Authors:

Rene de la Fuente[1], Ricardo Fuentes[1], Jorge Munoz-Gama[2], Arnoldo Riquelme[3], Fernando Altermatt[1], Juan Pedemonte[1], Marcia Corvetto[1], Marcos Sepúlveda[2].

[1] Department of Anaesthesiology. Faculty of Medicine. Pontificia Universidad Católica de Chile.
[2] Department of Computer Science. Faculty of Engineering. Pontificia Universidad Católica de Chile.
[3] Department of Gastroenterology, Centre for Medical Education and Health Sciences, Department of Health Sciences, Faculty of Medicine, Pontificia Universidad Católica de Chile.

Corresponding author:

Ricardo Fuentes
E-mail address: rfuente@med.puc.cl
Telephone Number: +56223543270
Fax Number: +56226327620

Title

Control-Flow Analysis of Procedural Skills Competencies in Medical Training through Process Mining

Short Title

Control-Flow Analysis of Procedural Skills through Process Mining

Abstract

Background: Procedural skills are key to good clinical results, and training in them involves a significant amount of resources. Control-flow analysis (i.e., the order in which a process is performed) can provide new information for those who train and plan procedural training.

Aims: This study outlines the steps required for control-flow analysis using process mining techniques in training in an ultrasound-guided internal jugular central venous catheter placement using a simulation.

Method: A reference process model was defined through a Delphi study, and execution data (event logs) were collected from video recordings from pre-training (PRE), post-training (POST) and expert (EXP) procedure executions. The analysis was performed to outline differences between the model and executions. We analysed *Rework (activity repetition)*, *Alignment-Based Fitness (conformance with the ideal model)* and *Trace Alignment Analysis* (visual ordering pattern similarities).

Results: Expert executions do not present repetition of activities (Rework). The POST rework is lower than the PRE, concentrated in the steps of the venous puncture and guidewire placement. The adjustment to the ideal model measure as Alignment-Based Fitness, expressed as a median (25-75 percentile) of PRE 0.74 (0.68-0.78) is less than POST 0.82 (0.76-0.86) and EXP 0.87 (0.82-0.87). There are no significant differences between POST and EXP. The graphic analysis of alignment and executions shows a progressive increase in order from PRE to EXP executions.

Conclusion: Process mining analysis is able to pinpoint more difficult steps, assess the concordance between reference mode and executions, and identify control-flow patterns in procedural training courses.

Introduction

The development of procedural skills is an essential component of the process of training medical doctors from surgical and non-surgical specialties. Superior technical performance positively affects patient outcomes [1]. Conversely, the presence of technical deficiencies is the most important factor associated with operator errors in hospitalized patients [2]. Simulation as a procedural teaching tool has proven to be effective to enable doctors to reach an adequate level, prior to contact with patients [3] and reduce complications associated with procedures [4],[5]. Its trade-offs are the

high costs associated with the material and human resources necessary for its implementation [6]. Therefore, it is important to have a wide range of analysis techniques that allow maximum knowledge to be mined from each session, or the learning process to be improved by reducing the number and eventually the cost of the sessions.
Usually, the procedural assessment tools used to analyse the executions in a simulation context have the objective of making individual evaluations of a formative or summative nature. Thus, checklists, for example, allow the execution of the steps of a procedure to be analysed dichotomically, without assessing quality. Global Rating Scales (GRS) carry out a global qualitative analysis with some inter-reliability deficiencies [7],[8]. These experiences have been used in various procedures, such as the installation of central venous accesses, lumbar punctures, cardiovascular resuscitations, paracentesis, etc. [9-13]. However, there are no tools that analyse the procedures from the perspective of control-flow, i.e., the order in which the activities of the process are executed. The use of process models - a simplified pattern of a medical procedure in a formal or semi-formal representation [14],[15] - allows the direct analysis of control-flow, but also indirectly benefits the learning process, as the process is clearly set out and broken down into a series of ordered activities; input that is required before training. [16],[17].

Process Mining is an emerging discipline which bridges the gap between traditional model-based process analysis and data-centric analysis, delivering tools to analyse and improve processes from factual data obtained from real instances of process execution [18]. In the healthcare domain, process mining has been used in different case studies e.g. medical treatment processes, organisational processes and non-elective and elective care with promising results [19],[20]. In the educational arena, process mining has been used to identify behaviours in online learning environments, feedback delivery, computer-based assessment, among others [21]. To the best of our knowledge, there are no studies describing the use of process mining as a tool for the analysis of educational experiences in healthcare disciplines, such as the learning of procedural skills.

Our hypothesis is that, by using process mining tools, we can obtain information about patterns of execution of a procedure and how it evolves during a training program, in addition to making a diagnosis based on comparisons with an ideal execution model. We will use process mining as a model for analysis in the training of an invasive medical procedure: ultrasound-guided internal jugular central venous catheter (UGIJCVC) placement; analysing it from a control-flow perspective before and after the course, as well as comparing trainees' and experts' performances in the procedure.

