Stochastic simulation of clinical pathways from raw health databases

Martin Prodel, Vincent Augusto, Xiaolan Xie, Baptiste Jouaneton, Ludovic Lamarsalle

Stochastic Simulation of Clinical Pathways from Raw Health Databases

Martin Prodel[2], Vincent Augusto[1], Xiaolan Xie[1,3], Baptiste Jouaneton[2] and Ludovic Lamarsalle[2]

***Abstract*— This paper presents a method to automatically create stochastic simulation models of clinical pathways from raw databases. We introduce an automatic procedure to convert a process model, discovered with process mining, into an actionable simulation model. The concept of state charts is used and enriched to incorporate the distinctive features of health-care processes into the model. The clinical pathway model is used to simulate new patients' sequence of events. The resulting model is validated by comparing key performances indicators with historical data. Finally, we use the model to perform an automatically setup sensitivity analysis. The whole process is automated and can be used with any input data.**

I. INTRODUCTION

Clinical Pathways (CP) are a collection of activities that serve a common goal, such as consultation, rehabilitation or chemotherapy sessions. A CP describes the whole care journey of a patient across various health-care structures. Data related to CP are collected in hospitals for various purpose: in France, the national hospitalization database is primarily used for the pricing of care activities in hospitals, but it also contains a large amount of valuable data about the patient, his/her pathology and treatment.

The study of such data is important to reveal patterns of CP and a better understanding of the processes and of its potential improvements through new treatments, medicines, or medical devices. Health authorities intend to propose standardization of care processes for various operational purposes: organization of care activities, assignment of human resources, reducing practice variability, minimizing delays in treatments or decreasing costs while maintaining quality. Today, there is a will to go further than experts' opinions to answer these challenges. As such, evidence-based medicine has become paramount to medical decision making and clinical judgment.

The work presented in this paper is the last part of a large study consisting in applying a combination of data analysis and process mining [1] to build automatically a model of a CP for a certain cohort of patients. The reader is referred to [2] and [3] for further details about the automatic generation of CP models and conversion into simulation models. The goal of the present study consists in providing a "simulation toolbox" that can be used by health-care practitioners to learn about available health databases.

[1]Vincent Augusto and Xiaolan Xie are with UMR CNRS 6158 LIMOS, Mines Saint-Étienne 158 cours Fauriel 42023 Saint-Étienne cedex 2, France augusto@emse.fr, xie@emse.fr

[2]Martin Prodel, Baptiste Jouaneton and Ludovic Lamarsalle are with HEVA, Lyon, France mprodel@hevaweb.fr, bjouaneton@hevaweb.fr, llamarsalle@hevaweb.fr

[3]Xiaolan Xie is also with Shanghai Jiao Tong University, Shanghai, China xie@emse.fr

Most simulation models are handmade: the perception of the actual process is influenced by the modeler's experience, creating modeling biases. To avoid such biases, the idea of integrating process mining results to automatically generate a complete simulation model was initiated by [4] and was taken over by [5] and [6]. In [4], the focus is on the simulation model validation (whether generated or hand-made) to ensure sufficient quality of simulation results. The authors also highlight the challenges of automatic discovery of simulation models from event logs, including creating not too complex models, adding other perspectives to the flow perspective and adjusting the model for real-time simulation. An example is shown using Petri Net as the representation of their process models. Concerning Sensitivity Analysis (SA), it is the study of how input variations induce output modifications. SA is either local or global [7]. Local SA study the variations of a single parameter while other parameters remain fixed [8], and global SA study the output changes when all the parameters vary simultaneously [8]. However, such approaches are never automated, is also time-consuming and subject to bias. In this paper, we propose an automated approach to perform a sensitivity analysis on a model discovered from raw health databases. It allows to determine data variables which have the highest impact on considered key performance indicators (KPI).

The scientific contribution of this paper is twofold: (i) an automated stochastic simulation of clinical pathways directly connected to a raw health database; (ii) a method to analyze and discuss KPI for the health-care area through automated SA. A new validation procedure is proposed to assess the results on a real case study. CP analysis is performed using an automatic sensitivity analysis, taking into account the characteristics of the health data recorded in database. An extension to **a formal sub-class of state-charts** is also provided to take into account all special features of CPs.

