Behavioural Operational Research approaches for the assessment of surgical voucher acceptance decisions

Madalena Comporta de Almeida
madalena.almeida@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2021

Abstract

A surgical voucher transfer system has been implemented in Portugal, with the aim of guaranteeing a timely surgical response by healthcare services. Despite its promising potential, this programme has not performed satisfactorily, mainly due to low acceptance rates among patients.

This study encompasses a comprehensive behaviour operational research of vouchers' lack of acceptance, starting with an exploratory factor analysis, followed by structural equation modelling techniques, which enabled the validation of the health belief model to represent the psychological determinants of patients' intention to accept a transfer offer. Results showed that especially perceived barriers, but also cues to action, were the key factors responsible for ultimate behaviour. Furthermore, a simulation model was developed so that different behavioural scenarios could be tested. It was shown that a 50% decrease in patients' perceived barriers was enough to produce a 75.5% voucher acceptance in the simulation, leading to about less 32 000 patients waiting for their treatment, after a one-year simulation run.

Such tools of behaviour investigation are valuable to health authorities, in order to plan effective policies that can change patients' behaviour and to predict their possible impact on the national surgery waiting list.
Keywords: Surgery waiting list, Health-related behaviour, Health belief model, Agent-based simulation

1. Introduction

Excessively long waiting lists and times for surgery have been a problem of the Portuguese healthcare system over the years, compromising one of citizens' most fundamental rights – the access to timely care treatments [1]. Since 2004, the management of the Portuguese surgery waiting list (SWL) has been executed based on an integrated system, SIGIC, where maximum waiting time limits for surgery scheduling and delivery (TMRGs) are defined, according to different levels of patients' clinical priority. These limits, displayed in Table 1, must be obeyed by all hospitals.

Table 1: TMRGs and maximum waiting times until surgery scheduling, for different levels of clinical priority and types of pathologies [1].

Clinical Priority Level	Type of Pathology	TMRG	Maximum waiting time until scheduling
1	General	180 days	135 days
1	Cardiac	90 days	68 days
1	Oncological	60 days	45 days
2	General	60 days	30 days
2	Oncological/Cardiac	45 days	23 days
3	General/Oncological/Cardiac	15 days	5 days
4	General/Oncological	72 hours	As soon as possible

When a certain hospital does not have enough delivery capacity to schedule a surgery within those limits, the law dictates that a transfer offer must be issued to the patient of concern, in the form of a surgical voucher (SV), so that he/she can be treated at another facility. The SV contains a list of both public and private providers, from which patients can choose where they want to be transferred to.

Patients receive their first SV after waiting 75% (for priority level one) or 50% (for priority level two) of their TMRG. They can refuse such transfer and have a second SV emitted to them later, at 100% of their TMRG. In case they send no positive or negative response to health authorities before the SV's expiration date, their SWL enrolment gets cancelled [1].

Despite the potential that the SVs' policy could have had, by transferring part of the high demand for surgery from public to private healthcare services, it has fallen short to the expectations. Proof of that is the lack of improvements registered in the SWL's statistics over the last years. Data from 2017 to 2019 is displayed in Table 2, where it is possible to observe that the total number of surgeries performed has been continuously smaller than the number of new entrances. Furthermore, waiting times have not registered significant improvements, just like the percentage of waiting patients who already surpassed their TMRG [2].

Table 2: Statistics of surgical demand and supply in Portugal, between 2017 and 2019 [2].

	Indicator	2017	2018	2019
DEMAND	Nb. of patients entering the SWL	699 132	706 103	724 234
	Nb. of patients enrolled on the SWL	231 250	244 501	242 949
	Median waiting time (in months)	3.6	3.5	3.5
	Percentage of enrolled patients surpassing TMRG	32.3%	30.0%	32.1%
SUPPLY	Nb. of patients whose surgery was performed in a public hospital (without PPPs)	478 961	469 986	528 780
	Nb. of patients whose surgery was performed in a PPP	55 584	59 772	33 163
	Nb. of patients whose surgery was performed in a *hospital convencionado* or in a *hospital protocolado*	54 268	65 220	66 339
	Total nb. of patients whose surgery was performed	588 813	594 978	628 282

The main reason behind the SVs' system poor results seems to be a very low acceptance rate, since the big majority of patients either refuses or ignores their transfer offer, preferring to wait longer for treatment at their home hospital. In 2019, for example, the percentage of accepted SVs was only 18.8% [2].

Therefore, in order to guarantee a timely provision of care, it is crucial to perform comprehensive studies regarding SIGIC methodologies and their current efficacy. More specifically, it should be of great value to understand the reasons why such a high percentage of SVs end up not being accepted by patients.

A study based on a telephonic enquiry made to patients from the five Portuguese ARSs (Regional Health Authorities) who declined being transferred, back in 2007, determined the main motives behind SVs' refusals [3]. These were grouped in the following four categories: patients did not want to be treated by a different doctor or at a different hospital (34% of the cases), patients were unavailable to use the SV within its expiration date (30%), patients did not want to be treated outside their residence area (26%), and lack of information regarding the transfer programme's functioning (10%).

Although this previous study has provided some insight regarding the main motives behind SVs' refusals, it is still imperative to investigate which factors determine patients' decision-making process regarding this matter, so that health authorities may be enlightened regarding possible impacting reforms, capable of increasing vouchers' acceptance levels and its effects on the national waiting list.

2. Literature Review

BOR applied to healthcare:

Behavioural Operational Research (BOR) stands for a special stream of Operational Research (OR) dedicated to the study of behaviour and OR models' interrelation. Namely, a big part of the BOR literature explores the degree to which stakeholders' behaviour can impact system's functioning and performance, with a final goal of making some kind of intervention or just to better understand a certain problem. Some of the BOR analysis techniques include modelling and simulation, which will be further explored in this chapter [4, 5].

Among the many contexts to which BOR can be applied, healthcare emerges as an important field of interest, since it is made out of multiple complex systems, dependent on different stakeholders' behaviour (whether it is patients, practitioners, or health services' managers). Studies have shown that individuals must be modelled according to their condition of, as the name suggests, individual entities, whose actions are the result of different backgrounds, beliefs, intuitions and personal biases, rather than as completely rational beings [4, 6].

Concerning patients, their individual health-related behaviour can influence a varied set of clinical procedures, services, and policies. Studying it can be of an immense value to health authorities, in order to provide them with information regarding possible interventions.