Context

For analysis, video extracts from a previous study [22] were used, oriented to the training of residents from different specialty programs (anaesthesiology, intensive medicine, cardiology, nephrology and emergency medicine). This study included a training program which was carried out in four stages:

1. Online Instruction: three web-based recorded lectures were available, each one with obligatory and complementary readings. At the end of this stage, a written evaluation and video recording of the procedure execution (PRE) were carried out.
2. Demonstration session: for the entire group of residents, an expert demonstration of the entire process of installing an internal jugular central venous access in Blue Phantom CVC torso [Blue Phantom®, Redmond, WA] was performed. In addition, there were 4 stations of deliberate practice: a) a station for the preparation of the ultrasonography equipment, patient and work tools, b) a station for the handling of the ultrasonography equipment, c) a station for venous puncture with ultrasonographic guidance, d) a station for catheter installation and fixation.
3. Deliberate Practice: Residents had to complete four sessions of deliberate practice accompanied by an instructor who supervised and gave immediate feedback.
4. After completing the course, each participant made a second video recording of the process of UGIJCVC (POST).
5. In addition, videos of the procedure were obtained by anaesthesiologists with at least 5 years of clinical experience at UGYCVC (EXP).

Methods

Prior to approval by the institutional ethics committee (Id: 16-194), videos obtained from previous training sessions were used and the proposed method was executed in three stages: 1) Model Sources and Modelling, 2) Data Sources and Collection, y 3) Control-Flow Analysis with Process Mining.

Model Sources and Modelling

The first stage of our method was the definition of a generic process model [14,15] of central venous catheter access with ultrasonography. As a modelling notation, we used Business Process Model Notation (BPMN) [23], which stands as the *de facto* standard most widely used due to its versatility, expressiveness and ease of interpretation for users in non-computing areas [24],[25]. This first model was submitted to a consensus procedure using Delphi methodology [26],[27], through an online survey. National experts were asked to score the appropriateness of including the activities proposed in the model using a 5 point Likert scale, and to propose new activities. The criterion for maintaining an activity was a consensus of 75% or more, and an average of the absolute percentage changes in all responses of less than 15% [28] in successive rounds was defined as the stopping criterion for ending the rounds, planning a maximum of three rounds if that criterion was not met.

Data Sources and Collection

The second stage was the generation of *event logs*, i.e., structured data on the execution of the process [29] [18]. Event logs were obtained from a secondary analysis of the video recordings from a previous study [22]. To obtain these *event logs*, video recordings of the execution to be analysed were used, from which the different activities observed were manually labelled and their start and end times noted. We considered 34 videos of the PRE stage, 30 of the POST stage of the training course, and 9 videos were obtained by anaesthesiologists with at least 5 years of clinical experience at UGIJCVC. Six video recordings (5 PRE and 1 POST) were not used because they were incomplete. The videos were tagged based on an observer-based approach [30], using the VcodeVdata tool [31], software developed to make annotations in video recordings. Two observers, blinded to the characteristics of the executor and moment of execution (PRE or POST), did the labelling of the activities in the videos. In an arbitrary way, it was decided that a set of 26 videos, chosen in a random manner, would be analysed by two observers. To evaluate agreement between observers, we used Normalized Levenshtein Distance Metric (NLD) [32-34], a metric that quantifies the similarity between strings, based on the number of eliminations, insertions, and substitutions necessary for two compared chains to be equal, which are normalised by the maximum possible number of changes in both chains and are expressed as differences of 1; considering this value as equal traces. An average value of NLD equal to or greater than 0.85 was defined as sufficient agreement to proceed to divide the rest of the videos between the two observers.

Process Mining Analysis

The third stage used the reference process model and event logs to analyse the control-flow of the executions using process mining. This paper proposes the application of 3 state-of-the-art analysis techniques appropriate to the proposed scenario in medical education.

1. Rework Analysis: For each activity, the Rework Metric was defined as *RM= af/nc*, where *af* represents the absolute frequency of the activity, and *nc* the number of cases where it appears. The metric allows the activities involved in the rework to be quantified and pinpointed, i.e., activities that are performed several times [18]. The metric was calculated for each of the individual runs, as well as for each of the groups as a whole, using the average values, standard deviation and range for PRE, POST, and EXP.

2. Conformance Analysis: In process mining, the techniques of conformance checking allow the degree to which a process execution conforms to the process model to be quantified [35]. State-of-the-art conformance checking techniques "align" each execution with the process model, and define metrics to numerically determine the adhesion to the model [36]. The most widespread metric, and the one used in this article is alignment-based fitness, defined as AF = co / cw, where co is the computed optimal alignment cost and cw the worst-case alignment cost computed for that model and execution [36]. A standard cost function is used as a cost function, where all the incorrect executions of activities have the same weight in the metric. For more details about the algorithm, we refer the reader to [36]. Alignment-based fitness was applied

to each of the study groups: PRE, POST, and EXP, values presented as median (25-75 percentile) [rank].