The paper is organized as follows. The global methodology is described in Section II. The state chart formalism used to simulate discovered CP is detailed in Section III. The CP stochastic simulation toolbox is provided in Section IV. A case study is proposed in Section V. Finally, conclusions and perspectives are given in Section VI.

II. METHODOLOGY

In a previous work [2], we proposed a new approach to discover Clinical Pathways (CP) from the French national hospitalization database using process mining. The objective was to create the most representative process model of an event log under a constraint on the size of the model. In the literature, CP analysis from raw data was mainly done with

data mining or process mining techniques, both receiving an increasing attention in medical informatics. The next step of this research, consists in proposing a model that can be executed using simulation and automatically analyzed regarding relevant KPI.

This paper provides a comprehensive methodology to analyze and simulate such CPs as described in Figure 1. It uses an existing process model discovered from an event log (step 1) [2], a set of features found using health-care data analytics tools (step 2) [9] and a set statistical distributions (step 3). For that, we propose (i) a new procedure to automatically build a simulation model of patient CP from an event log of hospital stays, and (ii) a new subclass of state charts called "Clinical Pathway State Charts" (CPSC) to capture all the required material to efficiently simulate and evaluate the performances of any CP. This subclass is an extension of the one proposed in [3] to include health-care decision point analysis. A simulation procedure is proposed to perform automatic analysis (step 4). Such methodology may be applied to any database and any cohort of patients. Simulation of CPs brings new knowledge and allows scenario evaluation through design of experiments.

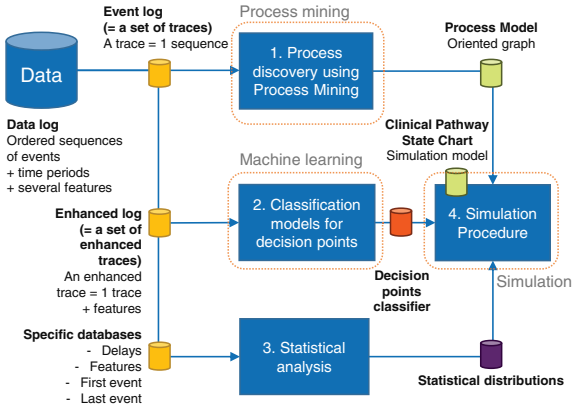

Fig. 1. Global scheme of our automatic modeling methodology

III. A SUBCLASS OF STATE CHARTS: CLINICAL PATHWAY STATE CHART

To simulate the clinical pathway of new patients, we use the general concept of state charts. It includes the definition of states, transitions, activation probabilities and state duration. We enrich this state chart definition with two new concepts: wait-states and care-states. Eventually, we introduce a new subclass of state chart that encapsulates all the specific features of a CP and simulates it.

Definition 1 (State chart): A state chart (SC) is a 4-tuple $M = (S, V, \zeta, \tau)$ where $S = \{s_1, s_2, \ldots, s_n\}$ is a finite set of states, $V \subseteq (S \times S)$ is a finite set of transitions, $\zeta : V \to [0, 1]$ is the probability of activating a transition, and $\tau : S \to \mathbb{N}$ is the time spent in a state.

We use state charts to model patient CPs. A patient is modeled using the concept of entity, defined by a set of features and an active state.

Definition 2 (Entity): An entity is a 3-tuple $u = (M, f, s)$, where $M = (S, V, \zeta, \tau)$ is a SC, $f = \{f_1(u), ..., f_x(u)\})$ is a set of assigned values for attributes from F (the set of trace's attributes) and s is its current state, $s \in S$.

Two types of states are defined to distinguish states related to hospital stays from states related to waiting periods between two stays.

Definition 3 (Care-state): A Care-state is a 2-tuple $s^c = (l, B)$ where l is a unique label and $B = \{(f_1, v_1), \ldots, (f_n, v_n)\}$ is the list of entities' features $\{f_1, ..., f_n\}$ to be updated in this state with new values $\{v_1, ..., v_n\}$, $(n \in \mathbb{N}^*)$. B includes at least a state-related cost that is used as a simulation performance indicator.

Definition 4 (Wait-state): A wait-state is a singleton $s^w = (l)$ where l is a unique label.

