Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

SIMULATION SOLUTION VALIDATION FOR AN INTEGRATED EMERGENCY POST

Martijn R.K. Mes
Martijn Koot

Faculty of Behavioural, Management and Social Sciences
University of Twente
Enschede, 7522 NB, THE NETHERLANDS

ABSTRACT

Simulation solution validation concerns the comparison between the expected and actual performance of a solution provided by a simulation model. Such a comparison might become challenging when not only the implementation of the solution changed the environment, but also the processes and data have changed. We illustrate this challenge using a case study at an Integrated Emergency Post (IEP), which is a collaboration between a general practitioners post and a hospital's emergency department to provide out-of-hours emergency care. After performing a simulation study, our solution has been implemented, after which data has been gathered for two years. We validated the solution by performing various comparisons, using simulated and realized performance, under the original and changed data and processes, and with and without the proposed solution. We propose a solution validation framework to structure these comparisons, and provide key insights regarding solution validation, using our case study at the IEP.

1 INTRODUCTION

During the last two decades, the organization of the out-of-hours emergency care has changed radically within the Netherlands. Originally, the primary out-of-hours care used to be organized in small groups of General Practitioners (GPs) with rotating out-of-hours shifts. The GPs are responsible for the delivery of primary care and should therefore operate as gatekeeper to the access of the more expensive secondary care offered by the hospital's emergency department (ED). However, patients could still decide for themselves to visit the GP or go directly to the ED. This type of organization resulted in an inefficient way of providing emergency care, which explains the introduction of the so-called Integrated Emergency Post (IEP). Although IEPs exists in many forms, they generally involve an integration between a GP post (cooperation of GPs) and an ED at the same location. An IEP would force self-referrals to contact the GP post first, which could reduce the workload experienced by the ED, improve resource utilization, and reduce patients' waiting times. These, and other potential benefits, have been argued by many authors, e.g., Kool et al. (2008). However, a quantitative analysis of the actual benefits resulting from an IEP implementation is hard to find. In this paper, we contribute to this gap by describing a simulation study conducted for the hospital Medisch Spectrum Twente (MST), The Netherlands.

MST started to build a new hospital in 2012, which created the possibility to integrate the ED and the GP post located on the same site. To investigate the feasibility and potential benefits of an IEP within the new building of MST, we performed a simulation study in 2014 (see the MSc graduation report of Koster, 2014). We applied a general and flexible discrete-event simulation model from Mes and Bruens (2012), which was originally developed for another hospital and IEP. After performing the simulation study, we concluded that the integration of the GP post and ED alone yields no positive effects. However, if the integration is associated with some other organizational changes, the average length of stay (sum of waiting and treatment times) could be reduced with 20% and 11% for the GP post and ED respectively. Based on this work, the hospital decided to implement the IEP, which became operational in January 2016. Since we

now have two years of data available of the operational phase of the IEP, we can compare the original results and recommendations of our simulation study with the results of the implemented solutions, i.e., perform solution validation.

Robinson (2014) states that solution validation is rarely carried out in practice or reported in the scientific literature, even though it is the only true test of the outcome of a simulation study. Typically, solution validation approaches assume that the implemented solution forms the only variable that has changed. This if often not realistic, as implementation of the solution might take considerable time during which many changes can occur. Hence, it is likely that solution validation is not simply a comparison between one recommended solution and one snapshot of the real system. Instead, many comparisons could be possible, depending on multiple real-world snapshots with changing data and processes.

Also for our case study at MST, many changes have taken place in the ED and GP post after implementation of the IEP. Besides changes that are a direct result of the IEP implementation, there are also changes related to data (e.g., patient arrivals, patient characteristics, staff characteristics) and processes (e.g., related to appointment planning, and decision rules for staff and room allocations). Our original simulation model does not govern all the required elements to realistically represent (i) the non-integrated situation after our simulation study during 2014-2015 and (ii) the integrated situation during 2016-2017. The conceptual model should be extended in order to gain reliable simulation results. All of these changes make solution validation challenging. Still, solution validation is important as it (i) provides insight into the reliability of the original simulation study, (ii) helps to improve the simulation model for reuse, and (iii) provides opportunities to improve today's emergency care even further.

The contribution of this research is threefold. First, we provide a quantitative comparison of the implementation of an IEP. Second, we provide insights into our validation activities over a period of more than six years, and describe the challenges we encountered. Third, we present a solution validation framework to structure the validation activities. The remainder of this paper is structured as follows. First, we present our case study in Section 2. Next, we present our solution validation framework in Section 3. In Section 4, we illustrate our validation framework using our case study, and provide quantitative insights into the effects of integrating the GP post and ED. We end with conclusions in Section 5.

2 CASE STUDY

In the following, we describe the hospital under study (Section 2.1), summarize our simulation study during the six-year period (Section 2.2), and present our simulation model (Section 2.3).

