Competence assessment of centres participating in clinical trials

Guillaume BOUZILLÉabc1, Pierre ASFARD, Marc CUGGIAabc, and Jean-François HAMELe

a INSERM, U1099, Rennes, F-35000, France

b Université de Rennes 1, LTSI, Rennes, F-35000, France

c CHU Rennes, CIC Inserm 1414, Rennes, F-35000, France

d CHU Angers, Service de réanimation médicale et médecine hyperbare, France

e CHU Angers, Maison de la recherche, France

Abstract. Procedures tested in clinical trials require adequate competence of one or more health professionals and an adequate environment of care. In the context of clinical trials, it is usually not possible to separate the effect of treatments studied from the effect of the proficiency levels. This work proposes a method to assess not individual proficiency but practice performance of centres participating in clinical trials. We used the LC-CUSUM test to retrospectively assess centres’ proficiency. We were finally able to ascertain one date for each centres from which they were considered as proficient. In conclusion, LC-CUSUM test seems suitable for centre’s performance assessment to avoid potential bias in analyses due to suboptimal achievements of studied procedures. Prospective proficiency analyses could also be considered in a preliminary stage to standard clinical trials.

Keywords. Learning Curve, Clinical Trials As Topic, Intention to Treat Analysis, Professional Competence

Introduction

Clinical trials are usually interested in testing a new treatment compared to a standard one. The goal is to assess benefits of the candidate treatment, with regard of a particular disease. Studied treatments may involve either new drugs or technical procedures like surgery, endoscopy or anaesthesia. Such procedures require adequate proficiency of physicians to be correctly performed. In the context of clinical trials, proficiency of physicians is not ascertained at the beginning of the study. In this case, it is not possible to separate the benefits of the treatment from the physician’s proficiency.

Several studies focused on the individual assessment of capability for a particular procedure. The goal is most frequently to identify when trainees become proficient or when experimented physicians acquire a new technique. These studies use statistical methods based on the cumulative summation test (CUSUM)1 or the learning curve cumulative summation test (LC-CUSUM)2. CUSUM is primarily developed to monitor if a loss of performance occurs in a process known to be efficient. At the opposite, LC-CUSUM, which is a variant of CUSUM coming from the medical field, is especially

1 Guillaume Bouzillé. LTSI, Campus de Villejean, Université de Rennes 1, 2 Avenue Du Professeur Léon Bernard - CS 34317 - 35043 Rennes Cedex, France. guillaume.bouzille@gmail.com
developed to assess physicians’ proficiency. This method allows to identify whether a process originally not efficient has an acceptable performance rate and to conclude when this performance rate is reached. To be monitored, studied procedures need to be summarized either with a qualitative process, such as a sequence of success or failure for the realization of a given technic, or with a quantitative measure, such as the time needed to perform a surgical procedure. LC-CUSUM may be useful in several medical fields and has already been used to monitor training of a new surgical or endoscopic technic, ultrasound interpretation or other skill acquisition.

In the context of clinical trials, some authors suggested to assess individual competence of physicians before their participation, to ensure a similar performance level at the beginning of the study. However, every procedure does not rely on individual proficiency but might involve several care professionals or an adequate environment. All centres may not have the same resources to get ready for participation in clinical trials. The latters include indeed large university hospitals as well as local hospitals. Several centres may also have already participated in trials whereas others are novices. This variability in the environment might play a role in the quality of care delivered among centres. For all these reasons, individual performance assessment may not be the most suitable in this case. A more relevant assessment would be to consider competence acquisition globally, in centres. We propose here a method to study not individual proficiency but practice performance of centres with regard to the particular context of clinical trials.

1. Methods

The only difference between studying individual or global proficiency is the statistical unit used for the LC-CUSUM test. Global performance can be summarized as well with a qualitative or quantitative process or even with a process made from mixed variables. Thus, we used the same method as individual assessment to build learning curves based on the LC-CUSUM test, but at a centre scale. The goal is to see whether centres are proficient from the beginning of the trial and if not, whether they reach an optimal predefined level of performance during the trial. Both standard and candidate procedures must be studied separately in each centre since their achievements are not likely to follow the same level of performance. The principle of LC-CUSUM is to test after each new realization if the process can be considered as being efficient regarding a predefined level of performance and the previous realizations. Several parameters are necessary to build learning curves, such as a predefined adequate level of performance, an acceptable deviation from the predefined level of performance and the standard deviation of the studied process. A threshold must also be set to conclude to adequate proficiency when the learning curve crosses it. Its value is usually chosen on simulated datasets to maximize the probability to ascertain proficiency for centres really proficient (sensibility) and to maximize the probability not to ascertain proficiency if centres are not efficient (specificity). Finally, this method allows to evidence one date from which centres can be considered as proficient.

To illustrate this work, we applied our method to the multicentre, randomized, controlled SEPSISPAM trial. It aimed to determine whether targeting a mean arterial pressure of 80 to 85 mm Hg would decrease 28-day mortality, as compared with targeting a mean arterial pressure of 65 to 70 mm Hg during initial resuscitation of patients with septic shock. The target mean arterial pressure was to be maintained for a
maximum of 5 days or until the patient was weaned from vasopressor support. The mean arterial pressure was measured every two hours during the 5-day protocol.