A care-state is related to a change in a patient's health condition and requires a medical response process during which the entity's attributes may change according to B. Finally, we propose a new subclass of state chart to describe clinical pathways, denoted Clinical Pathway State Chart.

Definition 5 (Clinical Pathway State Chart): A Clinical Pathway State Chart is a 6-tuple $CPSC = (S, V, \zeta, \tau, p, q)$:

1) $S = S_w \cup S_c$ where S_w is a finite set of wait-states and S_c is a finite set of care-states
2) $V \subseteq (S_c \times S_w) \cup (S_w \times S_c)$ is the set of transitions (vertexes) of the CPSC
3) ζ gives the probability of activating each transition given a state s and a set of features $F = \{f_1, ..., f_x\}$:

$$\begin{array}{rccc} \zeta & : & S \times F & \to & V^{|V|} \times [0,1]^{|V|} \\ & & s \times \{f_1, ..., f_x\} & \mapsto & (v_i, p_i) \end{array}$$

4) $\tau : S \to \mathbb{N}$ is the time spent in a state.
5) $p : S \to [0, 1]$ is the probability that the simulation starts at a given care-state, $\sum_{s \in S} p(s) = 1$
6) $q : S \to [0, 1]$ is the probability that the simulation stops after reaching a given state

A conversion procedure is proposed to automatically create an actionable CPSC. See [9] for a detailed methodological explanation of this step.

Example 1: We consider the process model given in Figure 2, formally defined as a causal net by $N = \{A, B, C, D\}$ and $E = \{e_1, e_2, e_3, e_4, e_5\}$. First, the conversion produces the state chart $CP = (S, V, \zeta, \tau)$ presented in Figure 2 with:

- $S = \{s^c_A, s^c_B, s^c_C, s^c_D, s^w_1, s^w_2, s^w_3, s^w_4, s^w_5\}$ where state s^c_i is a care-state related to node i and state s^w_j is a wait-state related to edge e_j. Care-states refer to hospital stays and wait-states to waiting between two stays.
- $V = \{(s^c_A, s^w_1), (s^w_1, s^c_B), (s^c_B, s^w_2), (s^w_2, s^c_C), (s^c_B, s^w_3), (s^w_3, s^c_D), (s^c_C, s^w_4), (s^w_4, s^c_D), (s^c_D, s^w_5), (s^w_5, s^c_D)\}$
- ζ and τ are initialized as null (defined at the 2nd step).

IV. CP STOCHASTIC SIMULATION TOOLBOX

In this Section, we provide the different elements provided in the CP Stochastic Simulation Toolbox, including an automatic setup of the simulation, validation and SA.

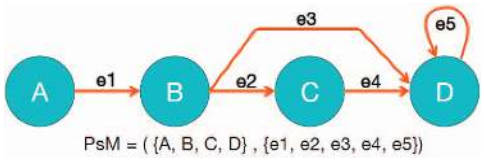

Fig. 2. 1st step of the conversion procedure - initial model

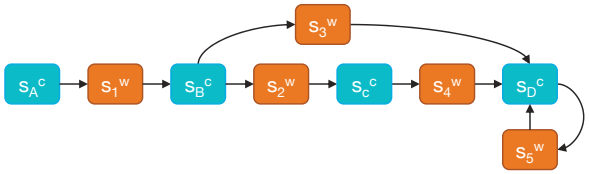

Fig. 3. 1st step of the conversion procedure - the output CPSC: care-states (blue), wait-states (orange) and transitions (black)

A. Simulation setup

1) Simulation procedure: The simulation procedure for a single entity is described as follows. First, a new entity is created. Its initial values of features and its initial state are drawn from the right random distributions. Then, the procedure computes the time spent in the current state and the next state based on the classifier. This is repeated until a stopping criterion is reached. Three stopping criteria are used: when a state has no outgoing transition, when the probability that a sequence stops within a given care-state is high and when an entity's sequence reaches the threshold of the maximal number of care-states. This threshold is set empirically as the size of the longest sequence seen in the data. When an entity enters a new care-state, its features (health condition, age, cost, medical history, etc.) are updated accordingly. In addition, whatever the state, the entity time-span is incremented with the time spent in this state.