As health-related behaviour can be a quite vague concept, it is important to better define it – it designates "overt behavioural patterns, actions and habits that relate to health maintenance, restoration and improvement", according to the Handbook of Health Behaviour Research [4]. Habits such as smoking, diet, physical activity, sexual behaviours, and alcohol consumption are examples of relatively well studied individual behaviours related to health matters. Nevertheless, BOR has also been applied to explore patients' dynamics with certain health services: resource consumption, screening compliance, vaccination intake, special programmes or tools' acceptance [6].

By studying the literature available regarding the application of BOR to healthcare settings, it is possible to observe common trends and, therefore, to group these articles into three subcategories, according to their main focus of analysis: modelling individual behaviour, simulating individual behaviour and, finally, a combination of both.

Modelling individual health behaviour:

Individual health-related behaviour can be modelled with the help of psychological models, which comprehend the factors that determine a certain action and how they might be intertwined. From those, the Health Belief Model (HBM) is one of the most often used in BOR research.

According to HBM (Figure 1), the likelihood of an individual taking a certain health-related action or adopting a given behaviour is related to seven dimensions: perceived susceptibility (perceived individual risk of contracting a disease/developing a condition), perceived severity (perception of the harmfulness of a disease and its consequences), perceived benefits (personal assessment of the benefits of adopting the behaviour to avoid a disease), perceived barriers (evaluation of the disadvantages of adopting the behaviour), cues to action (triggers that incentivize the adoption of the behaviour, such as physical symptoms, advice from family or health professionals, media campaigns), health motivation (someone's general interest and concern regarding personal health), and, in some cases, self-efficacy (someone's perceived capacity to carry out certain actions). HBM is deeply appreciated by its simplicity of concepts and its adequacy to cover health-related behaviour. On the other side, it lacks mathematical structure concerning the relationships between its constructs [5].

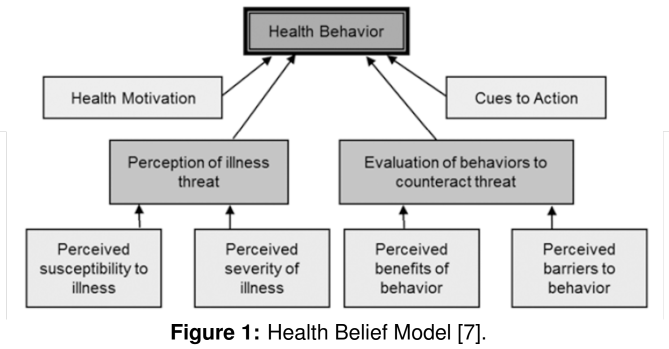

Figure 1: Health Belief Model [7].

Thus, some papers are mainly (or only) focused on modelling a specific health-related behaviour – dissecting it in detail, along with its contributing factors (which may be, for instance, psychologic, social or demographic), the complexity of the correlations among them, and their cumulative effects –, as an attempt to formally explain human actions and decisions concerning health. An example of that is a study conducted in Taiwan to understand the causes behind low acceptance rates concerning a technological tool for health self-management [8]. Almost four years after the implementation of the "My Health Bank", only about 3.2% of citizens were using it. A combination of HBM and other health and technology behaviour models was used, based on answers extracted from a survey, to identify the key factors behind people's intention to use (or not) their health passbook. The results proved that health belief factors, such as perceived susceptibility, rewards, perceived barriers, self-efficacy, and cues to action, as well as health literacy and technology acceptance elements (perceived utility and perceived easiness-of-use) were all key determinants of people's behaviour regarding health technology, more specifically, of their willingness to use the health passbook.

Simulation of individual health behaviour:

Simulation techniques are commonly used when applying OR to healthcare contexts, facilitating decision-making and problem-solving by mimicking how complex systems evolve over time. By simulating a certain reality, different policies can be tested before they are truly implemented, enabling cost-effectiveness assessments and comparisons of various possible scenarios. Furthermore, it allows the performance of clinical trials which would not be feasible to conduct in real-life [9, 10].

In the literature, it is possible to find many applications of different simulation techniques to healthcare settings. However, it is not so common for those simulation models to incorporate individual behaviour and, even when they do so, it usually lacks the deterministic dimensions that lead to those actions. That can be unexpected, especially if one thinks about how human health-related behaviour has the potential to influence health strategies' success. While superficially including human behaviour may be enough in some cases, particularly when the focus of study is on a larger dimension (like an evaluation of the overall functioning of a healthcare facility, for example, where it is important to include a few individual characteristics of patients in order to simulate how they will heterogeneously use that service and consume available resources, but there is no interest in studying the causes behind those individual behaviours), the same is not true when researchers intend to investigate ways of inducing behavioural change, since it is crucial to deeply understand how certain factors influence subjects' actions in order to plan and simulate this kind of interventions.

The two simulation techniques usually used to incorporate individual behaviour are Discrete-Event Simulation (DES) and Agent-Based Simulation (ABS). According to the literature, these are the most suitable for the task, since they admit the existence of individual units (entities or agents) which assume, at a given instant, a certain set of variables that characterize them. They can exchange messages with the environment surrounding them and take different actions accordingly to those stimuli and their personal traits. Therefore, each unit can act differently from the others, originating a heterogeneous set of "characters", appropriate to represent individual human behaviour. The main difference between the two methods is mostly related to the fact that, in ABS, agents can also exchange messages with each other [9, 10].

There are many examples in the literature of simulation studies which incorporate stakeholders' behavioural features. Lopes et al. [11] developed a simulation model to forecast the evolution of the medical workforce until 2050, in Portugal. Both supply and demand of physicians were taken into account in a robust model intended to help healthcare services planning. ABS was the chosen simulation approach, due to its capacity to account for doctors' individual preferences and decision-making, which play an essential part in the context of this study.

Modelling and simulating individual health behaviour:

As already stated, despite the fact that BOR has been applied to healthcare contexts multiple times over the years, few are the studies that contemplate both a detailed modelling and simulation processes of the psychological fundamentals of individual health behaviour. Two important examples of such comprehensive behavioural studies were related to patients' compliance with screening programmes – one for diabetic retinopathy, and another for breast tumors.

Regarding diabetic retinopathy, Brailsford et al. [12] developed a model to predict patients' intention to attend an examination session, mainly based on HBM. Such model enabled a complete description of various health-related behaviour dimensions (from threat perception and behaviour evaluation, to emotional, physical, cognitive and social aspects), although it lacked scientific structure. The mathematical expressions used to build the model were arbitrary and mainly based on assumptions; for that reason, the authors considered that study as a proof-of-concept, which showed the potential of including behavioural constructs on health policies' simulations, and did not claim the validity of the developed model. The results revealed that not considering individual variations of compliance with screening led to an overestimation of patients' attendance and, consequently, of that programme's beneficial effects. Furthermore, it showed potential to help health authorities understanding the ultimate impact of changes in different behavioural aspects and how to plan screening policies in order to achieve significant improvements.