We considered that the capability of centres to maintain patients in the predefined targets of mean arterial pressure was reflecting proficiency to correctly follow either standard or high-target protocol. We built a process from the original measures of mean arterial pressure representing the time spent in target. This was computed for each centre and for each arm of the study, as follow: we took the number of measures in target for all patients present in a centre a given day divided by the overall number of measures realized the same day. This was computed each day one or more patient from the study was present in a centre. This finally gave us a sequence of time spent in target values for both standard and candidate target of mean arterial pressure, in each centre. As no reference was available in literature, we chose a rather conservative value of at least 30% of time spent in target as the reference level of performance. We set the standard deviation for the time in target process and the acceptable deviation rate to 20% and a threshold representing adequate proficiency to -8. We developed an R package to build learning curves and to realize the preliminary stage on simulated dataset. We used the R Statistical Software, version 3.1.1.

2. Results

The SEPSISPAM trial included 776 patients randomized into two groups of 388 patients among 27 centres. We had 22607 measures of mean arterial pressure from which we computed 2445 values of time spent in target. It gave us on average 46 values of time spent in target for each arm and each centre of the study (standard deviation of 43). We highlighted three kinds of possible profiles (Figure 1). The first one (example: centre 11) is when the learning curve did not reach the threshold (8 centres out of 27 for the low-target group and 5 out of 27 for the high-target group). This occurred when centres did not become proficient during the overall time of the trial. The second case (example: centre 30) is when centres had a learning phase during the trial (12 centres out of 27 for the low-target group and 17 out of 27 for the high-target group). The last case (example: centre 9) is when the learning curve crossed the threshold within 8 days (6 centres out of 27 for the low-target group and 5 out of 27 for the high-target group). We considered that centres with a curve reaching the predefined level of performance in the 8 first days were in fact efficient from the beginning of the trial, without any learning phase.

![Figure 1](image1.png)

**Figure 1.** Curves of centres not proficient (a), with a learning phase (b) and proficient (c)

Figure 2 shows learning curves of centres according to the mean arterial pressure target and the number of inclusion performed in centres. The median delay to reach threshold was 6 days (IQR: [5 ; 10]) for the low-target group and 6.5 days (IQR: [4.25 ; 8]) for the high-target group. During the overall study, average time spent in
target was 29% (s.d: 23%). Patients included in proficient centres were significantly more often in target than patients included in centres not proficient (30 ± 24% versus 22 ± 23%; p < 0.0001).

We thereafter split patients into two groups, according to their date of recruitment compared to the date of the centre’s proficiency acquisition: 81 out of 388 patients from the low-target arm (21%) and 48 out of 388 patients from the high target arm (12%) were included before optimal centres’ performance. Conversely, 307 patients (79%) from the low-target arm and 340 patients (88%) from the high-target arm were included once centres were proficient. Original statistical analyses can then be performed stratified whether patients were included in proficient centres, or not. SEPSISPAM results from the subgroup statistical analysis will be the subject of a specific paper.

![Learning curves of centres, using the LC-CUSUM method.](image)

**Figure 2.** Learning curves of centres, using the LC-CUSUM method.

### 3. Discussion

Competence assessment is almost always studied with an individual perspective in medical field. We proposed here a method to assess proficiency at a centre scale in the context of clinical trials. We performed a retrospective analysis of centres participating to the SEPSISPAM trial. The LC-CUSUM test let us highlight different profiles regarding delays to be considered as proficient. We showed that centres had various performance levels regarding their ability to maintain patients in the predefined mean arterial pressure target. The LC-CUSUM test allowed us to ascertain the date from which centres had an optimal level of proficiency and so when they were ready to include patients. At the opposite, patients included before proficiency acquisition were in fact included in centres not ready to participate in the trial.

We thereafter performed data analyses stratified whether patients were included in centres proficient or not. Analyses performed only in patients included in proficient centres have the advantage to avoid bias due to the learning phase on the effect of the procedure studied. Conversely, analyses realized on patients included before proficiency of centres permitted to investigate potential issues due to suboptimal treatments during the learning phase.
However, retrospective analyses are not optimal. The design of the original trial does not take into account this ancillary study. Hence, the original sample size does not permit to manage power of subgroup analyses. Proficiency assessment in a prospective way would permit to overcome the possible lack of power. A two steps study could be realized with a preliminary inclusion phase to assess performance of centres aiming to participate to a clinical trial and finally the real inclusion phase for the comparison of procedures, once centres are known to be proficient. This kind of design would have several advantages. First, statistical analyses could be performed on an intention to treat basis with the necessary number of patients but included in proficient centres. This must avoid any potential bias due to suboptimal procedure’s achievement during potential learning phase. Second, the preliminary phase could permit to assess the delay to be efficient to the new procedure studied. Complementary analyses could be done on patients included in centres not yet proficient. This could help better understand the benefits and potential adverse events of procedures studied in real suboptimal conditions.

We did not consider possible loss of performance during the clinical trial since it was out of the objective of the study. However, other methods to monitor healthcare quality have been described in literature, using for example the CUSUM test\cite{1,11,12}. These methods could be complementary to a preliminary learning phase assessment with the LC-CUSUM test. Anyway, further investigations must be done to assess feasibility and the added of prospective analyses.

References