3. Trace Alignment Analysis: Finally, the third process mining technique applied is Trace Alignment Analysis. In this analysis, each execution of the procedure represents a trace. The algorithm described in Bose et col [37] allows a group of traces to be aligned, computationally and visually, highlighting the parts that are common among the different executions and highlighting the differences. This allows the similarity in control-flow of the execution of the procedure of different groups to be qualitatively evaluated. The algorithm allows the execution of a case to be analysed individually in comparison with others. The comparison can be evaluated with a global granularity (e.g., the procedure from beginning to end), or by analysing the similarity at specific stages of the procedure. Unlike the classic algorithm of Needleman and Wunsch [38] designed to optimally align two amino acid sequences, the trace alignment algorithm [37] adopts a heuristic strategy to align several process executions in reasonable computational time. In particular, the algorithm is based on the progressive alignment approach, where the idea is to iteratively construct a succession of pairwise alignments. Alignment is allowed between a pair of traces, a trace and an alignment and between alignments. The selection of traces for alignment at each iteration is based on their similarity. Traces that are most similar to each other are aligned first. Once similar traces have been aligned, the resulting clusters of traces are aligned against each other. The implementation of the algorithm used in this work corresponds to the tool ProM [39]. For more details about the algorithm, we refer the reader to Bose et al [37]. Trace alignment was applied to each of the study groups: PRE, POST, and EXP.

[*Footnote1] The concept of "alignment" used in "Alignment-Based Fitness" and "Trace Alignment" is totally different. While the first one refers to the alignment between models and executions, the second uses a notion of alignment closer to the one used in bioinformatics [36] [37].

Statistical Analysis

Statistical analysis is performed with the Mann-Whitney test for comparisons between training groups and Kruskal-Wallis for multiple comparisons when these include expert executions. A value of $p < 0.05$ is considered significant.

Results

Delphi Panel and Model Process.

The Delphi panel surveys were answered by 13 national experts from 3 specialties (anaesthesiology, critical care medicine and nephrology) and 8 different institutions (public, private, university, military and regional cities hospitals). It was done in 2 rounds because the absolute average variability of responses between the first and second round was 3.62%, with a standard deviation of 3.81%. The final generic model [Figure 1] considers 32 activities, including 28 activities proposed in the initial model, 4 new activities proposed by the experts and in other hands, 7 activities from the initial model were excluded.

Figure Nº 1. Central Venous Access with Ultrasonography Process Model in BPMN Notation.

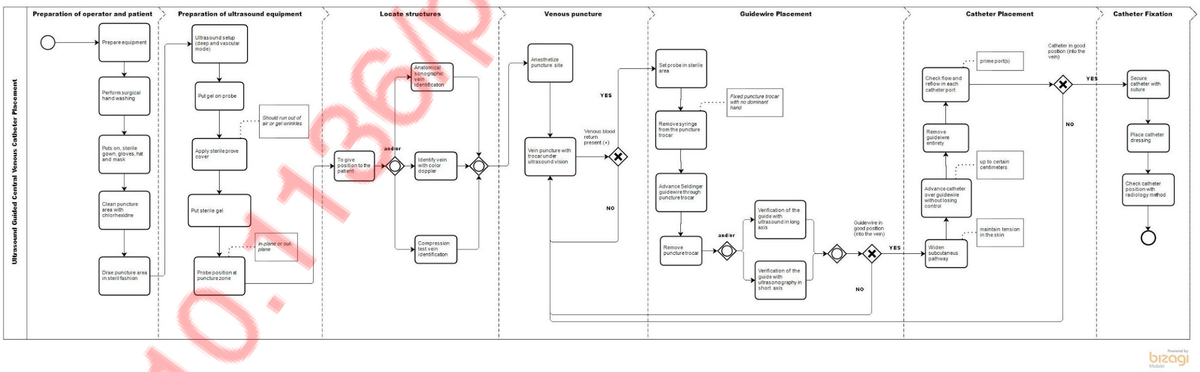


Figure 1. Representation of the generic model of a UGIJCVC installation in BPMN, where the outer rectangle (pool) encloses the complete process performed by an operator. Each column is a stage of the procedure. The activities are represented by squares joined by arrows representing the flow between each activity. Between some activities, there are gates that show flow to more than one activity, which can be inclusive (and/or) or exclusive (or); the beginning and the end of a procedure is represented by circles.

Process Mining Analysis

Rework Analysis

Optimal execution of the UGIJCVC process requires the absence of rework, i.e., a value of rework metric equal to 1 for all the activities of the process. This is the case of the EXP group, where all activities are executed only once. Figure 2 shows rework metric for all activities in the PRE and POST groups, whose averages, (DS) [range] are 1.1 (1-1.4) [1-1.8] to 1.1 (1-1.2) [1-1.3] respectively, p value 0.03. In most of the activities, there is a decrease of the rework in the POST group; this decrease is especially noticeable in the stages of Venous Puncture, Guidewire Placement, and Catheter Placement, objectively showing these stages as the most difficult stages of the process for the group being trained. Note, however, that in the POST records, a metric of 1 is not obtained for all the activities.