Later on, Brailsford et al. [13] applied an alternative psychological model (Theory of Planned Behaviour, TPB), once again to predict patients' compliance, but for breast cancer screening sessions. Due to the application of TPB, this model had a sounder mathematical structure; however, it lost some behavioural constructs, namely the ones related to individual perceptions regarding severity and susceptibility to the tumour. This study also incorporated a model for the screening programme stages, as well as for tumour's progression. Simulations proved how noteworthy improvements regarding the effectiveness of the programme could be achieved not only by modifying the time interval between screenings, but also through patients' behavioural changes operationalized in the model (attitude towards behaviour, subjective norms, and perceived behavioural control).

Both these articles represented an important starting point when developing the methodology employed in the present study, due to their complete approach, where both modelling and simulation techniques were used as complementary methods to fully dissect the problems at hand. As it is still not very common to find both techniques being used together in behavioural studies, this research intends to contribute to further explore this kind of approach. Furthermore, no relevant BOR articles were found regarding voucher acceptance behaviour, which constitutes another gap in the literature this dissertation intends to close.

3. Methodology

HBM conceptual model:

The HBM model was adopted to represent patients' individual acceptance behaviour regarding SVs. All seven constructs of the model presented in Figure 1 were included and an equal number of conceptual hypotheses were established:

– H1: Perceived susceptibility is positively associated to intention to accept an SV;
– H2: Perceived severity is positively associated to intention to accept an SV;
– H3: Perceived benefits are positively associated to intention to accept an SV;
– H4: Perceived barriers are negatively associated to intention to accept an SV;
– H5: Cues to action are positively associated to intention to accept an SV;
– H6: Self-efficacy is positively associated to intention to accept an SV;
– H7: Health motivation is positively associated to intention to accept an SV.

In order to evaluate the validity of such hypotheses, a measurement instrument was developed and a quantitative study was conducted to measure the model's dimensions and connections.

Quantitative study:

The measurement instrument created to operationalize the HBM conceptual model consisted of an online questionnaire, shared on social media and also door-to-door (in order to approach the eldery). The survey consisted of two main sections: an initial part, that entailed some brief demographic questions to better characterize the sample being studied (age, gender, district, education level) as well as three other items concerning subjects' overall health perception and past experience on the SWL, and a more extensive second part, which was intended to capture and measure HBM's main constructs.

Since HBM constructs (latent variables) are unmeasurable, they have to be implemented as sets of multiple questionnaire items (indicators), which can indeed be considered observed variables and that indirectly reflect the HBM dimension they relate to. Thus, in the last part of the survey, respondents were faced with 26 enquiry items, organized into seven groups, as follows: perceived susceptibility (measured by two items of the survey, Q1 and Q2), perceived severity (measured by four items, Q3, Q4, Q5, and Q6), perceived benefits (three items, Q7, Q8, and Q9), perceived barriers (six items, Q10 to Q15), cues to action (five items, Q16 to Q20), self-efficacy (two items, Q21 and Q22), and health motivation (four items, Q23 to Q26). To answer, respondents had to select their level of agreement with each item statement on a five-point Likert scale (ranging from "completely disagree", level one, to "completely agree", level five). The majority of the questions was adapted to this project's context from the literature, where similar surveys were used to develop HBM models to other health-related issues [8, 14].

In addition, a final question was used to enquire participants regarding their willingness to accept an SV if presented with that hypothesis ("If you would receive an SV in the future, would you be willing to use it?"), in order to measure intention to behave, the ultimate dimension of the HBM model, implemented also in a five-point scale, where category one referred to "definitely no" and category five to "definitely yes".

Given the impossibility of specifically directing the questionnaire to patients currently enrolled on the Portuguese SWL due to privacy policies, this questionnaire was distributed among the general population. It should not significantly compromise the results as, in the end, everyone is susceptible of entering the waiting list for surgery and be presented with the option to use an SV.

Answers were collected from June 28th 2021 to August 25th 2021. Participants from Azores and Madeira were excluded from the study, since these regions manage their own SWLs independently from the rest of the country.

To start, a sociodemographic analysis of the obtained sample was performed in order to assess if any of these variables had an influence on SVs' acceptance decisions. Age, gender, ARS, and educational level were evaluated by Fisher's exact test of independence. To simplify this comparative evaluation, answers regarding intention to accept (Q27, the last question of the survey) were aggregated from a five-point Likert scale and divided in only two groups: "intending to accept", which included answers from levels three to five, and "not intending to accept", that considered replies of both values one and two.

The following survey question referred to respondents' self-evaluation of their health status, on a scale from zero (very poor) to ten (excellent). Its independence evaluation regarding willingness to use an SV was performed using the Mann-Whitney U test.

To conclude this first part of the questionnaire, participants were asked about their previous history on the SWL and past experience using SVs ("Have you ever been enrolled on the SWL before?", "If yes, have you ever received an SV? Did you use it?"). As these are, once again, nominal variables, Fishers' exact test was used to evaluate independence from intention to behave.

To understand if the 26 (Q1 to Q26) HBM-related questions truly implemented the seven-construct structure of the model, Exploratory Factor Analysis (EFA) was performed, using half of the collected sample. This type of data-driven factor analysis is commonly employed to evaluate newly developed measurement instruments and models, since it unravels the latent structure underlying a group of observed variables [15].

If a certain subset of observed measures presents high correlations among themselves, that group of variables must be affected by a common latent variable. This common factor influences individual scores on each measurable indicator and accounts for the covariance among them. Relationships between observed variables and a set of constructs are computed as factor loading estimates, λ_{ij}. The non-shared variance between a set of indicators of a same construct is attributed to measurement errors, δ_i. Factors' correlations, ϕ_{jj}, may also be computed during EFA [15].

When performing factor analysis, the first step must be to check the factorability of the variables' correlation matrix, with the Bartlett's test of sphericity (that must produce a significant p-value, lower than 0.05) and the Kayser-Meyer-Okin (KMO) test (where a value ideally above 0.70 should be obtained) [15].

Mardia's estimates were computed in order to verify if polychoric correlation coefficients should be computed, due to the ordinal nature of the survey's data. The chosen method for factors' estimation was principal axis factoring. Parallell analysis was used to determine the ideal number of factors to extract from the data. Oblique factor rotation (*oblimin*) was performed to facilitate output interpretation.

In this study, the cut-off value that determined which loading coefficients were considered significant was set to 0.50. During an EFA analysis, researchers should aim at finding a solution where each factor is significantly loaded by several variables, which, in turn, present significant loading values on only one factor and negligible values on all others (no cross-loadings) [15].