2.1 System Description

Even before the integration of the GP post and ED, both organizations were closely located to each other. In the non-integrated situation, patients are recommended to first make a telephonic appointment with the GP, which acts as first point of access to emergency care, but can still choose to go directly to the GP post or ED as a so-called self-referral. Patients that make an appointment either get telephonic advice, are visited by a GP at home, or are invited to the GP post for a physical consult. After treatment at the GP post, patients go home, are referred to the ED, or are send to the radiology department if the GP suspects that the patient has one or more bone fractures. With respect to the arrival at the ED, patients are either referred by a GP, referred externally (ambulance, police, other hospital), or just decide themselves to go to the ED as self-referrals. Once the patient arrives at the ED and its urgency level is determined by an ED nurse during the triage activity, the patient is treated in three different steps: anamnesis (discuss patient's symptoms and medical history); diagnostics (e.g., X-ray, CT scan, Ultrasound, lab research, ECG); and the actual treatment (surgery, internal medicine, orthopedics, pulmonary medicine, neurology, and gastrointestinal & liver). After treatment at the ED, patients go home, are admitted to the hospital, are transferred to another hospital, or are send to the mortuary.

For the integrated situation, both the GP post and ED are accessible via the IEP's main entrance. At this shared entrance, there are two rooms reserved for the triage of self-referrals from both the GP post and

ED. An additional entrance is used for the arrival of the ED's emergency patients who travel by ambulance (or trauma helicopter and by police). The GP post can refer patients for further treatment to the ED, while the ED can admit patients into the hospital. All other activities or patient flows happen within the GP post or ED separately. Note that not all treated patients actually visit the IEP physically; some patients only receive telephonic advice or are visited by a GP at home. Most of the activities corresponding to the out-of-hours emergency care remained the same for both the separated and the integrated organization. However, some differences between the separated and integrated emergency care organization can be identified:

1. The locations of the GP post and the ED are not strictly separated anymore. The IEP can be seen as one location, which shelters two different stakeholders. Patients can be sent from the GP post to the ED directly, without leaving the building as before.
2. The GP post is now fully responsible for the triage of all out-of-hours patient arrivals. Calling patients and self-referrals are first helped by the GP assistant during the telephonic or physical triage respectively.
3. The GP post and ED work together on the physical triage during the night. However, at other times, a patient going to the ED will still receive triage from an ED nurse, even though the patient already had triage at the GP post.

Despite the integration, the GP post and ED share only a limited number of rooms. Furthermore, all staff members are employed by the GP post or the ED individually, which indicates that the two organizations still operate relatively independently from each other. For a detailed description of the typical processes at the GP post and the ED, both in the integrated and non-integrated situation, we refer to Mes and Bruens (2012). In the remainder of this paper, we use the abbreviations NIP and IEP to denote the non-integrated situation (until January 2016) and the integrated emergency post respectively.

2.2 Simulation Study

During the period 2012-2018, we performed a simulation study at the ED of the hospital MST and the GP post located on the same site. Our simulation study consisted of three phases, as outlined in Figure 1. We now briefly describe each of these phases.

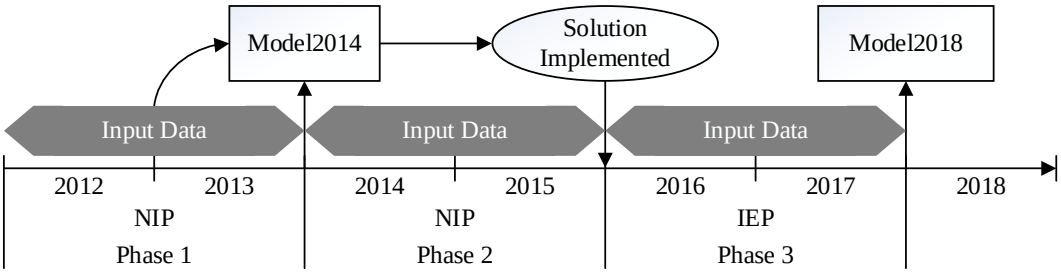

Figure 1: Time line of our simulation study.

In phase 1, data has been gathered during 2012-2013 where the ED and GP post where still separately organized. This data has been used in a simulation study in 2014 on the benefits of an IEP introduction. The conceptual model has been based on a snapshot of the real system at the beginning of 2014. The findings of this study have been reported to MST in 2014. Our simulation study from Phase 1 resulted in a recommendation to implement an IEP, but only when combined with several other interventions. It took two years before the IEP was fully implemented, we denote this period by phase 2. During this period, various changes took place besides implementing the IEP. We gathered data of the period 2014-2015 where those changes took place and where the ED and GP post were still organized separately. At the beginning

of 2016, MST implemented the IEP, although not with all of our recommended interventions. Afterwards, we collected two years of data of this IEP, we denote this period by phase 3. In 2018, we performed a new simulation study, updating the original conceptual model, updating the simulation model in accordance with the updated conceptual model, and performing experiments using the most recent data sources.

The main differences in data during the three simulation phases are the following:

1. In general, patient arrival rates decreased over the years (see, e.g., Figure 2 showing the decrease in patient arrivals during the weekends; the out-of-hours on weekdays show a similar decrease).
2. After the integration, relatively less patient are redirected to the ED by the GP post, which is unexpected given that a larger group of self-referrals arrives at the GP post.
3. The total number of urgent patients contacting the GP post increases over the years, while the number of ED patients decreases for all urgency classifications available.
4. The number of GP consults and GP visits at home decreased after the integration, while the number of GP assistant consults and nurse practitioner (NP) consults increased. More patients are also provided with advice, preventing them from visiting the GP post or ED unnecessarily.
5. The average duration of all activities at the GP post increases, except for the GP visits at home and the GP assistants' consults, for which the duration has decreased a bit.
6. The average total cycle time of both the GP post and the ED increased over the years, despite the reduction in patient arrivals (more complex care). The average GP post cycle time increased from 29 minutes in 2014-2015 to 34 minutes and 2016-2017. The ED cycle time increased from 138 minutes in 2014-2015 to 159 minutes in 2016-2017.