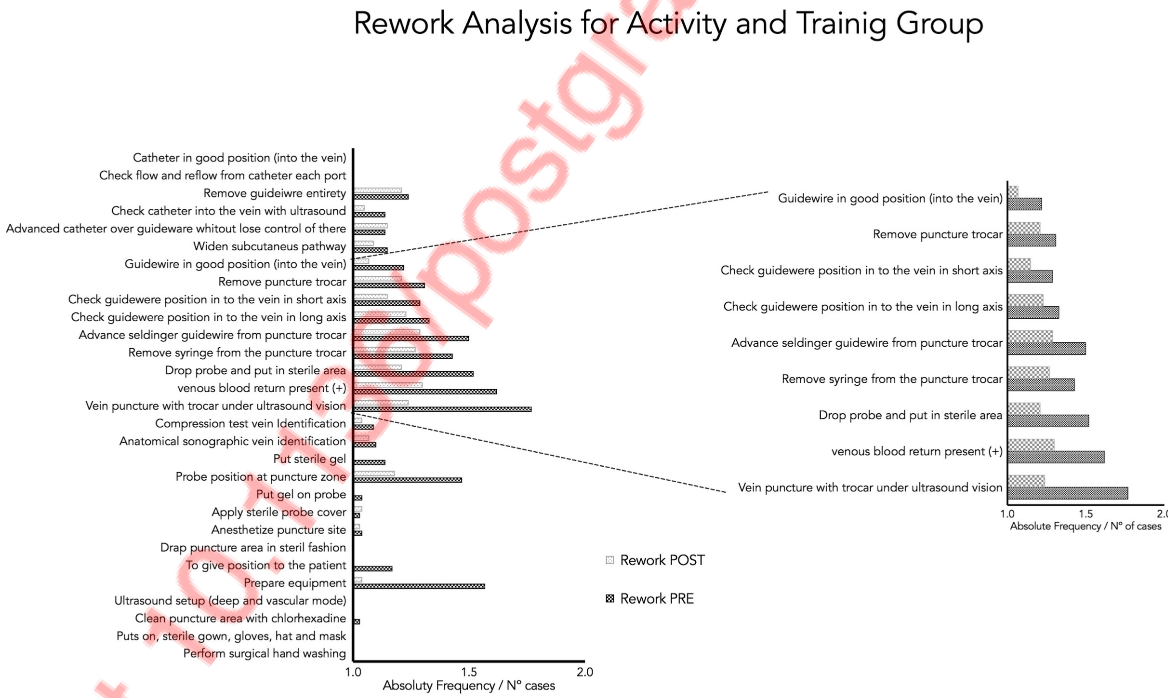


Figure Nº 2. Each horizontal bar represents Ultrasound Guided Central Venous Catheter placement activities in the generic model, ordered from first activity (Bottom) to last activity (Top), in pre-training (PRE) and post-training records (POST). Zoom of activities which concentrate mayor Rework Metric as RM= af/nc, where af represents the absolute frequency of the activity, and nc the number of cases where it appears.

Conformance Analysis

The conformance analysis (Figure Nº 3) shows the alignment-based fitness metric for the PRE, POST, and EXP groups in relation to the process model. The PRE group has a lower adhesion to the ideal process model, with an Alignment-Based Fitness value of 0.74 (0.68-0.78) [0.62-0.87] expressed as a median (25-75 percentile) [range], than the group POST with 0.82 (0.76-0.86) [0.62-0.9], and EXP with 0.87 (0.82-0.87) [0.80-0.93]. There are significant differences in control-flow between PRE, POST and EXP with values of p 0.0006 and <0.0001, respectively. However, there are no significant differences between POST and EXP.

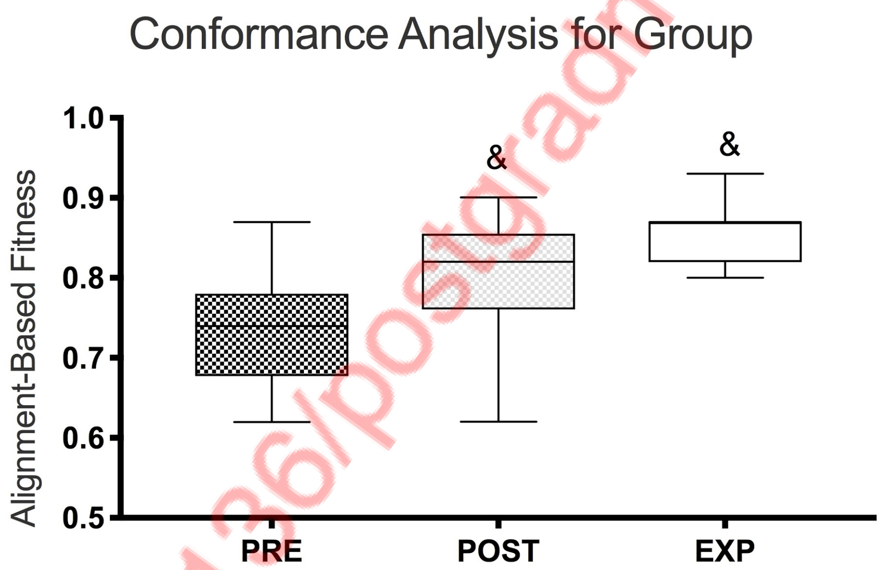


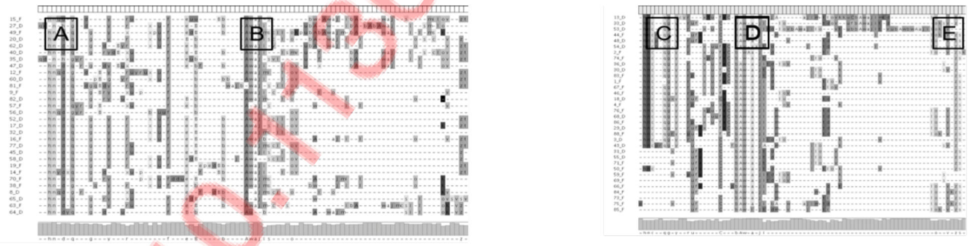
Figure Nº 3. Alignment-based fitness respect to the generic model of execution an Ultrasound-Guided Central Venous Catheter placement between resident pre-training (PRE), post-training (POST) and expert group. Values presented as median (25-75 percentile) [range]; & significative difference of PRE versus POST and EXP. Considerer statistically significant difference p ≤ 0,05.