For an EFA solution to be considered acceptable, there is still one more criterion to be met. Each factor should obtain a Cronbach's alpha coefficient higher than 0.70, in order to prove to have internal consistency [16]. This measure examines if the whole subset of items concerning a certain construct measures the same thing consistently, based on their average correlations.

After EFA was used to investigate if the underlying factor structure of the data matched the one intended for the HBM-based model, it was necessary to confirm if such structure was also present among a different subsample (the remaining half of the total dataset). Cross-validation was used to ensure that the results were replicable and reliable. The causal relationships existent in the model also needed to be computed. To do all that, Structural Equation Modelling (SEM) was used, consisting on a two-step study: first with Confirmatory Factor Analysis (CFA), and afterwards with path analysis.

Unlike what happens during EFA, in CFA the factor structure under study is defined a priori. The model obtained is usually designated by measurement model and it specifies each factors' set of corresponding indicators (x_i), as well as their loading coefficients (λ_{ij}). Besides that, this model also includes error components (δ_i), and correlations among different factors (ϕ_{jj}) [17, 18]. Such errors can be defined as:

$$\delta_i = 1 - R_i^2, \qquad (1)$$

where R_i^2 stands for an item's determination coefficient (the proportion of that item's variance explained by its respective factor in the model). Some authors argue that R_i^2 is a measure of items' internal reliability (ideally close to one), while others attribute it to the square value of an indicator's standardized loading factor, λ_{ij}^2. Nevertheless, in general, R_i^2 and λ_{ij}^2 assume approximate values for a same item *i* [17, 18].

In this case, where the measurement scales were ordinal and a non-normal distribution of the data was plausible (that was, nevertheless, examined by Mardia's estimates), diagonally weighted least squares was the method chosen for the analysis.

Multiple model fit indices can be used to assess how well a sample supports an hypothesized conceptual model. As it is recommended in the literature, for this analysis there were produced different types of fit indices: absolute – χ^2/df –; relative indices – Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) –; and, finally, an index of populational discrepancy – Root Mean Square Error Approximation (RMSEA) [18, 19].

The χ^2/df estimate classifies a data-model adjustment as acceptable when it is lower than five and higher than two, good if it is between one and two, and very good if it is close to one. Regarding CFI and TLI, values between 0.8 and 0.9 are considered acceptable, between 0.9 and 0.95 good, and higher or equal to 0.95 very good. Lastly, RMSEA must produce a significant result (p-value below 0.05) lower than 0.10 in order to the adjustment to be considered acceptable – values between 0.05 and 0.10 indicate a good fit, and under 0.05 a very good fit [18].

Factors' composite reliability (CR) was computed to judge internal reliability of each construct. This estimate evaluates the consistency with which a set of indicators represents a certain latent variable. Values higher or equal to 0.7 are considered adequate. CR of a *k*-item factor is defined as:

$$\widehat{CR_j} = \frac{\left(\sum_{i=1}^{k} \lambda_{ij}\right)^2}{\left(\sum_{i=1}^{k} \lambda_{ij}\right)^2 + \sum_{i=1}^{k} \delta_{ij}}, \qquad (2)$$

where each items' error term, δ_{ij}, is given by equation (1) [18].

Regarding the evaluation of this study's validity, to examine if the measurement instrument operationalizes or not SVs' acceptance behaviour, different aspects need to be taken into account: factor validity, convergent validity, and, lastly, construct validity. Factor validity respects to items' factor loadings, which should be higher or equal to 0.5. Items' internal validity contributes to factor validity and it's given by the portion of an items' variance explained by its corresponding factor, which should be higher or equal to 0.25. Convergent validity ensures that the majority of items' variance is explained by the factor they are related to in the model. To assess that, factors' average variance extracted (AVE) must be determined, through the following mathematical expression:

$$\widehat{AVE_j} = \frac{\sum_{i=1}^{k} \lambda_{ij}^2}{\sum_{i=1}^{k} \lambda_{ij}^2 + \sum_{i=1}^{k} \delta_{ij}}, \qquad (3)$$

considering, once again, a construct defined by k indicators [18]. Obtained AVE values higher or equal to 0.5 are a sign that this type of validity is verified. To conclude, a model is said to possess discriminant validity when items related to a certain factor are not correlated to any other factors. Such property must be verified by confirming that a factor's AVE is higher or equal to the square values of its correlations with all other factors of the model [17, 18].

During this CFA, all forms of validity were examined. If the measurement instrument turns out to be valid, the identified factor structure should meaningfully represent HBM's dimensions.

After CFA, SEM proceeded with path analysis, where causal relationships between latent variables are established in a structural model [18]. Upon this analysis, the HBM relational model theoretically developed is further explored. Path trajectories between the HBM causal constructs defined during EFA (perceived threat, perceived benefits, perceived barriers, cues to action, and health motivation) and the ultimate behavioural dimension – intention to accept an SV –, associated to its one measurement item (Q27), were estimated. Therefore, their impact on behaviour could be evaluated.

SEM's global model results from the union of both the CFA's measurement model and the path analysis' structural model. Just like in CFA, goodness-of-fit tests (χ^2/df, CFI, TLI, and RMSEA) were performed to evaluate the quality of model's adjustment to the data.

All the above analysis was performed in R, from the demographic characteristics' independence tests, to EFA (*psych* package) and SEM (*Lavaan* package).

Simulation model:

Within the ambit of the second part of this dissertation's research, an agent-based simulation model of the waiting process for surgery implemented in Portugal was built in AnyLogic software. The goal of this simulation was to better understand the effects of behavioural changes in SVs' acceptance levels and in the reduction of the national waiting list. Patients were represented by individual agents, who assumed different states and underwent multiple transitions, representing the dynamics of the waiting process.

The model was based on data referring to 2018 and 2019. During 2019 – the year mimicked during the simulation run –, a total of 724 234 patients entered the SWL and joined the 244 501 that had transited from 2018 [2]. Since these different groups of patients were, at the beginning of 2019, in different points of the waiting process, they had to be modelled according to the different conditions they were in. Therefore, different agent-types and their corresponding subpopulations had to be implemented, in order to discriminate the different pathways possible for them to follow in the model.

The patients that entered the waiting list in 2019 formed the *patients19* population. The remaining patients, who were already on the waiting list at the start of the modelling year, were subdivided into three distinct groups, according to the stage of the waiting process they were in: *patients181* who, although already on the waiting list before 2019, had still not surpass 75% of their TMRG waiting; *patients182*, that had waiting times between 75% and 100% of TMRG; and, finally, *patients183* were the agents who had surpassed 100% of TMRG waiting for surgery. Data from a Portuguese public healthcare hospital was used to define the distribution of waiting patients between these three groups.

