	11	57.1%
F3 with steatosis	10	11	10.0%	4	5	25.0%
F0 with NASH	40	41	2.5%	21	20	4.8%
F1 with NASH	34	33	2.9%	18	16	11.1%
F2 with NASH	18	15	16.7%	11	8	27.3%
F3 with NASH	16	19	18.8%	6	9	50.0%
Cirrhosis	18	16	11.1%	9	9	0.0%
Global error	1207	-	4.5%	594	-	8.8%

Supporting Table 2 – Parameter estimates for the alternative analysis using the Kleiner scale: baseline transition rates and hazard ratios of the covariates impacting these transition rates. Baseline transition rates correspond to a woman without diabetes, who became overweight during the post-puberty period and have a BMI at age 20 indicating that they are overweight.

	Parameters	95% confidence intervals
Baseline transition rates		
Normal liver → F0 with steatosis	8.02%	7.48%-8.58%
F0 → F1, with steatosis	1.79%	1.60%-2.03%
F1 → F2, with steatosis	2.65%	2.09%-3.36%
F2 → F3, with steatosis	5.60%	3.72%-8.47%
F3 with steatosis → cirrhosis	0.73%	0.16%-3.03%
Steatosis → NASH, regardless of fibrosis	0.59%	0.54%-0.64%
F0 → F1, with NASH	19.55%	18.17%-20.95%
F1 → F2, with NASH	7.39%	6.85%-7.93%
F2 → F3, with NASH	12.29%	11.40%-13.21%
F3 with NASH → cirrhosis	1.73%	1.55%-1.92%
Hazard ratios		
Male sex	1.50	1.41-1.61
Age of onset of overweight		
Before 7 years old	0.63	0.63-0.63

Between age 7 and puberty	0.63	0.62-0.63
During puberty	0.77	0.77-0.78
Post puberty	1	(ref)
BMI at 20 years old		
Normal weight	1	(ref)
Overweight	1.21	1.20-1.22
Obese class I	1.36	1.35-1.37
Obese class II & III	1.57	1.55-1.58
Presence of type 2 diabetes	1.68	1.66-1.70

Supporting Table 3 – Results of the alternative analysis using the Kleiner scale: adequacy and error rates in the training and testing cohorts

	Training cohort			Testing cohort		
	Observed number of patients	Predicted number of patients	Error score	Observed number of patients	Predicted number of patients	Error score
Normal liver	179	211	17.7%	87	101	16.1%
F0 with steatosis	648	625	3.5%	324	311	4.0%
F1 with steatosis	160	163	1.9%	77	81	5.2%
F2 with steatosis	38	37	1.8%	19	19	2.8%
F3 with steatosis	23	25	7.7%	10	13	25.1%
F0 with NASH	21	17	17.0%	11	9	20.9%
F1 with NASH	29	38	30.1%	13	19	44.0%
F2 with NASH	20	17	15.3%	13	8	35.7%
F3 with NASH	29	25	14.5%	13	13	4.0%
Cirrhosis	18	7	60.4%	9	4	56.6%
Global error	1165	-	7.8%	576	-	9.1%

Supporting Table 4 – Parameter estimates for the alternative analysis fitted on the non-imputed cohort only: baseline transition rates and hazard ratios of the covariates impacting these transition rates. Baseline transition rates correspond to a woman without diabetes, who became overweight during the post-puberty period and have a BMI at age 20 indicating that they are overweight.

	Parameters	95% confidence intervals
Baseline transition rates		
Normal liver → F0 with steatosis	8.42%	7.46%-9.42%
F0 → F1, with steatosis	1.61%	1.39%-1.88%
F1 → F2, with steatosis	1.72%	1.28%-2.30%
F2 → F3, with steatosis	3.44%	1.90%-6.17%
F3 with steatosis → cirrhosis	15.18%	12.02%-19.08%
Steatosis → NASH, regardless of fibrosis	0.72%	0.60%-0.85%
F0 → F1, with NASH	10.98%	8.41%-14.02%
F1 → F2, with NASH	7.41%	5.36%-10.14%
F2 → F3, with NASH	10.58%	7.90%-14.44%
F3 with NASH → cirrhosis	3.69%	2.18%-6.43%
Hazard ratios		
Male sex	1.35	1.22-1.48
Age of onset of overweight		
Before 7 years old	0.70	0.61-0.80

Between age 7 and puberty	0.79	0.68-0.91
During puberty	0.89	0.78-1.00
Post puberty	1	(ref)
BMI at 20 years old		
Normal weight	1	(ref)
Overweight	1.14	1.01-1.28
Obese class I	1.16	1.01-1.33
Obese class II & III	1.21	1.05-1.40
Presence of type 2 diabetes	1.53	1.31-1.78

Supporting Table 5 – Parameter estimates for the alternative analysis using simplified age of onset of overweight: baseline transition rates and hazard ratios of the covariates impacting these transition rates. Baseline transition rates correspond to a woman without diabetes, who became overweight during the post-puberty period and have a BMI at age 20 indicating that they are overweight.