Trace Alignment Analysis

In the graphic display of the Trace Alignment Analysis (Figure Nº4), each block represents each of the three groups identified: PRE, POST and EXP. In each block, the rows represent a particular execution, a trace, identified in the first column with an id. At the same time, each activity is represented throughout the block by a rectangle of a defined grey-scale colour and a letter inside it. In this way, trace alignment allows a visualization of all the executions of each group, making it clear within each one of them where there are areas with sequences of activities carried out in the same order in several executions.

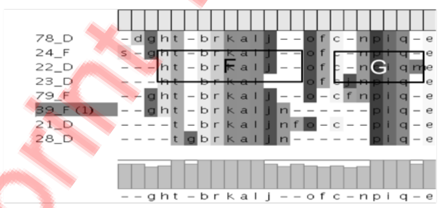
In this case, Trace Alignment Analysis shows considerable control-flow differences between each execution in the PRE group, showing two zones of sequential activities, first preparative activities (Zone A) and second, from when the blood is drawn until when the Seldinger guidewire is advanced (Zone B). However, the executions of the POST group increase the number of common areas to three: venous puncture (Zone C), preparation of implements (Zone D), and removal of the guide (Zone E). Finally, the executions of the EXP group show a much greater order, expressed in clear areas and longer sequences of activities carried out in the same pattern (Zones F and G).

Figure Nº 4. Trace Alignment Analysis



a) Trace alignment visualization Pre Training group

b) Trace alignment visualization Post Training group



c) Trace alignment visualization Expert group

FIGURE Nº 4.
Trace Alignment Analysis visualization of procedure execution in different groups. Each row represent an execution of UGIJCVC placement. Columns describe positions of activities in traces. Common executions patterns are captured in the form of well conserved zones (A, B, C, D, E, F, G).

Discussion

To the best of our knowledge, this is the first study using process mining methods to analyse data from the learning process of procedural skills in the healthcare field. The proposed analysis provides information that generates new perspectives of the execution of the procedure during the training process. We can identify which activities are more difficult from a control flow perspective, the adherence of the actual executions in relation to the ideal model of execution and common patterns of executions at different levels of procedural competencies.

The proposed control-flow analysis provides important information regarding execution patterns at different levels of competence to be obtained. A specific expert performance pattern does exist, which is different from the observed pattern in the novice level participant. A progression between these two stages can be inferred from the POST results. This way we can appreciate, when analysing the images of the Trace Alignment Analysis (Figure Nº4), that the Venous Puncture, Guidewire Placement and Catheter Placement stages are consistently executed in the same order by all the experts, and that the training demonstrates that the course allows an increase in order in these stages and in the overall execution of the procedure with a decrease in variability among residents. At the same time, we can see that the repetitions of activities are concentrated in the same stages (Figure Nº2), with higher values of rework prior to training, which decrease considerably once finished, which makes the beneficial effect of the course clear, even when they did not reach the level of the experts, who do not repeat any activity.

The ideal way in which a procedure should be executed should be made explicit prior to the start of the training [40] [41] [42]. This is not trivial, especially when trying to extract this knowledge from experts, because they can omit up to 70% of the information needed for proper execution [16]. To avoid this problem, Cognitive Tasks Analysis (CTA) has proven to be an effective tool, allowing the necessary steps, critical points, cognitive decisions, etc. to be clearly set out based on expert interviews [17][16][43]. Control-flow analysis offers new information based on a factual approximation of the executions of experts, which allows the CTA to be enriched, based on how the procedure is actually executed. In this way, we can consider that for this procedure, the experts do not make mistakes that oblige them to repeat activities and they perform activities consistently in the same order from venous puncture to catheter placement. An analysis of this type makes it possible to begin to characterize the executions of the procedures. We can identify key features at procedural level [44],[45], the technical skills that are critical and specific for the good result of each procedure, with the implications this knowledge has from a perspective of training and also of evaluation [46].

A recent review that aimed to collect the best evidence available for training in procedural skills from medium to greater complexity [47] points out the positive role of decomposition into units of knowledge and skill. It also highlights the need to limit the number of skills taught in each session to avoid student cognitive overload [48],[49],[50]. In the specific case of the our UGIJCVC training program, we subdivided the training into the 4 stations referred to in the materials and methods section, but the control-flow results suggest a new subdivision of the Venous Puncture, Guidewire Placement and Catheter Placement; because they represent the stages where residents make the most mistakes, forcing them to repeat these activities (rework metric more than 1). Including them all in a single session implies exposing the trainee to the most difficult stages at the same time. Our analysis, allows to create subdivisions based not only on logical considerations, but also upon factual procedural evidence of the motor skills load that the task implies. In addition, recognizing these difficulties would allow us to focus on the development of lower fidelity phantoms at a lower cost, in order to train doctors in each of these skills separately, and, at a later time, integrate these steps into a fluid activity in a high fidelity phantom. Furthermore, the good results in the preparation and cleaning stages, show that the online training section fulfils its cognitive role since the activities are consistently carried out without error and do not present important modifications after training. This online module prepared our students to start working on the next steps that put more demands on motor skills.