2) Key Performance Indicators: Key performance indicators are used for simulation model validation and to test new situations through SA. Most KPIs are specifically chosen for a case study. For instance, in a lung cancer care process, the time between diagnosis and death is of major interest. Still, based on the definition of a $CPSC = (S, V, \zeta, \tau, p, q)$, we define a set R of generic KPIs:

- **KPI-1 :** The total (cumulative) time spent in care-states
- **KPI-2 :** The total time spent in wait-states
- **KPI-3 :** The number of visited care-states
- **KPI-4 :** The number of visits in each state

A 95% confidence interval is ensured when collecting such KPIs in the toolbox. To do so, the simulation procedure is replicated for many entities.

B. Validation

The model validation is done by comparing output values of KPIs with the same measure from historical data. More formally, let $CPSC$ be a clinical pathway state chart, let L be a log of historical patient sequences and let $R = \{KPI_1, \ldots, KPI_n\}$ be the set of key performance indicators chosen to validate $CPSC$, with $n \geq 1$. Then, for each KPI we compute the absolute difference between the model value and the data value:

$$\delta_i = |KPI_i^{CPSC} - KPI_i^L| \quad \forall i \in \{1, n\}$$

where KPI_i^{CPSC} is the average value of the Monte-Carlo replications for the KPI#i, and KPI_i^L is the value from the data. The simulation also produces an error value ϵ_i which gives the simulation confidence interval $[KPI_i^{CPSC}$ +/- $\epsilon_i]$.

Based on the difference δ_i between the model and the data, we propose to assess the model validity with a binary validation process: if the KPI value from the data belongs to the simulation confidence interval, we conclude that the model is valid regarding this KPI. Formally, for each KPI we define a validation function v_i:

$$\begin{array}{rcccl} v_i & : & \mathbb{R}^2 & \rightarrow & \{0,1\} \\ & & (\delta_i, \epsilon_i) & \mapsto & \begin{cases} 1 & \text{if } \delta_i \leq \epsilon_i \\ 0 & \text{else} \end{cases} \quad \forall i \in \{1, n\} \end{array} \quad (1)$$

The v_i function is computed for each KPI of R, thus validating or not the model for each KPI independently. Then, we aggregate these results to determine if the model is globally valid. One aggregation method is to choose a threshold on the minimum percentage of KPIs on which the model shall be independently valid. All the KPIs not being equally important for the validation, we introduce weighting factors $\beta_i \in [0, 1]$ for that purpose. Let $T_{min} \in [0, 1]$ be such a threshold, then a simulation model $CPSC$ is valid if inequality 2 stands:

$$\sum_{i=1}^{n} \beta_i . v_i(\delta_i, \epsilon_i) \geq T_{min} \qquad \text{with} \quad \sum_{i=1}^{n} \beta_i = 1 \quad (2)$$

To summarize, a new validation approach is proposed. A binary measure against the original data is proposed on predefined KPI using a user-defined validation threshold.

C. Automatic sensitivity analysis

A SA is the study of how input parameter variations impact the model outputs. It is a technique used to determine how an independent variable impact a dependent variable. In this paper, we propose an automatic generation of a SA for simulation models. First, we select eligible input variables that may impact the model outcomes. For each selected variable, a variation range is then determined and a systematic SA is performed for each value of the range.

1) Automatic selection of variables to evaluate: Variables related to clinical pathways are either **care-state attributes** or **instance attributes**. They were used to enrich the process model into a simulation model (instance attributes to learn decision point choices [9], state attributes to generate random distribution and evaluate indicators). Examples of care-state attributes are the length of stay, the medical diagnosis and the cost. Examples of instance attributes are age, gender and size. Each variable is one of the 3 following types: textual, categorical or numeric. Here, we do not consider textual variables (e.g. medical reports). Once we have identified these eligible variables, mainly based on the available field in the data set, we need to determine their variation range.

2) Variation range of the variables: A variable variation range depends on its type. The variation range of **categorical variables** and of **discrete numeric variables** can be determined automatically. These variables are described by a probabilistic distribution where each probability belongs to the interval $[0-1]$ and their sum is equal to 1. An incremental step Δ is set based on the available computation power (e.g. $\Delta = 0.01$). Then, for each possible value i of the variable, the associated probability p_i is set to 1 (other $p_j = 0$), and successively decreased such as $p_i = p_i - \Delta$ (and $p_j = p_j + \frac{\Delta}{n}$).