It is important to note that the used software did not support such big numbers of agents, which imposed the use of approximated values. In order to facilitate the interpretation of results, it was determined that each agent in the model represented 1 000 patients in real-life.

Agents of all categories were characterized by seven parameters (Sex, Age, Region, Priority, PBr, CtA, ItA). The distribution of the four demographic parameters was defined according to the description of the SWL population reported in a SIGIC's study [3]. PBr, CtA, and ItA represented each patient's behaviour constructs of interest – perceived barriers, cues to action and intention to accept. PBr and CtA values were set according to their respective indicators' scores observed in the HBM questionnaire sample. The mathematical expression defining ItA from PBr and CtA was implemented according to the results of SEM's global model, after path analysis.

Although more than one agent-type were considered in order to represent patients in different circumstances of their waiting journey, the sequence of states and transitions such agents experienced was always based on adaptations of the same overall representation of the waiting list's universe – Figure 2 shows the statechart scheme which illustrates the functioning of the general model and, more specifically, the one followed by agents from the *patients19* subpopulation.

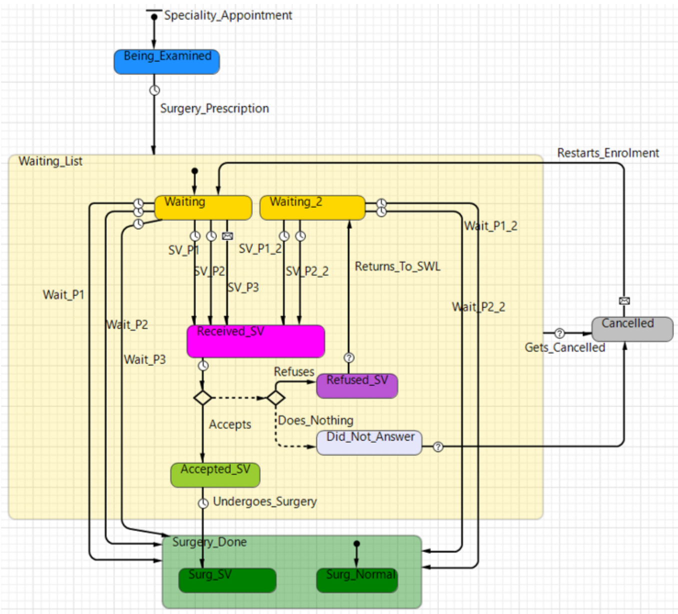

Figure 2: Statechart developed in AnyLogic to define agents' states and transitions in the simulation, as a representation of patients' possible pathways on the Portuguese SWL.

Agents enter the model environment at the "Speciality_Appointment" starting point, which symbolizes the first speciality medical appointment they go to. Once enrolled for surgery, they enter the "Waiting_List" state, which includes multiple sub-states. At the beginning of the waiting process, the agent is automatically directed to the

"Waiting" state, from where it can either proceed to "Received_SV" (in pink), in case that patient's surgery does not get scheduled before the deadlines defined by law causing an SV to be emitted, or to "Surgery_Done" (in green), if surgery is performed within the expected time period.

What determines if a certain patient is going to directly undergo surgery or, instead, to receive an SV, offering the option to be transferred to another facility, is a database table of true waiting times registered at a Portuguese hospital for that specific patient's priority level, which is defined in both transitions "Wait_P1" and "Wait_P2" (for priority levels one and two, respectively). On the other hand, transitions "SV_P1" and "SV_P2" have a fixed countdown of 135 and 30 days, respectively and according to the information displayed in Table 1 regarding the timestamps for the emission of SVs.

In cases where surgery waiting time is longer than the time for emission of a transfer document, patients move to the "Received_SV" state. There, agents' destiny is dependent on their personal decision regarding the SV they were offered – they can either accept it, decline it, or not answer. The conditions on which patients proceed to one of these states are defined by the transitions "Accepts", "Declines", and "Does_Nothing". As the subject of interest in this simulation study, "Accepts" was not fixed to a given percentage of acceptance, but rather implemented as a condition dependent on each agent's ItA score. ItA above a specific threshold denoted that patients' intention to accept an SV was high enough to originate that same behaviour, while ItA values lower than that limit would result in SV's refusal or in the absence of response.

To define that cut-off value, the survey's sample was used: participants' ItA estimates were computed and the top 18.8% of those scores was considered to result in effective acceptance (in order to correspond to the reality of 2019). The remaining SV-related transitions, "Refused_SV" and "Did_Not_Answer", were modelled as fixed parameters and also corresponded to percentages reported by the Ministry of Health in 2019: from the non-accepted vouchers, 67.2% were due to justified refusals and the remaining caused by an absence of answer from the patient [2].

From "Did_Not_Answer", agents are moved to the "Cancelled" state, representing the cancellation of patients' enrolment on the waiting list, as it is predicted in SIGIC's guidelines [1]. They can then be re-enrolled on the list by their hospital.

When a patient accepts the SV, he/she advances to the "Surgery_Done" state (in green), more specifically to "Surg_SV", which indicates that surgery was performed within the ambit of the transfer programme.

Finally, patients have the option to decline their transfer. When patients answer negatively, they are returned to their hospital's waiting list ("Returns_To_SWL"), without losing their previous position. Thus, this time they are not sent to "Waiting", but to "Waiting_2" (in yellow). From there, the destiny options are similar to the ones previously explained for the "Waiting" state.

Guards implemented as Java code in "Received_SV" guarantee that each patient gets a maximum of two SV offerings, as it is defined in SIGIC's regulation [1].

At any moment patients' enrolment on the waiting list can be cancelled (for a variety of reasons). The cancellation percentage was set to 13.7%, according to what was experienced in 2019 [2].

Controls were implemented to vary behaviour determinants' estimates. In an initial run, PBr and CtA were both set to 100%, in order to achieve the baseline acceptance rate registered in 2019, 18.8% – scenario A. After that, six other simulation runs, where behavioural components PBr and CtA were modified (PBr decreased and CtA increased, so that acceptance would also grow), were executed, in order to evaluate its effects on the SWL's outcomes after one year of simulation: scenario B (PBr set to 75%, CtA set to 100%), scenario C (PBr set to 50%, CtA set to 100%), scenario D (PBr set to 100%, CtA set to 125%), scenario E (PBr set to 100%, CtA set to 150%), scenario F (PBr set to 75%, CtA set to 125%), and scenario G (PBr set to 50%, CtA set to 150%).

