MONITORING THE FUTURE

Leo Huberts

When the heart stops pumping, the body is in trouble. Immediate action is required to reboot the heart and get the blood circulation flowing. Sometimes these actions are taken on time and a life is saved (e.g., football player Christian Eriksen at the European Championship in 2021), sometimes the actions are late or insufficient and have lasting consequences (e.g., football player Abdelhak Nouri of Ajax at an exhibition match in 2017). We cannot determine the problem by simply observing from the outside. We do not see the heart rate and have only indirect hints towards blood flow (the color of the skin). Often there are no clear warning signs. The person seems fine, until they don't. Detecting physical problems such as a heart attack as quickly as possible requires monitoring of vital signs such as heart rate, blood pressure, and flow, oxygen levels, etc. The functioning of a body can be described as a complex proces. The vital signs give us an indication of the overall state of that process.

Monitoring the present

A large variety of processes are being monitored in industry and services. When we talk about monitoring a process, we generally mean closely watching the process in real time to find problems. The current state of the process is assessed to determine if intervening action is required. In factories the weight and dimensions of products are monitored, the temperatures of machines, and the production time. In banking, the location and amounts of transactions are monitored to detect fraud and theft, financial markets are monitored to mitigate risk and the duration of service inquiries is monitored to improve customer service. In healthcare we monitor the vital signs of patients, but also the length of stay to detect inefficiencies and medicine stock to optimize storage and drug availability. To support public policy, we monitor crime statistics, family incomes and Covid-19 cases and hospital admissions.

Looking forward

The existing monitoring applications overwhelmingly focus on the current state of a process. Problems are detected when they occur. However, using the explosion in available data and computing power, it is becoming more and more attainable to monitor the future states of a process. Instead of monitoring the current state of a process, we can model future outcomes from the effects of current and past patterns in process indicators. For example, instead of monitoring heart rate to detect heart failure, we model the risk of heart failure as a function of indicators such as blood pressure, heart rate, and oxygen levels. The patterns of these indicators can hint towards future problems before they occur, enabling preventive action. Such action can prevent defective batches of products and machine damage in factories, fraud in financial markets, and large-scale outbreaks of infectious diseases.

Monitoring Processes in the Age of Big Data

In *Statistical and Predictive Process Monitoring; Monitoring Processes in the Age of Big Data* (Huberts, 2021) we outline the shift from monitoring the current state (statistical process monitoring) to the future state of a process (predictive process monitoring). Firstly, the increase in available data and processing power enables improvements when monitoring the current state. For example, using the central limit theorem, fewer assumptions are needed to monitor non-normal data when using subgroups of sufficient size. This means it becomes easier to monitor a process indicator that has a non-normal distribution. Furthermore, updating the process parameters during monitoring can greatly improve the performance of the monitoring procedure.

Let's illustrate this with an example. Suppose we want to monitor the time it takes to perform a common surgical