Medical education has progressively migrated towards a competence-based training approach, so valid evaluation tools are needed to be able to account for the achievement of these competences [51],[52],[53]. There are different ways of evaluating the acquisition of procedural competences and none of them are perfect, since each one has its strengths and weaknesses [8], [54], [13]. The result of the compliance analysis presented in this research shows the potential of using it as an evaluation tool from the control-flow perspective. This makes a comparison between the execution performed and a generic reference model. As described in methods, this metric is based on penalizing violations of the ideal pattern in terms of not executing an activity, thus collecting the information captured by the checklists, but it also penalizes the activities executed in a different order than expected, thereby a qualitative aspect usually captured by the global scales is collected. In the case of UGIJCVC, the Alignment-Based Fitness metric shows an increase as a result of the course that is also concordant with the performance of the experts. To demonstrate that its use as an evaluation tool is valid, a psychometric validation process ('assessing the assessment tool') [55-58] must be performed. In light of our results, we believe that it can be a good tool, for which there are still possibilities to improve their discriminatory capacity, working on differentiated weights of violations to the ideal pattern based on patient or clinician safety, patient comfort, and procedure outcome, a proposal that has also been suggested as a way to improve the checklist [59].

The feedback process, defined as information from any source about an apprentice's performance with respect to a task goal [60], plays an essential role in training in procedural skills. A recent meta-analysis oriented to its use in simulation demonstrates that it has a moderate effect on the training of these skills [61]. At the same time, it raises concerns about what, who, how, and when it should be delivered to maximize its positive impact on learning. In this context, the control-flow information obtained from the process mining analysis can be a source of standardized and valuable information to the trainee. For example, in the UGIJCVC training case, the results of trace alignment and rework analysis allow the learner to perceive the activities he/she executes properly, repeats, does not perform and even identify stages of greater difficulty when graphically visualizing more than one execution. This could be considered process-oriented feedback [62]. In any case, the possibility of using this information as a source of feedback should be planned considering that not only the content plays a role in its effectiveness, but also the execution method and how it has an impact on the motivation of the trainees [60].

Limitations

In our research, we retrospectively used data corresponding to the training of residents of several specialties, analysed before and after exposure to simulation training. Given that our objective was to explore the potential of control-flow analysis in a new context, the findings are a good approximation for the planning of prospective studies that will allow us to demonstrate the final usefulness of our approach. On the other hand, the executions of the EXP group were all carried out by anaesthesiologists from the same centre, which is a bias that may limit the generalization of the conclusions derived from the analysis of this group.

Conclusion

In conclusion, the control-flow analysis based on process mining tools can make explicit information about the executions carried out in different stages during training in the simulation context. This information, which can be obtained both from single executions and from the analysis of a complete group, allows the identification of the activities and stages that cause residents greater difficulty and how these are modified with the training. In addition, it allows the specification of some other patterns in the different levels of competence, making it a promising metric for the evaluation of a reference pattern.

References

1 Fecso AB, Szasz P, Kerezov G, *et al.* The Effect of Technical Performance on Patient Outcomes in Surgery. *Annals of Surgery* 2017;**265**:492–501. doi:10.1097/SLA.0000000000001959

2 Leape LL, Brennan TA, Laird N, *et al.* The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med* 1991;**324**:377–84. doi:10.1056/NEJM199102073240605

3 Cook DA, Hamstra SJ, Brydges R, *et al.* Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Medical Teacher* 2013;**35**:e867–98. doi:10.3109/0142159X.2012.714886

4 Barsuk JH, Cohen ER, Feinglass J, *et al.* Use of Simulation-Based Education to Reduce Catheter-Related Bloodstream Infections. *Arch Intern Med* 2009;**169**:1420–3. doi:10.1001/archinternmed.2009.215

5 Cohen ER, Feinglass J, Barsuk JH, *et al.* Cost savings from reduced catheter-related bloodstream infection after simulation-based education for residents in a medical intensive care unit. *Simul Healthc* 2010;**5**:98–102. doi:10.1097/SIH.0b013e3181bc8304

6 Zendejas B, Wang AT, Brydges R, *et al.* Cost: the missing outcome in simulation-based medical education research: a systematic review. *Surgery* 2013;**153**:160–76. doi:10.1016/j.surg.2012.06.025

7 Ilgen JS, Ma IWY, Hatala R, *et al.* A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015;**49**:161–73. doi:10.1111/medu.12621

8 Bould MD, Crabtree NA, Naik VN. Assessment of procedural skills in anaesthesia. *British Journal of Anaesthesia* 2009;**103**:472–83. doi:10.1093/bja/aep241

9 Hartman N, Wittler M, Askew K, *et al.* Validation of a performance checklist for ultrasound-guided internal jugular central lines for use in procedural instruction and assessment. *Postgraduate Medical Journal* 2017;**93**:67–70. doi:10.1136/postgradmedj-2015-133632