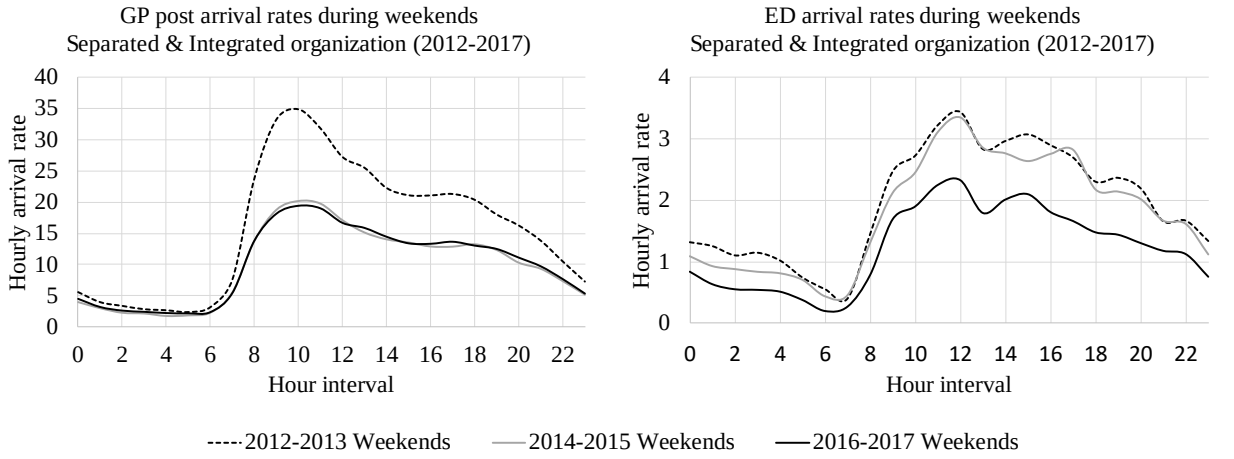

Figure 2: Patient arrival patterns and arrival rates during the weekends for each data set.

After the introduction of the IEP, also many process changes have taken place. Obviously, we have the foreseeable changes due to the implementation of the IEP, but also other changes have occurred, which we briefly describe below:

1. The ED implemented the same triage system used by the GP post as opposed to the different systems used in the past, resulting into slightly different urgency classifications for all ED patients.
2. A new appointment strategy is in use for the GP post's physical consults.
3. A new staff allocation strategy is in use for physical GP consults (e.g., NPs and GP assistants are allocated to a physical consult more frequently).
4. The ED's staff allocation strategy has been updated in accordance with the changes in patient characteristics (e.g., reduction in surgery/orthopedic hours and increase in dedicated care).

2.3 Simulation Model

Our simulation models for phase 1 and 3 are based on the flexible framework for discrete-event simulation modelling of (integrated) emergency departments developed by Mes and Bruens (2012). The model has been implemented in Technomatix Plant Simulation from Siemens, a simulation software package for integrated, graphical and object-oriented modelling and simulation. The model contains the following main components: (i) moving entities, which mainly represent the patients requesting emergency care; (ii) recourses, e.g., treatment rooms, staff, and diagnostic testing equipment, and (iii) processes, i.e., the decision logic implemented in order to represent the patient's logistic care pathways.