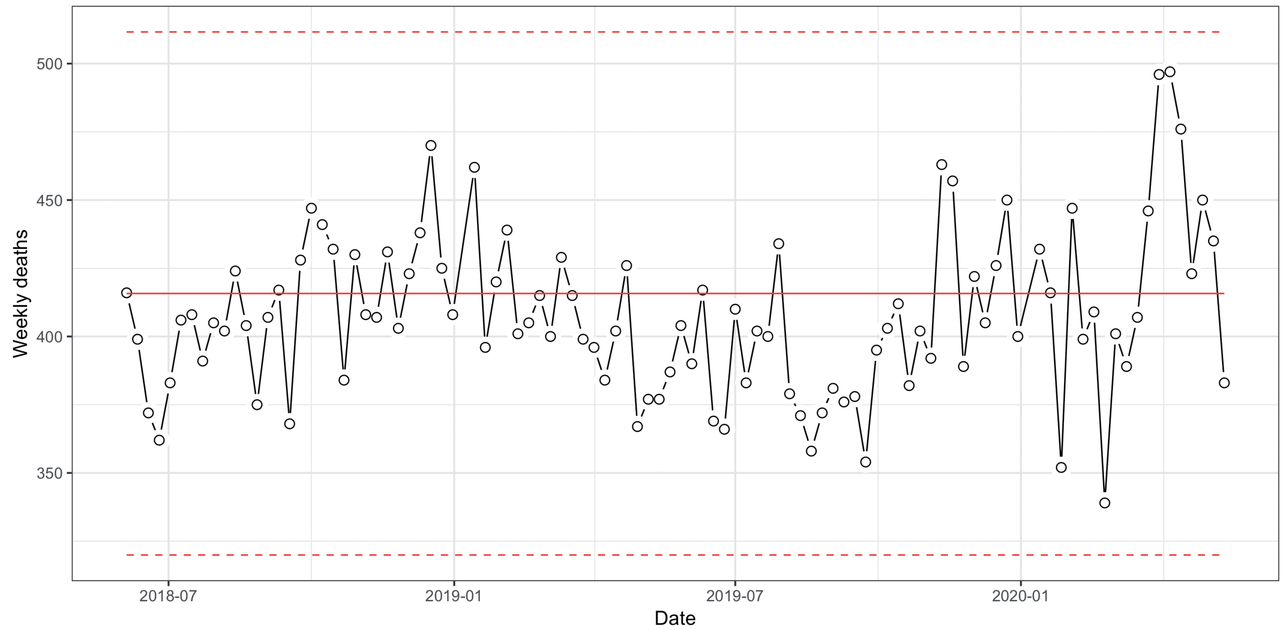

Figure 1. An example of a Shewhart control chart showing the weekly death rate among people of ages 0 to 65 in the Netherlands

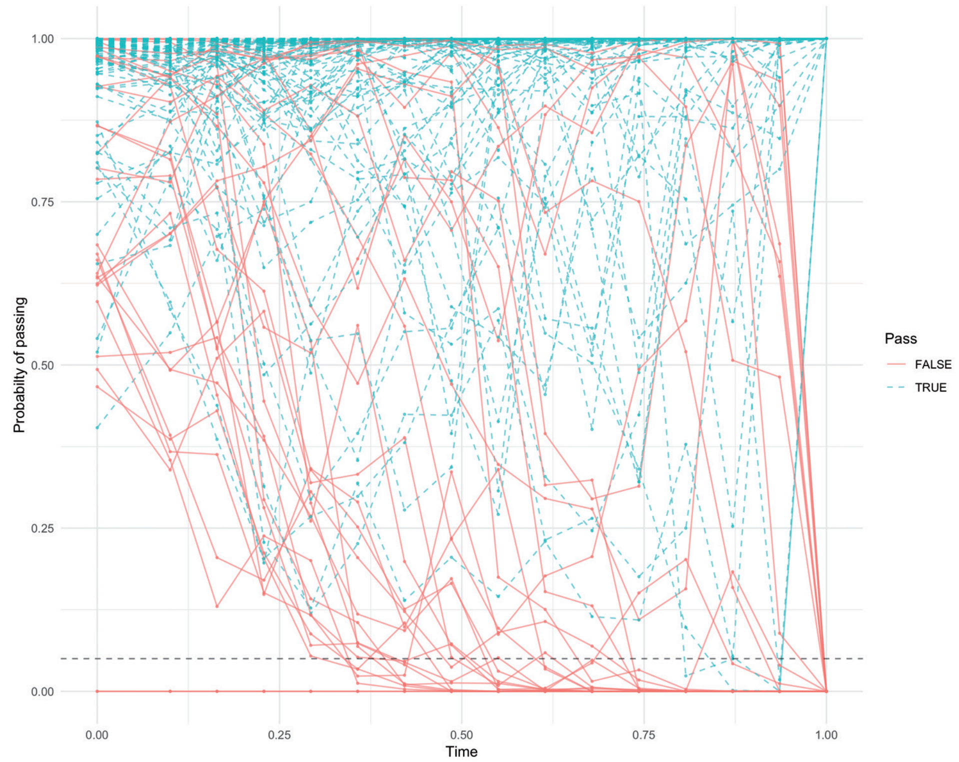

Figure 2. Graph showing the estimated probabilities of passing for a set of students throughout a school year