4. Results & discussion

Characteristics of the sample:

The questionnaire obtained a total of 170 valid responses. Table 3 shows the demographic characterization of the sample.

Table 3: Demographic characterization of the studied sample.

Demographic variables		Total Sample	
		Frequency (total=170)	Percentage (%)
Age (years)	0-20	3	1.8
	21-40	79	46.5
	41-60	59	34.7
	above 60	29	17.1
Gender	Female	125	73.5
	Male	45	26.5
ARS	North	5	2.9
	Centre	15	8.8
	Lisbon and Tagus Valley	145	85.3
	Alentejo	3	1.8
	Algarve	2	1.2
Educational level	None	0	0.0
	Primary school	19	11.2
	Middle school	7	4.1
	High school	44	25.9
	University	100	58.8

The sample used in this study was compared to the one used in a SIGIC's study, conducted in 2007 [3]. It was observed an over-representation of young adults (21-40 years old) and an under-representation of older patients (above 60) in the present group of citizens enquired, as a result of the survey's distribution having been mainly carried out in online platforms. The middle-aged group (41-60) were equally represented in both studies. Regional representation also fell a little short in this specific research. Nevertheless, it is expected that these differences do not tremendously affect the results.

Age, gender, ARS, and educational level all reached p-values lower than the alpha significance level (0.05) in Fisher's test. Therefore, these demographic variables were considered independent from intention to accept.

Regarding health self-evaluation, a similar conclusion was drawn from the Mann-Whitney U test. Previous enrolments on the SWL also proved to be independent from intention to accept a future SV, and the same happened when patients were divided into two different groups depending if they had received an SV somewhere in the past or not. On the contrary, the fact that one had previously accepted or not an SV showed influence on the last question of the survey, presenting a significant p-value of 0.021, which suggested dependence between past and future SV decisions. This shows that, in spite of the low acceptance rates obtained among patients, the ones that took advantage of the transfer opportunity were not disappointed by the SVs system's functioning and were willing to use it again.

Factor structure:

Answers from one of the two sub-samples created and imported to R (containing 87 responses) obtained statistically significant values ($p<0.05$) for the absence of a multivariate normal distribution on variables Q1 to Q26, indicating polychoric estimates as the adequate correlation coefficients to compute during EFA. The KMO value (0.68) and Barlett's test ($p<0.05$) proved the factorability of the correlation matrix. Parallel analysis suggested five as the ideal number of factors to extract from the data.

All variables achieved relevant loading coefficients in a certain factor with the exception of Q21 and Q22. Q1 to Q6 were significantly loaded in the same factor, which was, therefore, named perceived threat (PTh) as a reference to the HBM super-construct they relate to. Following the same reasoning, variables Q7 to Q9 constituted the perceived benefits construct (PBn); Q10 to Q15, perceived barriers (PBr); Q16 to Q20, cues to action (CtA); and, lastly, Q23 to Q26 represented the health motivation (HM) dimension. No cross-loadings were identified.

Questions Q21 and Q22 were written with the intent of representing the self-efficacy variable of HBM, but did not obtain significant loading results. There seems to be no theoretical arguments that could justify the inclusion of such variables in further analysis since the health-related behaviour here considered concerns a personal medical decision, that does not require any special kind of skill or capacity (either psychological, physical, or technical) to be executed. That caused the exclusion of those two items along with the self-efficacy construct. This also meant that research hypothesis H7 was not supported by the model. New factor loading estimations for the remaining 24 variables were then obtained, but no significant differences were observed for the lasting variables' loadings. In total, the five factors structure accounted for 62% of the measurement instrument's variance.

High Cronbach's alpha values (between 0.82 and 0.93) were obtained for all factors, signalling that good internal consistency was achieved. Such results are important to support the reliability and validity of this model.

To wrap up the exploratory part of the study, it is possible to conclude that EFA exposed an HBM-similar latent structure from the answers collected with the questionnaire developed within this dissertation. The above described results show that a five-factor solution is adequate to represent the SVs' HBM-based model, at least for the subsample being used. In order to infer about the robustness of this instrument and its corresponding model when a different dataset is considered, the following step was to perform CFA with the remaining subsample of survey's answers.

Measurement model:

The remaining half of the original survey's sample (made out of 83 subjects) was imported to R. Mardia's multivariate normality revealed, once again, that polychoric correlation coefficients should be considered.

The results confirmed that the 24 considered indicators represented satisfactorily well the structure of five factors exposed during EFA. Nevertheless, some items showed standardized factor loadings below the cut-off of 0.50 and, for that reason, were eliminated at this stage of the study – Q14 and Q15, from PBr. Then, CFA was redone witout them.

As it can be observed in the a) part of Figure 3, factor validity is assured by items' standardized loadings (all higher than the defined cut-off) and by their satisfactory individual validity (all items exhibited λ_{ij}^2 values of at least 0.25, and the majority of them also presented R^2 estimates higher than 0.50, meaning that a major part of their variance was explained by their corresponding factor in the model). However, some exceptions were detected: indicators Q19, Q20, and Q24 obtained lower determination coefficients, of 0.293, 0.248, and 0.371, respectively. These items were, nonetheless, satisfactorily correlated to their corresponding factors – CtA and HM –, with loading coefficients equal to or slightly over the inclusion threshold. Since factor validity was not compromised, they were kept in the analysis.

PTh and PBn were the two constructs with higher correlation (0.53), followed by PBn and HM (0.37), CtA and HM (0.36), PBn and CtA (0.34), and PTh and CtA (0.22). PBn and PBr showed a negative correlation (-0.26). The remaining values were all below $|0.09|$.

Regarding the quality of model's fit, tests' results indicate a good to very good level of adjustment ($\chi^2/df = 1.463$, $CFI = 0.994$, $TLI = 0.993$, $RMSEA = 0.075$).

CR estimates (0.96 for PTh, PBn, and PBr; 0.89 for HM; 0.83 for CtA) indicate good composite reliability for all constructs.

Besides the already addressed (and proved) factor validity, convergent and discriminant validity were also evaluated. All factors' AVE values were higher than 0.50, therefore suggesting the presence of convergent validity. In addition, for every factor, its AVE was considerably higher than the square values of its correlations with any other factor, as it may also be observed in Table 4. Thus, the developed model also presented discriminant validity.

Table 4: Comparison of each factor's AVE value (in bold) and its correlation square values with all other factors.