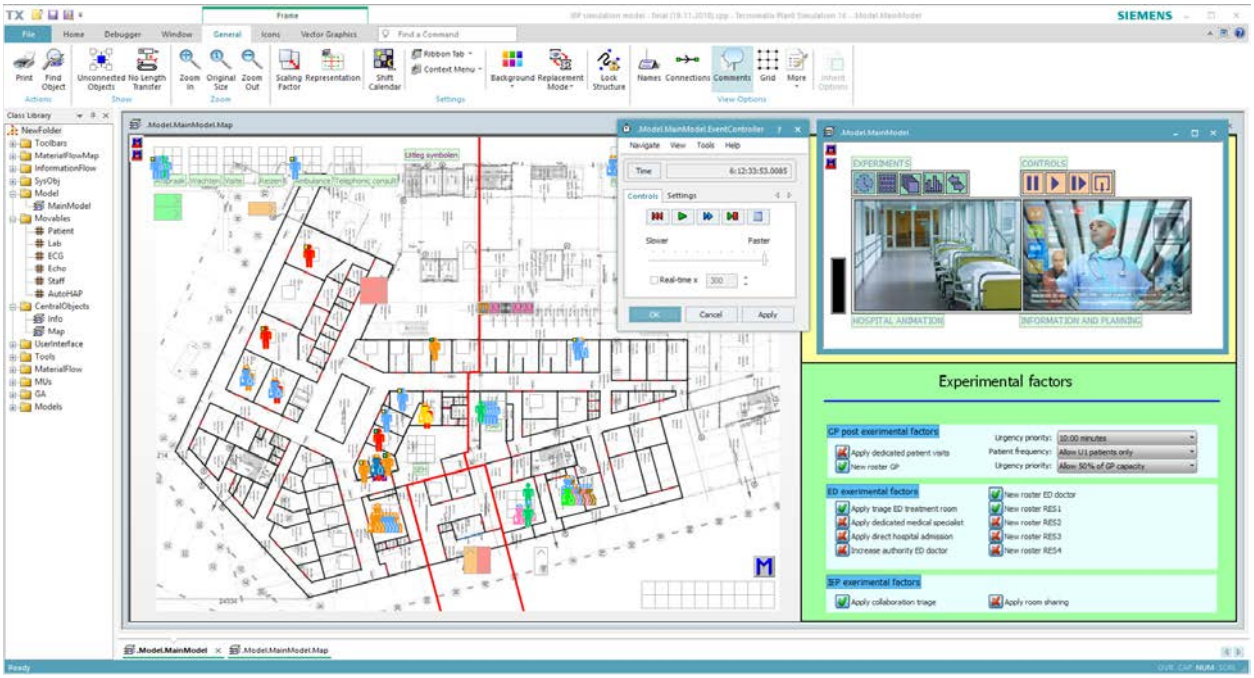

Figure 1: Visualization of the simulation model's graphical user interface (GP post to the right of the red line, ED to the left of the red line, and shared entrance between the red lines).

The entities move through the simulation model's objects according to predefined processes, while resources act on those entities. The resources in the model are represented by a map of the hospital denoting all treatment rooms, rosters for all employees, and a list of medical equipment. The processes prescribe which entities and resources are brought together into which treatment rooms. The simulation progresses depending on generated events, e.g., related to patient arrivals, activation of processes, task completions, staff schedules, and patient departures. The model contains many details with many dependencies, e.g., dependencies between time, number of patient arrivals, patients' urgencies, patients' care pathways, required diagnostics tests, etc. The main inputs are: patient arrivals, patient characteristics, care pathways of patients, distribution of treatment activities, staff rosters, resource allocation rules, planning rules for diagnostic tests, and appointment planning rules. The main outputs are: patients' length of stay, resource utilization rates and patients service levels (exceeding maximum waiting times depending on type). Figure 3 gives an impression of the simulation's graphical user interface, consisting of (i) a main frame with animation using the map of the hospital, (ii) a control panel that allows the user to activate, pause or exit the simulation experiments, and (iii) a dashboard and configuration panel that allow the user to experiment with the model.

3 SOLUTION VALIDATION

A wide variety of simulation modelling frameworks exists in literature (Robinson 2014). However, most frameworks in some form consider the following stages: (i) conceptual model (a description of the model that is to be developed), (ii) computer model (the simulation model programmed into a computer), (iii) solutions and/or understanding (the results obtained from experimentations), and (iv) real-world improvements (the implementation of the solutions). Between the different stages, we have to ensure that the model is designed or implemented correctly, using verification and validation. Verification is the process of evaluating if the conceptual model is correctly programmed into the simulation model, while validation includes the process of ensuring that the implemented simulation model and the underlying conceptual model are able to represent the real system and the simulation objectives adequately (Law 2015; Robinson 2014). Obviously, a simulation model is a simplification of reality, but it should represent the real world with sufficient accuracy in order to meet the simulation study's objectives (Robinson, 2014).

Verification and validation activities should result into a quantified level of agreement between the experimental data and the predictions made by the simulation model with sufficient accuracy (Thacker et al. 2004). However, Robinson (1997) states that "Verification and validation is far from straightforward and is often not performed as thoroughly as it might be". A large number of verification and validations methodologies is recognized in scientific literature, but no unique validation test exists that can easily be applied to determine the model's correctness (Sargent 2011). Graphical data comparisons and the usage of confidence intervals will help the validity activities (Kleijnen 1999), but no standardized format exists in literature so far. Some well-known methodologies for performing a simulation study, including verification and validation, are: Law's model to build valid and credible simulation models (Law 2015), Landry's validation model (Landry 1983), Sargent's model on simulation modelling, verification and validation (Sargent 2015), and Balci's life cycle for modelling and simulation (Balci 2012). In our study, we relied on the model from Robinson (2014), adapted from Landry et al. (1983), that consists of the following main validation techniques.

1. Conceptual model validation: determining if the content, assumptions and simplifications proposed are sufficiently accurate.
2. Data validation: determining if the data required is gathered, processed and applied sufficiently accurate.
3. White-box validation: determining that separated sub-modules of the computer model represent the real-world elements sufficiently accurate.
4. Black-box validation: determining that the overall model represents the real system sufficiently accurate.
5. Experimentation validation: determining that the experimental procedures provide sufficiently accurate results.
6. Solution validation: determining whether the performance of the recommended solution corresponds sufficiently accurate with the realized performance of the real system after implementing the solution.

Especially the activity of solution validation is rarely carried out in practice, whereas this is the only true test of the simulation study's outcome (Robinson 2014). This final activity is required to compare the expected and actual performances of the recommendations made, allowing the investigators to adapt the implementations made over time, but also to improve the simulation model for reuse. Reason for not investing in solution validation range from costs related to time, data gathering, and staff availability, to the fact that many studies involve small-scaled simulation models, which are less relevant to decision makers to validate (Robinson and Brooks 2009).