This procedure allows to test various configurations for the variable, without being fully exhaustive. The advantage is to at least test high values of each p_i. For a single categorical variable with K possible values, the required number of simulation runs for the SA is $K \times \frac{1}{\Delta}$.

The variation range of a **numeric continuous variable** is based on the distribution of historical data. Data are fit with the closest theoretical random distribution. The SA is performed by shifting this distribution: the same function (e.g. Weibull) and parameters are kept, but a translation factor T is added. For a given variable x, based on the standard deviation σ, the range for the translation factor T is $[-\sigma, +\sigma]$, with an incremental step Δ. In addition, we use truncated random distributions for all variables with a (semi-)bounded domain of definition (e.g. age is positive).

Finally, for any type of variable, the SA provides the outcomes of the simulation model derived for any of the tested input configuration. An example of SA is presented in the next Section (case study).

3) Summary: Our "automatic sensitivity analysis" **determines the most impacting variables on each output measure** by achieving the following: **(i) Automatic selection of variables to evaluate:** modeling variables (e.g. size of the model, confidence level), and case study variables, including care-state attributes and instance attributes; **(ii) Automatic generation of a variation range** for these variables. Depending on their type, we developed a procedure to generate relevant intervals; **(iii) Computation of single input-output relationship** for one output KPI.

SA give decision makers new insights about the uncertainties and their potential impact. It can discover hidden input-output relationships that were not straightforward to determine without a comprehensive model. Such information can be used to organize an action plan with the most relevant leverages regarding the target.

V. CASE STUDY

A. Cardiovascular diseases and implantable defibrillators

Heart diseases are one of the major health problems today. It was ranked as the first leading cause of death in the world in 2012 by the WHO. More specifically, cardiac arrhythmia is the most important cause of sudden cardiac death, affecting about 40,000 people per year in France and 300,300 in the USA. Implantable Cardioverter Defibrillators (ICDs) are medical devices that are indicated in two cases of severe cardiac arrhythmia: after a patient has experienced sudden death due to a ventricular tachycardia, or in prevention of it.

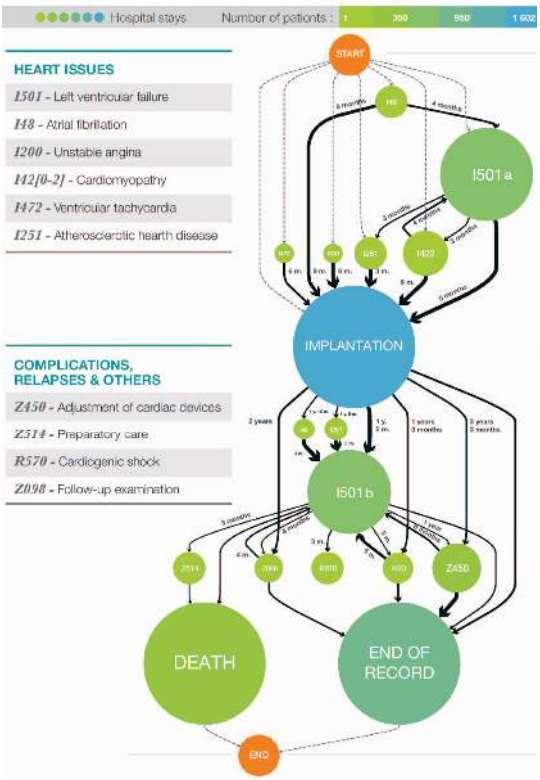

Fig. 4. Heart failure process model

Data were obtained from a single source: the French hospitalization database. It contains records of all hospitalization stays in France from 2006 to 2014 included, both in public and private sectors. It is an exhaustive database that represents 27 million hospitalization stays for 11 million patients annually. We selected the 1,602 patients implanted in France in 2008 and all their stays during a follow-up period of 2-years backward (2006) and 5-years afterward (2013). It represents a total of 16,931 hospital stays.