	PTh	PBn	PBr	CtA	HM
PTh	**0.73**				
PBn	0.29	**0.89**			
PBr	0.00	0.07	**0.67**		
CtA	0.05	0.11	0.01	**0.60**	
HM	0.01	0.14	0.00	0.13	**0.69**

To conclude, this CFA corroborated the structure unravelled by EFA of correlations between HBM's indicators and constructs, with the exception of two items which did not produce significant correlations to remain in the model. A good adjustment to the data was achieved, showing sound proofs of reliability and validity – assuring that the final considered questionnaire items did indeed measure the HBM constructs they were related to in the measurement model, and did so in a consistent way.

Global model:

The measurement model was completed with a representation of the causal relationships among latent variables (which include the behaviour predictor, intention to accept (ItA), measured by Q27), originating the SEM's global model presented in Figure 3.

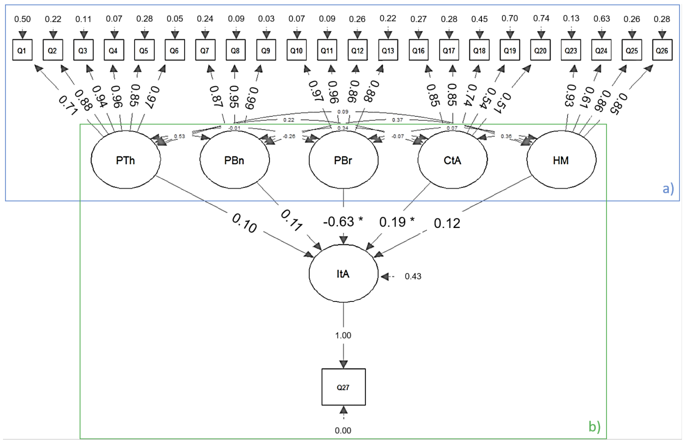

Figure 3: Global model obtained during after path analysis, resulting of the union between the measurement model, a), and the structural model, b). Correlations among factors (ϕ_{jj}) are represented as double-ended arrows, while factor loadings (λ_{ij}) are directed arrows from each factor to their corresponding indicators. Directed arrows from the five independent constructs to ItA are represented along with their respective path coefficients (γ_j), signalled with an * when statistical significance was achieved (p<0.05). Indicators' error components (δ_i) are represented as dotted arrows pointing to each measured variable's square.

The global model presented a good fit to the data, as it can be confirmed based on the quality adjustment tests' results ($\chi^2/df = 1.368$, $CFI = 0.995$, $TLI = 0.994$, $RMSEA = 0.067$). Thus, the measurement instrument here developed, and operationalized through the model displayed in Figure 3, proved to be adequate to represent SVs' acceptance behaviour.

Path analysis revealed that "PBr -> ItA" was the most significant trajectory of the model, with a standardized weight of -0.63 (p=0.000). "CtA -> ItA" came in second place, with a trajectorial weight of 0.19 (p=0.025). All the other three constructs (PTh, PBn, and HM) did not obtain significant weights in relation to ItA: 0.10 (p=0.179), 0.10 (p=0.346), and 0.12 (p=0.126), respectively.

Therefore, regarding the six research hypotheses still under study, only H4 ("PBr are negatively associated to ItA") and H5 ("CtA are positively associated to ItA") could be accepted. Although PTh, PBn, and HM had reached positive causal weights in relation to ItA, as hypothesised, such values did not show statistical significance and, because of that, H1, H2, H3, and H7 could not be accepted.

The fact that PBr significantly influenced patients' ItA an SV corroborated the conclusions obtained in a SIGIC's study [3], where some of those barriers were pointed by patients themselves as the main reasons for SVs refusals. Furthermore, incentives from people whose opinions patients value, such as family, friends, or their family doctor, as well as a deterioration of their health condition while waiting for surgery – which together form CtA – seemed to also have a somewhat important impact on their final decision to use an SV.

Simulation model:

In order to evaluate the AnyLogic model's capacity to reliably simulate the reality of the Portuguese SWL, it was mandatory to start this simulation study with scenario A, since it is equivalent to the real conditions experienced in 2019 and reported by the Ministry of Health. At the one-year mark of the simulation running, 18.8% of the emitted vouchers had been accepted, just like it was registered in 2019 [2]. The real and simulated surgical delivery statistics were considerably close – the total number of surgeries performed was estimated with an error of 0.5%, while the number of patients leaving the SWL differed in 1.7% and the number of patients remaining on the SWL in 6.6%. Slightly bigger differences were registered in the number of SVs emitted (10.4%) and accepted (10.7%), as well as in the number of enrolment cancellations (15.3%). The outputs of the model seemed acceptably close to the ones observed in reality, confirming the adequacy of the developed SWL's model to study the effects of changes in patients' behaviour regarding a transfer offer.

Figure 4 shows the changes in SVs' general acceptance percentage among patients, detected for scenarios B to G. As expected, variations in PBr (B and C) had a much stronger impact in ItA than modifications of the CtA construct (D and E). In fact, while a 25% decrease in PBr made the acceptance percentage more than duplicate, an equivalent growing in CtA did not have a significant impact on the considered output. It is also important to mention the impressive value of 75.5% of acceptances obtained after a 50% reduction of PBr.

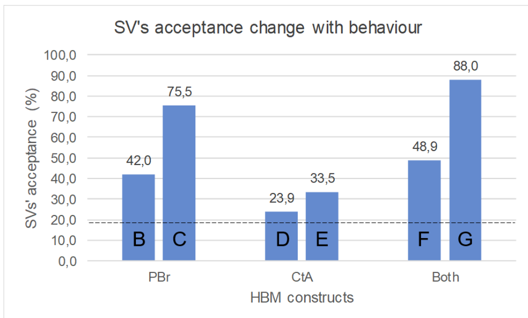

Figure 4: SVs' acceptance percentage for simulation scenarios B to G. The black dashed line marks the reference acceptance percentage, registered for scenario A.

Figure 5 shows the effects of the simulation scenarios on the total number of surgeries performed. With the exception of scenario D, where the number of surgeries was close to the one reached in scenario A, all the others presented a considerable increase in surgery delivery. In particular, scenarios E and F obtained similar results, with 687 and 690 surgeries performed – which would correspond to approximately 687 000 and 690 000 in reality. In scenario C, 720 surgical procedures were conducted (more 95 than in scenario A). Scenario G, as expected, reached an even more impressive improvement, with 110 more surgeries than in the baseline situation. Once again, these values would have to be converted to thousands in order to foresee their possible impact in reality.