10 Barsuk JH, Cohen ER, Nguyen D, *et al.* Attending Physician Adherence to a 29-Component Central Venous Catheter Bundle Checklist During Simulated Procedures*. *Critical Care Medicine* 2016;**44**:1871–81. doi:10.1097/CCM.0000000000001831

11 Ma IWY, Pugh D, Mema B, *et al.* Use of an error-focused checklist to identify incompetence in lumbar puncture performances. *Med Educ* 2015;**49**:1004–15. doi:10.1111/medu.12809

12 Brydges R, Nair P, Ma I, *et al.* Directed self-regulated learning versus instructor-regulated learning in simulation training. *Med Educ* 2012;**46**:648–56. doi:10.1111/j.1365-2923.2012.04268.x

13 Ma IWY, Zalunardo N, Pachev G, *et al.* Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. *Adv in Health Sci Educ* 2011;**17**:457–70. doi:10.1007/s10459-011-9322-3

14 Neumuth D, Loebe F, Herre H, *et al.* Modeling surgical processes: a four-level translational approach. *Artif Intell Med* 2011;**51**:147–61. doi:10.1016/j.artmed.2010.12.003

15 Neumuth T, Jannin P, Strauss G, *et al.* Validation of Knowledge Acquisition for Surgical Process Models. *AMIA* 2014;**16**:72–80. doi:10.1197/jamia.M2748

16 Sullivan ME, Yates KA, Inaba K, *et al.* The Use of Cognitive Task Analysis to Reveal the Instructional Limitations of Experts in the Teaching of Procedural Skills. *Academic Medicine* 2014;**89**:811–6. doi:10.1097/ACM.0000000000000224

17 Wingfield LR, Kulendran M, Chow A, *et al.* Cognitive Task Analysis: Bringing Olympic Athlete Style Training to Surgical Education. *Surg Innov* 2015;**22**:406–17. doi:10.1177/1553350614556364

18 van der Aalst WMP. *Process Mining. Data Science in Action*. Second. Springer 2016.

19 Rojas E, Munoz-Gama J, Sepúlveda M, *et al.* Process mining in healthcare - A literature review. *J Biomed Inform* 2016;**61**:224–36. doi:10.1016/j.jbi.2016.04.007

20 Mans RS, van der Aalst WMP, Vanwersch RJB. *Process Mining in Healthcare*. Cham: : Springer 2015. doi:10.1007/978-3-319-16071-9_3

21 Bogarín A, Cerezo R, Romero C. A survey on educational process mining. *WIREs Data Mining Knowl Discov* 2018;**8**. doi:10.1002/widm.1230

22 Corvetto MA, Pedemonte JC, Varas D, *et al.* Simulation-based training program with deliberate practice for ultrasound-guided jugular central venous catheter placement. *Acta Anaesthesiol Scand* 2017;**61**:1184–91. doi:10.1111/aas.12937

23 Object Management Group. Business Process Model and Notation™ (Bpmn™) Version 2.0. omg.org. http://www.omg.org/spec/BPMN/2.0/ (accessed 9 Aug2017).

24 Scheuerlein H, Rauchfuss F, Dittmar Y, *et al.* New methods for clinical pathways-Business Process Modeling Notation (BPMN) and Tangible Business Process Modeling (t.BPM). *Langenbeck's Archives of Surgery* 2012;**397**:755–61. doi:10.1007/s00423-012-0914-z

25 Rolón E, Chavira G, Orozco J, *et al.* Towards a Framework for Evaluating Usability of Business Process Models with BPMN in Health Sector.

Procedia Manufacturing 2015;**3**:5603–10. doi:10.1016/j.promfg.2015.07.748

26 Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Information & Management* 2004;**42**:15–29. doi:10.1016/j.im.2003.11.002

27 Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000;**32**:1008–15.

28 Diamond IR, Grant RC, Feldman BM, *et al.* Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol* 2014;**67**:401–9. doi:10.1016/j.jclinepi.2013.12.002

29 Adriansyah A, Buijs JCAM. Mining Process Performance from Event Logs. In: *Business Process Management Workshops*. Berlin, Heidelberg: : Springer, Berlin, Heidelberg 2012. 217–8. doi:10.1007/978-3-642-36285-9_23

30 Lalys F, Jannin P. Surgical process modelling: a review. *Int J Comput Assist Radiol Surg* 2014;**9**:495–511. doi:10.1007/s11548-013-0940-5

31 Hagedorn J, Hailpern JM, Karahalios K. VCode and VData - illustrating a new framework for supporting the video annotation workflow. *AVI* 2008;:317–21. doi:10.1145/1385569.1385622

32 Levenshtein VI. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 1965;**10**:707–10.

33 Weigel A, Fein F. Normalizing the weighted edit distance. IEEE 1994. 399–402vol.2. doi:10.1109/ICPR.1994.576958

34 Yujian L, Bo L. A normalized Levenshtein distance metric. *IEEE Trans Pattern Anal Mach Intell* 2007;**29**:1091–5. doi:10.1109/TPAMI.2007.1078