An abstract and simplistic view on solution validation is shown in the left side of Figure 4, where solution validation is performed similarly to black-box validation: when we use the same input data for our

simulation model as was used in the real system, then the outputs (performance) of the simulation and real system should be comparable. For solution validation, it would mean that we perform a simulation on the recommended solution and compare it with the realized performance once this solution has been implemented in the real system and sufficient data has been collected during a period in which the solution was operational. This approach gives a good indication of the quality of the original simulation study, in which we could have taken into account possible changes that might occur in data and processes (e.g., by the application of a sensitivity analysis). However, it gives less insight into the quality of the simulation model itself, as well as into the effects of the proposed solution compared to alternatives. Hence, for solution validation we might want to take into account the fact that systems change, not only with respect to the recommended solution, but also in terms of other processes and data. Although this might hold between each of the stages from Figure 4, it particularly plays a role between the generated solutions from the simulation study and the implementation of these solutions in the real system, as implementation of the solution might take considerable time. During this time, the conceptual model on which the simulation is based might become outdated, and we might need to update our model. Therefore, we see solution validation as a potential multi-dimensional comparison, distinguishing the following components:

1. Data modifications: a collection of different data sets collected over time.
2. Model modifications: a collection of process modifications resulting in updated conceptual models and corresponding implemented simulation models over time.
3. Solutions: a collection of recommended solutions.

We can perform experiments using each combination of data set, model, and solution. For our case study, this would mean we can choose between 3 data sets (corresponding with the 3 phases), 2 models (those from 2014 and 2018), and 2 solutions (NIP and IEP), resulting in 12 experiments. Theoretically, we could compare each of these experiments with the 3 snapshots of the real system (corresponding with the 3 phases), resulting in 36 comparisons. However, for solution validation comparing with a given snapshot of the real system, it makes sense to consider the same solution in the model as was implemented in the real system. Because the solution is now determined by the real system we compare with, we now have a two-dimensional solution validation framework as shown on the right side of Figure 4.

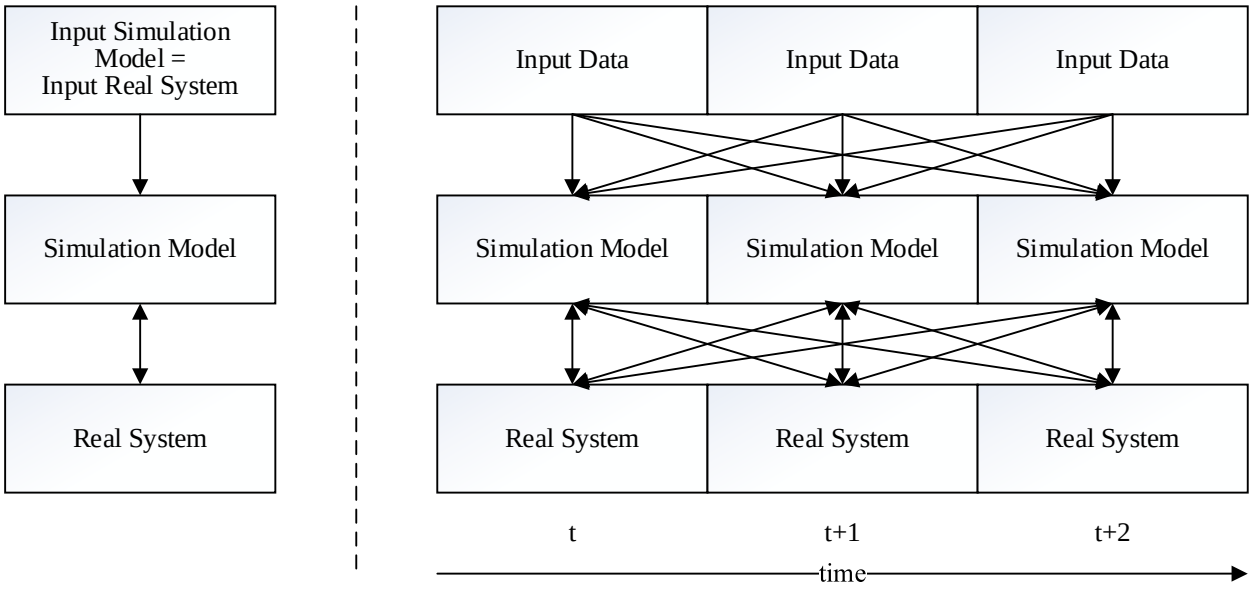

Figure 4: Model and solution validation: static (left) and dynamic (right).

This framework makes it possible to evaluate and compare the simulation results for all model descriptions and data sets used. In this way, insights are gained by comparing the impacts of changed processes, input variables, and model assumptions. In the next section, we illustrate these comparisons using our case study.

4 SOLUTION VALIDATION FOR OUR CASE STUDY

When applying the solution validation framework from Section 3 to our case study, we have to deal with three data sets, two simulation models, and two possible solutions, as also shown in Figure 1 and Table 1. As mentioned before, our solution validation framework might result in large number of comparisons, which are not necessarily all relevant. For our case study, we distinguish between 8 comparisons, grouped into 6 comparison categories, as shown in Table 1.

Table 1: Model and solution validation comparisons.

Nr	Input Data	Real System	Model	Solution	Purpose
1	2012-2013	2012-2013	Model2014	NIP	Model validation phase 1.
2	2014-2015	2014-2015	Model2018	NIP	Model validation phase 2 and 3 for the NIP and IEP respectively.
	2016-2017	2016-2017	Model2018	IEP	
3	2014-2015	2014-2015	Model2014	NIP	Solution validation: assess the quality of the original model using the same inputs as the real system.
	2016-2017	2016-2017	Model2014	IEP	
4	2012-2013	2014-2015	Model2018	NIP	Solution validation: assess the impact of the changed data.
	2012-2013	2016-2017	Model2018	IEP	
5	2012-2013	2014-2015	Model2014	NIP	Simulation study evaluation.
	2012-2013	2016-2017	Model2014	IEP	
6	2016-2017	2016-2017	Model2018	NIP	Solution validation: assess the benefits of the IEP implementation.