	Parameters	95% confidence intervals
Baseline transition rates		
Normal liver → F0 with steatosis	7.93%	7.24%-8.64%
F0 → F1, with steatosis	1.59%	1.39%-1.79%
F1 → F2, with steatosis	1.29%	0.96%-1.73%
F2 → F3, with steatosis	5.44%	3.22%-9.12%
F3 with steatosis → cirrhosis	4.25%	1.62%-10.59%
Steatosis → NASH, regardless of fibrosis	0.60%	0.50%-0.72%
F0 → F1, with NASH	7.07%	5.16%-9.69%
F1 → F2, with NASH	8.76%	6.18%-12.96%
F2 → F3, with NASH	11.92%	8.70%-16.33%
F3 with NASH → cirrhosis	7.98%	5.84%-11.07%
Hazard ratios		
Male sex	1.44	1.33-1.55
Age of onset of overweight		
Before puberty	0.73	0.65-0.81

At/after puberty	1	(ref)
BMI at 20 years old		
Normal weight	1	(ref)
Overweight	1.08	0.99-1.17
Obese class I	1.13	0.99-1.29
Obese class II & III	1.40	1.24-1.58
Presence of type 2 diabetes	1.94	1.65-2.27

Supporting Table 6 – Parameter estimates for the alternative analysis using background mortality: baseline transition rates and hazard ratios of the covariates impacting these transition rates. Baseline transition rates correspond to a woman without diabetes, who became overweight during the post-puberty period and have a BMI at age 20 indicating that they are overweight.

	Parameters	95% confidence intervals
Baseline transition rates		
Normal liver → F0 with steatosis	8.00%	7.08%-9.17%
F0 → F1, with steatosis	1.53%	1.28%-1.78%
F1 → F2, with steatosis	1.24%	0.88%-1.76%
F2 → F3, with steatosis	3.77%	1.83%-7.66%
F3 with steatosis → cirrhosis	3.58%	0.87%-14.68%
Steatosis → NASH, regardless of fibrosis	0.63%	0.52%-0.77%
F0 → F1, with NASH	7.14%	4.99%-10.44%
F1 → F2, with NASH	7.69%	5.05%-11.89%
F2 → F3, with NASH	11.51%	8.16%-15.91%
F3 with NASH → cirrhosis	4.62%	2.25%-9.07%
Non-cirrhosis → Mortality	0.80%	Fixed
Cirrhosis → Mortality	8.00%	Fixed
Hazard ratios for NAFLD progression		
Male sex	1.50	1.34-1.67

Age of onset of overweight

Before 7 years old	0.67	0.57-0.79
Between age 7 and puberty	0.72	0.61-0.86
During puberty	0.78	0.67-0.92
Post puberty	1	(ref)

BMI at 20 years old

Normal weight	1	(ref)
Overweight	1.07	0.94-1.21
Obese class I	1.13	0.96-1.34
Obese class II & III	1.28	1.08-1.53
Presence of type 2 diabetes	1.69	1.45-1.98

Hazard ratios for background mortality

Male sex	1.61	Fixed
Presence of type 2 diabetes	1.61	Fixed

Supporting Table 7 – Parameter estimates for the alternative analysis using a 3-category covariate for diabetes: baseline transition rates and hazard ratios of the covariates impacting these transition rates. Baseline transition rates correspond to a woman without diabetes, who became overweight during the post-puberty period and have a BMI at age 20 indicating that they are overweight.

	Parameters	95% confidence intervals
Baseline transition rates		
Normal liver → F0 with steatosis	7.93%	7.07%-8.91%
F0 → F1, with steatosis	1.60%	1.38%-1.87%
F1 → F2, with steatosis	1.31%	0.97%-1.79%
F2 → F3, with steatosis	4.25%	2.39%-7.72%
F3 with steatosis → cirrhosis	5.49%	2.13%-14.58%
Steatosis → NASH, regardless of fibrosis	0.62%	0.51%-0.75%
F0 → F1, with NASH	7.14%	5.02%-10.02%
F1 → F2, with NASH	9.44%	6.39%-13.55%
F2 → F3, with NASH	11.81%	8.22%-16.70%
F3 with NASH → cirrhosis	6.25%	4.56%-8.27%
Hazard ratios		
Male sex	1.46	1.32-1.61
Age at onset overweight		
Before 7 years old	0.64	0.56-0.74
Between age 7 and puberty	0.69	0.59-0.81
During puberty	0.78	0.67-0.90
Post puberty	1	(ref)
BMI at 20 years old		
Normal weight	1	(ref)
Overweight	1.14	1.01-1.29

Obese class I	1.28	1.10-1.50
Obese class II & III	1.46	1.24-1.71
Presence of type 2 diabetes		
Absence of type 2 diabetes	1	(ref)
Presence of controlled diabetes	1.83	1.32-2.53
Presence of uncontrolled diabetes	2.08	1.73-2.49