35 Munoz-Gama J. *Conformance Checking and Diagnosis in Process Mining*. Springer 2016. doi:10.1007/978-3-319-49451-7

36 Carmona J, van Dongen B, Solti A, *et al. Conformance Checking*. 2018.

37 Bose RPJC, van der Aalst WMP. Process diagnostics using trace alignment: opportunities, issues, and challenges. *Information Systems* 2012;**37**:117–41. doi:10.1016/j.is.2011.08.003

38 Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* 1970;**48**:443–53. doi:10.1016/0022-2836(70)90057-4

39 Verbeek HMW, Buijs JCAM, van Dongen BF, *et al.* XES, XESame, and ProM 6. Cham: : Springer International Publishing 2011. 60–75. doi:10.1007/978-3-642-17722-4_5

40 Grantcharov TP, Reznick RK. Teaching rounds - Teaching procedural skills. *British Medical Journal* 2008;**336**:1129–31. doi:10.1136/bmj.39517.686956.47

41 Spruit EN, Band GPH, Hamming JF, *et al.* Optimal training design for procedural motor skills: a review and application to laparoscopic surgery. *Psychological Research* 2013;**78**:878–91. doi:10.1007/s00426-013-0525-5

42 Nicholls D, Sweet L, Hyett J. Psychomotor Skills in Medical Ultrasound Imaging. *Journal of Ultrasound in Medicine* 2014;**33**:1349–52. doi:10.7863/ultra.33.8.1349

43 Clark RE, Pugh CM, Yates KA, *et al.* The Use of Cognitive Task Analysis to Improve Instructional Descriptions of Procedures. *Journal of Surgical Research* 2012;**173**:e37–e42. doi:10.1016/j.jss.2011.09.003

44 Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine* 1995;**70**:194–201. doi:10.1097/00001888-199503000-00009

45 Page G, Bordage G. The medical council of Canada's key features project: A more valid written examination of clinical decision-making skills. *Academic Medicine* 1995;**70**:104–10.

46 Bordage G, Page G. The key-features approach to assess clinical decisions: validity evidence to date. *Adv Health Sci Educ Theory Pract* 2018;**23**:1–32. doi:10.1007/s10459-018-9830-5

47 Nicholls D, Sweet L, Muller A, *et al.* Teaching psychomotor skills in the twenty-first century: Revisiting and reviewing instructional approaches through the lens of contemporary literature. *Medical Teacher* 2016;**38**:1056–63. doi:10.3109/0142159X.2016.1150984

48 Fraser KL, Ayres P, Sweller J. Cognitive Load Theory for the Design of Medical Simulations. *Simul Healthc* 2015;**10**:295–307. doi:10.1097/SIH.0000000000000097

49 Young JQ, Van Merrienboer J, Durning S, *et al.* Cognitive Load Theory: Implications for medical education: AMEE Guide No. 86. *Medical Teacher* 2014;**36**:371–84. doi:10.3109/0142159X.2014.889290

50 Naismith LM, Cavalcanti RB. Validity of Cognitive Load Measures in Simulation-Based Training. *Academic Medicine* 2015;**90**:S24–S35. doi:10.1097/ACM.0000000000000893

51 Chin M, Lagasse RS. Assessment of competence. *Current Opinion in Anaesthesiology* 2017;**30**:236–41. doi:10.1097/ACO.0000000000000431

52 Epstein RM. Medical education - Assessment in medical education. *N Engl J Med* 2007;**356**:387–96. doi:10.1056/NEJMra054784

53 Davis MH, Harden RM. Competency-based assessment: making it a reality. *Medical Teacher* 2003;**25**:565–8. doi:10.1080/0142159032000153842

54 Chuan A, Wan AS, Royse CF, *et al.* Competency-based assessment tools for regional anaesthesia: a narrative review. *British Journal of Anaesthesia* 2018;**120**:264–73. doi:10.1016/j.bja.2017.09.007

55 Cook DA, Brydges R, Ginsburg S, *et al.* A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ* 2015;**49**:560–75. doi:10.1111/medu.12678

56 Pearce J. Psychometrics in action, science as practice. *Adv in Health Sci Educ* 2017;**20**:1–11. doi:10.1007/s10459-017-9789-7

57 Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)* 2016;**1**:31. doi:10.1186/s41077-016-0033-y

58 Boulet JR, Murray DJ. Simulation-based assessment in anesthesiology: requirements for practical implementation. *Anesthesiology* 2010;**112**:1041–52. doi:10.1097/ALN.0b013e3181cea265

59 Yudkowsky R, Tumuluru S, Casey P, *et al.* A patient safety approach to setting pass/fail standards for basic procedural skills checklists. *Simul Healthc* 2014;**9**:277–82. doi:10.1097/SIH.0000000000000044

60 Wulf G, Lewthwaite R. Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic Bulletin & Review* 2016;**23**:1382–414. doi:10.3758/s13423-015-0999-9

61 Hatala R, Cook DA, Zendejas B, *et al.* Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis. *Adv Health Sci Educ Theory Pract* 2014;**19**:251–72. doi:10.1007/s10459-013-9462-8

62 Sedrakyan G, De Weerdt J, Snoeck M. Process-mining enabled feedback: "tell me what i did wrong" vs. “tell me how to do it right.” *Computers in Human Behavior* 2016;**57**:352–76. doi:10.1016/j.chb.2015.12.040