B. Model creation

The clinical pathway model of Figure 4 is a process model in the form of a causal net. We uses the conversion procedure presented in [2] to obtain a Clinical Pathway State Chart $CPSC = (S, V, \zeta, \tau, p, q)$. S and V are directly derived from the nodes and arcs of the causal net, ζ is made of the decision trees generated using machine learning approaches and τ was obtained with distribution fitting. The two last elements of the CPSC are p and q. They were obtained from the historical data and they are presented in Table I.

C. Model validation

The model was validated using the 4 Key Performance Indicators presented in Section IV.

The results for all the KPIs are presented in Table II, based on the simulation of 100,000 patients. Regarding **KPI-1 and KPI-2** (time related measures), the validation was

TABLE I

STARTING AND STOPPING PROBABILITIES OF THE CP STATE CHART

Wait States	Starting probability	Stopping probability
I48 (before ICD)	6.5%	0
I472	3.9%	0
I200	4.7%	0
I251 (before ICD)	11.1%	0
I422	11.0%	0
I501a	49.2%	0
Implantation	13.6%	0
Death	0	1
End of record	0	1

challenging because of the large variability of these measures in the original data.

The simulation model seems to underestimate the time spent by patients in care states (KPI-1) and in wait states (KPI-2) when using the mean and the standard deviation. However, the simulation results show a significant decrease in the variability (standard deviation) compared to historical data. The high variability of the data is explained by the presence of some outliers (e.g. a patient spent 4 years at hospital). However, it is difficult to remove outliers from the data since these patients may bring other interesting data to the case study. This is a difficulty when dealing with health data because some individuals may carry important information for the study. Variability reduction is an asset for the simulation model.

TABLE II

VALIDATION RESULTS FOR 5 MEASURES (100,000 PATIENTS)

KPI	Historical data Mean (+/- STD)	Simulation model Mean (+/- STD)	Simulation model 95% CI
KPI #1	65.80 days (+/- 88.10)	45.07 (+/- 29.18)	+/- 0.15
KPI #2	4 years 1 month (+/- 2 years 1 month)	3 years 8 months (+/- 9 months)	+/- 1.55 days
KPI #3	13.2 care states (+/-18.8)	11.7 (+/- 4.8)	+/- 0.025
KPI #4	Figure 5	68.5%	-

Regarding **KPI-3**, we obtained a close value of the number of care states in a trace sequence (11.7 versus 13.2). **KPI-4** is presented in detail in Figure 5. For each care state, the histogram shows the historical data (orange), the simulation result (blue) and the 95% confidence interval (red line). Based on a binary validation approach, the simulation model gets a validation score of 68.5% for KPI-4, which is above regular thresholds (50% or 66% for binary validation).

D. Sensitivity analysis

An automatic SA of input parameters was then performed for the simulation model described above, as described in Section IV. The input parameters are the patient features available in the case study data. It includes the 5 comorbidities, 2 non-medical patient characteristics and 1 variable related to defibrillators: (1) Hypertension, (2) Diabetes, (3) Obesity, (4) Kidney failure, (5) Cancer, (6) Age at implantation, (7) Gender and (8) Replacement rate.

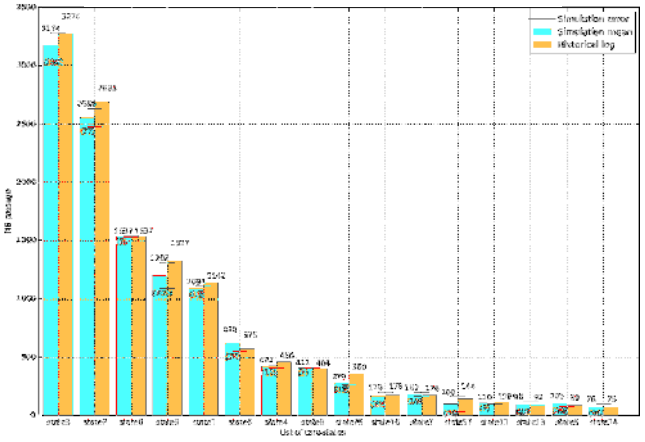

Fig. 5. Validation of the CPSC on KPI#4. States legend: 0 (implantation), 1 (end of record), 2 (I501a), 3 (I501b), 4 (death), 5 (I200), 6 (Z450), 7 (I420), 8 (Z098), 9 (I422), 10 (I251), 11 (I48-before), 13 (I472), 14 (I48-after), 25 (Z514), 57 (R570)