The first two comparisons correspond with model validation. The third comparison corresponds with the definition of solution validation from Robinson (2014), in which the final model of the proposed solution is compared with today's real system (left side of Figure 4). The alternative is to use the original data within the new model. The combination of both comparisons provides insights into the effects of changed data and processes. The fifth comparison provides insight into the quality of the original simulation study as a whole, in which changes in data and processes could have been taken into account. In the final comparison we study the differences between the current situation of the IEP and a simulated NIP using the current processes and data. This comparison gives an impression of the benefits of the current IEP implementation, while all other independent variables are held constant (ceteris paribus). In the following, we briefly illustrate the comparisons of Table 1. A summary of the results of all the comparisons can be found in Table 2; we use the superscripts to refer to the corresponding comparisons. In our simulation study, we performed the comparisons using various KPIs, distinguishing between patient types and waiting times before various processes, considering both averages and distribution of results. However, for the sake of brevity, we focus in this paper on the average Length of Stay (LoS) of patients, given by the sum of all waiting and treatment times. For all experiments, we used the batch means method with a batch size of one week, a warm-up period of one week, and a simulation run length of 100 weeks. To achieve a relative error of at most 5% for the LoS, a run length of 37 weeks would have been sufficient. For the comparisons, we used a significance level of 5% for our t-tests.

4.1 Comparison 1: Model Validation Phase 1

To contribute positively to the validity of the model, we collected high-quality information and historical data on the system from the years 2012 and 2013. For processes with limited historical data, we used expert

opinion. We used monthly meetings to discuss our assumptions and establish credibility with the stakeholders. Next, we used white-box validation for specific components of our simulation model. An important KPI is the waiting time of patients. We compared various waiting times resulting from the simulation model with those in practice. For example, using the waiting time before first contact (triage), which is 5.10 and 4.60 minutes in reality and the simulation respectively.

Finally, we used black-box validation to compare the overall simulation model with reality. For the ED, the average difference in LoS between reality (120.4) and the simulation (118.8) is 1.3%. When zooming in on the differences in LoS per urgency level (there is a high correlation between urgency level and care pathway), the differences range between 1% and 3%, except for the lowest urgency level (denoted by blue level). The blue patients are patients that, according to professionals, do not really belong to the ED. In theory, and hence in our simulation model, these patients are treated after all the other patients are helped. However, in practice these patients are helped in between the other patients. We decided not to adapt the simulation model, as the blue patients only account for 0.32% of the total ED population, concerning less than one arrival per week. We performed a similar analysis for the GP post. Here we found larger and significant differences; the average LoS at the GP post is 15.7% lower in the simulation model (27.3) compared to reality (32.4). After discussing these results with stakeholders from the GP post, we found that this discrepancy was caused by (i) the fact that patients could arrive before their appointment time (voluntarily waiting time) and (ii) delays in registration of leaving patients (treatment is officially registered as finished after the patient left the GP post). Experts approved that the difference in LoS at the GP post could be explained by the before mentioned delays. Hence, we conclude that the simulation model is valid for the purpose of representing the NIP organization in 2012-2013.

Table 2: Results of all model and solution validation experiments (time in minutes). Superscripts refer to model and solution comparisons (Table 1) where the corresponding result is used).

	GP post LoS			ED LoS		
	2012-2013	2014-2015	2016-2017	2012-2013	2014-2015	2016-2017
Reality	32.4[1]	28.7[2,3,4,5]	33.8[2,3,4,5,6]	120.4[1]	138.2[2,3,4,5]	159.1[2,3,4,5,6]
Model1-NIP	27.3[1,5]	60.7[3]		118.8[1,5]	86.3[3]	
Model1-IEP	41.6[5]		66.2[3]	154.5[5]		101.2[3]
Model2-NIP	29.7[4]	30.3[2]	28.8[6]	162.8[4]	140.1[2]	178.1[6]
Model2-IEP	37.3[4]		31.4[2]	251.3[4]		160.9[2]

4.2 Comparison 2: Model Validation Phase 2 and 3

To validate the model for phase 2 and 3, we used similar procedures as for phase 1. As input we use the data from 2014-2015 for the NIP and from 2016-2017 for the IEP. We first used white-box validation and concluded that the simulation model works properly without any bugs. Next, we used black-box validation. Here we compare the real systems, for both the separated (2014-2015) and the integrated (2016-2017) situations, with the outcomes of the simulation model using the same inputs as the two real systems. Only the simulation model's input variables were changed for both the separated and integrated organization, the underlying decision logic remained the same for the two data sets simulated. Therefore, the results are validated twice, which will support the simulation model's reliability. Again, we validate the models based on the LoS for both the GP post and the ED.

Separated organization (2014-2015). The differences between the actual and simulated average LoS are too small to conclude any significant difference for both the GP post and the ED. The LoS variances simulated also look similar to the values obtained from the patient records. We also compared the patients' LoS at the ED and GP post for each of the 6 urgency levels. The simulated ED's LoS matches the actual LoS quite well (on average 1% difference, largest difference of 6%). The GP post's results are not perfectly

aligned (on average 6% difference). The simulation model treats high urgent patients too fast, because they always receive priority over low urgent patients, without considering any other attributes that are taken into account in reality (e.g., entrance complaints). Because the effect of this difference is limited, our model is assessed as being valid for the purpose of this study.