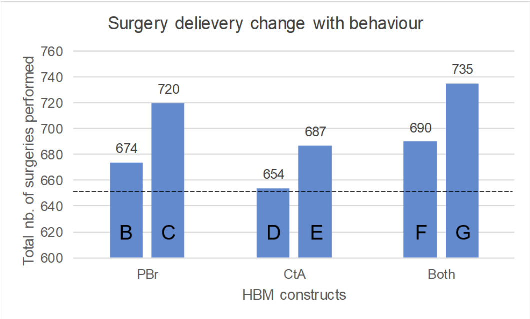

Figure 5: Total number of surgeries performed for simulation scenarios B to G. The black dashed line marks the reference number of surgeries, registered for scenario A.

When comparing scenarios B and C with the reference A, for example, it was observed an increase of 19 and 32 patients leaving the SWL, respectively. Such improvements, especially when converted to thousand units to predict real-life impact, show how behavioural changes regarding PBr have the potential to significantly decrease the SWL's length.

Therefore, and in order to achieve the improvements explored above, public health authorities should make efforts to refine the SVs' programme policies, especially aiming at minimizing the obstacles felt by patients to accept their transfer (and, consequently, at reducing their PBr, the construct with higher impact on SWL's numbers). Thus, offering more options of hospitals closer to patients' residential area, as well as more transportation and accommodation aids are examples of possible measures to reduce the percentage of transfer refusals. Besides PBr, CtA also proved to be an important contributing factor to patients' ItA. Strategies targeting this latter construct could also be effective – some examples may be: recommending family doctors (or the doctors who initially proposed surgery at the home hospital) to incentivize patients' acceptance of the SV, as well as promoting the general population's knowledge about the benefits of this programme.

Nevertheless, it is important to reinforce that the estimations presented in the above graphs were extracted from a simulation model which has a certain degree of error associated to it, as it was observed with the outputs obtained for scenario A that, although close to what was experienced in reality, showed, nonetheless, slight discrepancies. Several approximations were adopted when developing the model, which might have originated these differences. For example, the fact that, while SVs' acceptance, refusal, and cancellation statistics were based on national values, waiting times were extracted from a single hospital's database and may, therefore, not exactly represent the national scenario. Furthermore, as any model representation of a real system, some aspects of the SWL's functioning were simplified or not included (such as patients' reenrolment after ignoring an SV).

5. Conclusions

Despite the important results achieved in this study, it also presents some limitations. For instance, it would be beneficial to perform a similar analysis with a larger sample of patients currently enrolled on the SWL, which would allow researchers to monitor their real behaviour in case their are offered an SV. Moreover, the simulation model could be refined, for example, by including national-wide data regarding waiting times.

All things considered, and despite the identified shortcomes, this research contributed with sound tools to help policymakers developing effective behavioural policies and understanding the impact that patients' behavioural changes could have on the evolution of the SWL.

Following the strategy implemented in Portugal with SIGIC, of better distributing surgical demand among the already existing (public and private) supply capacity, this decrease in the SWL's length should also cause a meaningful reduction in patients' waiting times, which, ultimately, would result in more accessible and responsive healthcare services, as well as in a healthier population.

References

[1] *Manual de Gestão de Inscritos para Cirurgia*, vol. 1 - Princípios Gerais. UCGIC, 2011.

[2] *Relatório Anual - Acesso a Cuidados de Saúde nos Estabelecimentos do SNS e entidades.* Ministério da Saúde, 2019.

[3] A. Ferreira, C. Antunes, E. Pires, M. L. M. Viegas, and P. Gomes, *Estudo dos motivos de recusa de transferência - Inquérito Telefónico.* UCGIC, 2008.

[4] M. Conner and P. Norman, "Health behaviour: Current issues and challenges," 2017.

[5] M. Conner, "Health behaviors," *International Encyclopedia of the Social & Behavioral Sciences*, 12 2015.

[6] G. Royston, A. Dost, J. Townshend, and H. Turner, "Using system dynamics to help develop and implement policies and programmes in health care in england," *System Dynamics Review: The Journal of the System Dynamics Society*, vol. 15, no. 3, pp. 293–313, 1999.

[7] S. C. Brailsford, "Healthcare: Human behavior in simulation models," in *Behavioral operational research*, pp. 263–280, Springer, 2016.

[8] P.-J. Hsieh and H.-M. Lai, "Exploring peoples intentions to use the health passbook in self-management: An extension of the technology acceptance and health behavior theoretical perspectives in health literacy," *Technological Forecasting and Social Change*, vol. 161, p. 120328, 2020.

[9] A. P. G. Scheidegger, T. F. Pereira, M. L. M. de Oliveira, A. Banerjee, and J. A. B. Montevechi, "An introductory guide for hybrid simulation modelers on the primary simulation methods in industrial engineering identified through a systematic review of the literature," *Computers & Industrial Engineering*, vol. 124, pp. 474–492, 2018.

[10] S. C. Brailsford, "Modeling human behavior-an (id) entity crisis?," in *Proceedings of the Winter Simulation Conference 2014*, pp. 1539–1548, IEEE, 2014.

[11] M. A. Lopes, Á. S. Almeida, and B. Almada-Lobo, "Forecasting the medical workforce: a stochastic agent-based simulation approach," *Health care management science*, vol. 21, no. 1, pp. 52–75, 2018.

[12] S. Brailsford and B. Schmidt, "Towards incorporating human behaviour in models of health care systems: An approach using discrete event simulation," *European journal of operational research*, vol. 150, no. 1, pp. 19–31, 2003.

[13] S. C. Brailsford, P. R. Harper, and J. Sykes, "Incorporating human behaviour in simulation models of screening for breast cancer," *European Journal of Operational Research*, vol. 219, no. 3, pp. 491–507, 2012.

[14] M. C. Wong, E. L. Wong, J. Huang, A. W. Cheung, K. Law, M. K. Chong, R. W. Ng, C. K. Lai, S. S. Boon, J. T. Lau, Z. Chen, and P. K. Chan, "Acceptance of the covid-19 vaccine based on the health belief model: A population-based survey in hong kong," *Vaccine*, vol. 39, no. 7, pp. 1148–1156, 2021.

[15] M. W. Watkins, "Exploratory factor analysis: A guide to best practice," *Journal of Black Psychology*, vol. 44, no. 3, pp. 219–246, 2018.

[16] D. George and P. Mallery, *IBM SPSS statistics 26 step by step: A simple guide and reference.* Routledge, 2019.

[17] M. de Fátima Salgueiro, *Modelos com Equaçoes Estruturais.* Sociedade Portuguesa de Estatística, 2012.

[18] J. Marôco, *Análise de equações estruturais: Fundamentos teóricos, software & aplicações.* ReportNumber, Lda, 2010.

[19] E. Knekta, C. Runyon, and S. Eddy, "One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research," *CBE—Life Sciences Education*, vol. 18, no. 1, p. rm1, 2019.