Figure 6 shows the result of the sensitivity analysis on KPI-1, the total time spent by patients in care-states. The impact of the 8 input variables is displayed on the same graph (8 curves), even if each variable varies independently (anything else equal). The y-axis represents the possible values of KPI-1 and the x-axis represents variations on the input variables. In order to plot and to easily compare the 8 curves, we normalized the possible values of each variable. The baseline point is when the modification coefficient of all variables is 1 (Green Arrow).

Among the 8 inputs, only 2 influence the time spent by patients in care-states: the age at implantation (red line) and the presence of kidney (grey line) failure. First, the impact of kidney failure is linear. The fewer patients have kidney failure (caution, a high coefficient of this variable means fewer patients), the shorter the total time spent in care-states (i.e. at hospital) will be. It can be explained by the necessity of having regular dialyses sessions (half a day) when having kidney failure. Regarding the age at implantation, the curve's shape appears more atypical. From the left, there is a fast increase in KPI-1 when the implantation age increases, then it stagnates, and it finally slowly decreases. This shape illustrates the fact that an increase in age is totally correlated with the need for more cares (the initial increase). After a certain threshold (mean age at implantation is 75), need for care on a 4-year term decreases because patients die faster.

Figure 7 shows the result of the sensitivity analysis on KPI-4, the number of times that state *cardiomyopathy before implantation* was visited by a patient. The outcome values are standardized for 1,602 patients. For this KPI, it is interesting to notice that no input variable significantly impacts the output values. It means that such cardiac issues are not dependent on factors that we incorporated in the model. A more in-depth backward analysis of patient history might turn out more relevant (more than 2 years before implantation).

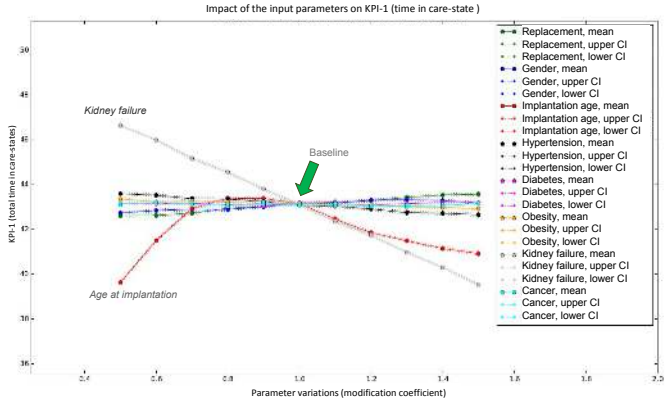

Fig. 6. Sensitivity analysis result - impact of 8 input variables on KPI-1

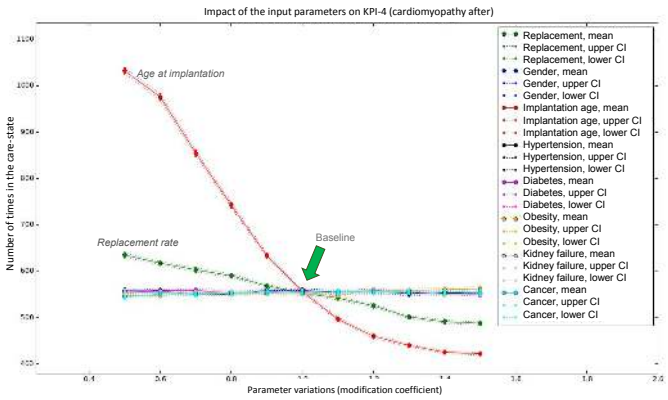

Fig. 8. SA: impact of 8 input variables on KPI-4 (b)