Integrated organization (2016-2017). Again, the differences between the actual and simulated average LoS are too small to conclude any significant difference for both the GP post and the ED. The GP post's results are not perfectly aligned across the five different urgency classifications. Again, the differences can be explained by the appointment strategy implemented. The GP post LoS differences are significant, but acceptable for the purpose at hand. The simulated ED's LoS values match the actual LoS almost perfectly, the results only differ ±2% for all urgency classifications.

We conclude that the simulation model for Phase 2 and 3 is valid for the solution validation purposes at hand. The model properly represents both the separated (2014-2015) and integrated (2016-2017) emergency care organization.

4.3 Comparison 3: Solution Validation with Original Model and Recent Data

To assess the quality of the original simulation model, we compare its outcomes with reality using recent inputs. This comparison represents the black-box validation approach for solution validation as proposed by Robinson (2014). We first validate the NIP, by comparing the results during 2014-2015 with those of the original simulation model using the data of 2014-2015. Next, we validate the IEP, by comparing the results during 2016-2017 with those of the original simulation model using the data of 2016-2017.

Separated organization (2014-2015). The simulation model reveals significant different results for both the GP post LoS and the ED LoS: 112% higher for the GP post LoS (60.7 instead of 28.7) and 38% lower for the ED LoS (86.3 instead of 138.2).

Integrated organization (2016-2017). Again the differences in LoS are significant as seen in Table 2. The GP post LoS is clearly overestimated by the original simulation model (96%). The simulated ED LoS is underestimated in comparison with today's actual ED LoS (-36%).

The differences regarding the expected effects of the implemented solution are also significant. The simulation models predicted a 9% increase in LoS at the GP (60.7 to 66.2) and a 17% increase in LoS at the ED (86.3 to 101.2) due to (i) the implementation of an IEP and (ii) the changing data between phase 2 and phase 3. However, in reality, these differences where 18% (28.7 to 33.8) and 15% (138.2 to 159.1) respectively. Given the differences in LoS values between the simulation and the real system, both for the NIP and IEP, we conclude that the underlying conceptual model changed too much to reliably simulate the GP post and ED activities. Hence, the original model is not valid anymore.

4.4 Comparison 4: Solution Validation with New Simulation Model and Original Data

Although we concluded in Section 4.3 that the original simulation model does not accurately describe the current processes, it is still interesting to find out if the differences can be contributed to the changing processes only or whether changed data also has a major impact. Furthermore, the results do not show whether the effects can be contributed to the implementation of the IEP or to the changing data. Therefore, the new simulation model will be used to simulate both the separated and the integrated situation, based on the patient arrivals, processes and resource allocations during 2012-2013. The simulation model results are compared to the patient records of 2014-2015 and 2016-2017 for the NIP and IEP situation respectively.

Separated organization (2012-2013). Again, there are major significant differences (see Table 2), although they are now considerably smaller than before. Hence, changed processes affected our simulation more than changed data. Our model overestimates the GP post and ED LoS by 4% and 18% respectively for the NIP, which is caused by the decreasing number of patients and changed staff allocations.

Integrated organization (2012-2013). Over the years, the actual number of patient arrivals for both the GP post and the ED decreased (Figure 2). If today's patient arrivals would be equal to the number of

patient arrivals in 2012-2013, a significant increase in the patients' LoS would be obtained for both the GP post and the ED: our model overestimates the GP post and ED LoS by 10% and 58% respectively.

The differences regarding the expected effects of the implemented solution are also significant. The simulation models predicted a 26% increase in LoS at the GP (29.7 to 37.3) and a 54% increase in LoS at the ED (162.8 to 251.3) due to (i) the implementation of an IEP and (ii) the changing data between each of the three simulation phases. However, in reality, these differences where 18% (28.7 to 33.8) and 15% (138.2 to 159.1) respectively.

4.5 Comparison 5: Simulations Study Evaluation

This comparison might be the most intuitive one, as we compare the original recommendations with today's practices. Our original study underestimated the LoS at the NIP for both the GP post and ED with 5% and 14% respectively. For the IEP, the original study overestimated the LoS at the GP post (23%) but estimated the LoS at the ED relatively well (-3%). These differences are caused by changes in data (e.g., patient arrivals and characteristics) and changes in processes (e.g., staff allocations). As discussed in Section 4.3 and 4.4, these changes have different effects: the decrease in patient arrivals results in overestimation, more complicated care results in underestimation, and the effect of changes in processes depend on the data set used in the simulation as new processes are often the result of changes in patient arrivals and characteristic. Given the opposite effects of these changes, a comparison between the results and recommendations of the original simulation study and the realized results do not make any sense. This illustrates the importance of taking estimates of future changes in data and processes into account within a simulation study, e.g., by means of a scenario and sensitivity analysis.

4.6 Comparison 6: Solution Validation Using Recent Data

The previous sections revealed that the patient arrivals, processes and resource allocations changed significantly for both the GP post and the ED, such that our original estimated effects of an IEP introduction are not valid anymore. As a result, we are still not able to draw conclusions on the effects of the integration. To achieve this, we need to compare the current situation with the hypothetical situation in which the IEP would not have been implemented, i.e., simulate the NIP with the new model using the inputs of 2016-2017, see Table 2. Clearly, the LoS at the ED decreased as a result of the IEP introduction, but at the expense of an increase in LoS at the GP post. It can be concluded that the ED benefits most from the integration. The transfer of self-referrals to the GP post allowed the emergency physicians to decrease their workload by approximately 10% on average, which decreases the average ED LoS by 11% (178.1 to 159.1). The GP post however has to take care of more unexpected patient arrivals, which increases the average GP post LoS by 11% (31.4 to 33.8). However, these changes are acceptable for two reasons: (i) the GP post LoS is absolutely smaller than the ED LoS and (ii) the GP post's appointment strategy allows the GP assistants to schedule arriving patients, reducing the negative impact that self-referrals have on the GP post LoS.

Given that the new simulation model has been validated for both the NIP and IEP situation (see Section 4.2), we decided also to study some interventions to further improve the real system. The GP post LoS can be reduced with 9.7 minutes (-29%) by increasing the time slots for scheduling patients and creating new rosters by shifting the working times of some staff members. The ED LoS can be reduced with 52.6 minutes (-33%), by implementing new staff rosters for the emergency physicians and surgery/orthopedic residents, changing the way patients are admitted to the hospital, changing the authority of the emergency physicians, performing the physical triage in the ED treatment rooms, and sharing the triage results between the GP post and the ED.

5. CONCLUSIONS

In this paper, we described a simulation study of an Integrated Emergency Post (IEP) at a hospital within The Netherlands. After performing an initial simulation study for this IEP, our solution has been implemented, after which data has been gathered for two years. With this data, we compared the original

results and recommendations of our simulation study with the results of the implemented solutions, i.e., performed solution validation. We showed that solution validation becomes challenging when not only the implementation of the solution changed the environment, but also the processes and data have changed considerably. In this case, the typical black-box solution validation approach might not provide useful insights. Therefore, we performed various other comparisons, using simulated and realized performance, under the original and changed data and processes, and with and without the proposed solution. We proposed a solution validation framework to structure these comparisons, and showed that this enabled us to identify the factors that influenced the actual IEP performances the most. More specifically, it enabled us to (i) judge the quality of the performed simulation studies, (ii) judge the quality of the simulation models themselves and improve these models for reuse, and (iii) provide insight in past and current bottlenecks, providing opportunities for further improvement of the real system.

REFERENCES

Balci, O. 2012. "A Life Cycle for Modeling and Simulation". *Simulation* 88(7):870–883.

Kleijnen, J. P. C. 1999. "Validation of Models: Statistical Techniques and Data Availability". In *Proceedings of the 1999 Winter Simulation Conference*, edited by P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 647-654. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kool, R., D. Homberg, and H. Kamphuis. 2008. "Towards Integration of General Practitioner Posts and Accident and Emergency Departments: A Case Study of Two Integrated Emergency Posts in the Netherlands". *BMS Health Services Research* 8(1): 225.

Koster, D. 2014. "Simulating the Effect of an Integrated Emergency Post in Enschede - Verification and Application of a General and Flexible Discrete-Event Simulation Model". Graduation Report, Master Industrial Engineering & Management, University of Twente, Enschede, The Netherlands.

Landry, M., J-L. Malouin, and M. Oral. 1983. "Model Validation in Operations Research". *European Journal of Operational Research* 14(3): 207-220.

Law, A. M. 2015. *Simulation Modeling & Analysis*. 5th ed. New York: McGraw-Hill, Inc.

Mes, M. and M. Bruens. 2012. "A Generalized Simulation Model of an Integrated Emergency Post". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1–11. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Robinson, S. 1997. "Simulation Verification, Validation and Confidence: A Tutorial". *Transactions of the Society for Computer Simulation International* 16(2): 63-69.

Robinson, S. (2014). *Simulation: The Practice of Model Development and Use*. 2nd ed. Palgrave Macmillan.

Robinson, S. and R. J. Brooks. 2009. "Independent Verification and Validation of an Industrial Simulation Model". *Simulation* 86(7): 405-416.

Sargent, R. G. 2011. "Verification and Validation of Simulation Models". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, 183-198. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sargent, R. G. 2015. "An Introductory Tutorial on Verification and Validation of Simulation Models". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 1729-1740. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Thacker, B. H., S. W. Doebling, F. M. Hemez, M. C. Anderson, J. E. Pepin, and E. A. Rodriguez. 2004. "Concepts of Model Verification and Validation". Technical Report. Los Alamos National Lab., Los Alamos, NM (US).

AUTHOR BIOGRAPHIES

MARTIJN MES is Associate Professor within the department of Industrial Engineering and Business Information Systems at the University of Twente, The Netherlands. He holds a MSc in Applied Mathematics (2002) and a PhD in Industrial Engineering and Management at the University of Twente (2008). His research interests are healthcare logistics, transportation, multi-agent systems, stochastic optimization, discrete event simulation, and simulation optimization. His email address is m.r.k.mes@utwente.nl.

MARTIJN KOOT is PhD Candidate within the department of Industrial Engineering and Business Information Systems at the University of Twente, The Netherlands. He holds both a BSc and MSc in Industrial Engineering & Management at the University of Twente, received in 2016 and 2018 respectively. His research interests are healthcare logistics, operations research, data science, enterprise architectures, and artificial intelligence in logistics. His email address is m.koot@utwente.nl.