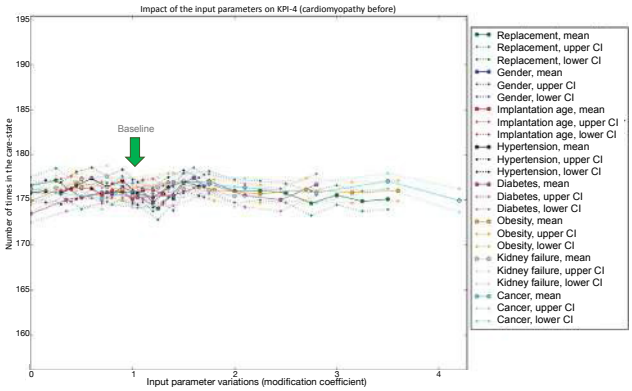

Fig. 7. SA: impact of 8 input variables on KPI-4 (a)

Similarly to the previous graph, Figure 8 shows the result of the sensitivity analysis on KPI-4, the number of times that state *cardiomyopathy after implantation* was visited by a patient. This time, two input variables show a direct impact on the output values: the age at implantation and the replacement rate. An increase in the age of patients when being implanted induces a substantial decrease in the number of times they have a cardiomyopathy (red line). This is probably explained by an edge effect of the long-term follow up of patients (4-5 years). Older patients with severe heart conditions have "less time" to develop other issues as the 2-year death rate is extremely high for patients over 75.

Regarding the replacement rate, an increase (i.e. more patients have a defibrillator replacement after few years) induces a linear decrease in the risk of having a cardiomyopathy (green line). It shows the importance of a close follow-up of patients and of anticipating the device malfunctioning.

VI. CONCLUSIONS

In this paper, we proposed the final phase of a formal procedure for the automatic conversion of a process model discovered from a health database into a simulation model. Our objective was to be able to generate new patients that are close enough to the historical data. We used the concept of state charts to integrate several perspectives of clinical pathways into a single simulation model. After the simulation model creation, we introduced several generic key performance indicators that can be used for model validation. We run the model to simulate the pathway of new patients so that we compare the output KPIs with the historical values from the event log. A validated model is finally used to perform sensitivity analysis and what-if scenarios evaluation. Sensitivity analysis provides insights about the determinant factors (input variables) that most impact the model's behavior (output measures). The resulting toolbox is a "ready-to-use" software for health practitioners. For future works, we plan to integrate resources in the model which means a deep modification of the simulation procedure since patients will be linked together. A further analysis of the new complexity of such update will be required. Also, we plan on creating a loop in order to automatically clean health data by removing outliers from the original data. Finally, an extension to multi-way sensitivity analysis may be an interesting contribution for practitioners.

REFERENCES

[1] W. M. van der Aalst, "Workflow mining: Discovering process models from event logs," *Computers in industry*, vol. 16, pp. 1128–1142, 2004.

[2] M. Prodel, V. Augusto, X. Xie, B. Jouaneton, and L. Lamarsalle, "Discovery of patient pathways from a national hospital database using process mining and integer linear programming," in *Automation Science and Engineering (CASE), IEEE Int. Conf. on*, 2015, pp. 1409–1414.

[3] ——, "Evaluation of discovered clinical pathways using process mining and joint agent-based discrete-event simulation," in *Proceedings of the 2016 Winter Simulation Conference*, Dec 2016, pp. 2135–2146.

[4] A. Rozinat, R. Mans, M. Song, and W. M. van der Aalst, "Discovering simulation models," *Inf. Syst.*, vol. 34, no. 3, pp. 305–327, 2009.

[5] N. Martin, B. Depaire, and A. Caris, "Event log knowledge as a complementary simulation model construction input," in *Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH), 2014 International Conference on*, Aug 2014, pp. 456–462.

[6] ——, "The use of process mining in a business process simulation context: Overview and challenges," in *Computational Intelligence and Data Mining, 2014 IEEE Symposium on*, Dec 2014, pp. 381–388.

[7] D. Maljovec, B. Wang, P. Rosen, A. Alfonsi, G. Pastore, C. Rabiti, and V. Pascucci, "Rethinking sensitivity analysis of nuclear simulations with topology," in *2016 IEEE Pacific Visualization Symposium (PacificVis)*, April 2016, pp. 64–71.

[8] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Introduction to Sensitivity Analysis*. John Wiley & Sons, Ltd, 2008, pp. 1–51.

[9] M. Prodel, "Process discovery, analysis and simulation of clinical pathways using health-care data," Ph.D. dissertation, Mines Saint-Etienne, 2017.