procedure such as an Appendectomy. The durations are right-skewed, thus not normally distributed. Furthermore, we want to start monitoring as soon as possible. We can monitor this process using a control chart (see Figure 1 for an example of a Shewhart control chart). Much of the existing control chart theory uses the assumption of normality. By taking subgroups of sufficient size, we can circumvent the normality assumption (see chapter 2 of Huberts, 2021, for more details). To start monitoring, we estimate the in-control situation using the available data. These data should be for operations that are representative for the in-control process. The estimates are imperfect and updating these estimates during monitoring will improve the performance of the monitoring procedure (see chapter 3 of Huberts, 2021, for more details). Incorporating a delay in the updates can prevent problems (see chapter 4 of Huberts, 2021, for more details). In the long run, these adjustments to traditional process monitoring will reduce the number of false alarms, improve the detection rate, and increase usability.

As discussed, the increase in the amount of available data and computing power can facilitate predictive monitoring. In the second part of *Huberts (2021)*, we discuss this transition to 'monitoring the future'. Two predictive process monitoring case studies are presented, one in mental healthcare and one in education.

Monitoring mental health crises

In the first case study, we investigate monitoring mental health crises to assist healthcare workers to improve planning. Using a unique big data set on all Dutch citizens, we zoom in on people diagnosed with schizophrenia. Schizophrenia is a debilitating disease estimated to affect 1% of the population. One of the properties of schizophrenia is that people often relapse into crisis care. In Huberts et al. (2021) we set out to predict these crises, to enable preventive care and improved resource planning. Using a wide variety of variables on healthcare activities, diagnoses, income, and personal information over 250 variables were constructed based on input from practitioners. These variables were modeled to predict the probability of a crisis. Gradient boosting outperformed other methods (regression, support vector machines, random forest, etc.) in terms of predictive accuracy and efficiency. The next step was to set up a monitoring procedure based on the predictions. A threshold is needed to determine when a predicted probability is too high. It was unclear how to determine this threshold. Therefore, in Huberts et al. (2021), we propose a simple tuning procedure to find a threshold that will produce an acceptable false alarm rate. The results are promising, with performance depending on the time frame and chosen false alarm rate.

		C					
		0.05	0.1	0.25	0.5	0.75	0.999
TIME	0.0	1 (0.07)	1 (0.07)	1 (0.07)	0.67 (0.13)	0.74 (0.47)	0.25 (0.93)
	0.1	1 (0.07)	1 (0.07)	1 (0.07)	1 (0.27)	0.71 (0.40)	0.71 (0.40)
	0.3	1 (0.10)	1 (0.20)	0.85 (0.37)	0.76 (0.53)	0.67 (0.67)	0.27 (1)
	0.5	1 (0.33)	1 (0.43)	0.94 (0.53)	0.79 (0.63)	0.67 (0.67)	0.34 (0.97)
	0.7	1 (0.57)	1 (0.63)	0.88 (0.73)	0.77 (0.70)	0.70 (0.77)	0.40 (0.97)
	0.9	0.90 (0.63)	0.86 (0.63)	0.88 (0.70)	0.81 (0.70)	0.81 (0.73)	0.59 (0.90)
	1.0	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)

Table 1. Precision (Recall) for monitoring student failure for various values of threshold C and time t

Monitoring student performance

In Huberts, Schoonhoven & Does (2020) we investigate predictive process monitoring for high school students. This challenge was proposed by high school managers and consists of identifying over- and underperforming students as early as possible. This enables targeted intervention by teachers and management. In Huberts, Schoonhoven & Does (2020) we propose a hierarchical Bayesian model for statistical process monitoring on the one hand and predictive process monitoring on the other. Using this model, schools can monitor at the individual course grade level, student level, and school level either signaling current problems (statistical process monitoring) or early warnings of future problems (predictive process monitoring). The model identifies much more structure in student performance than other investigated modeling techniques do. Figure 2 shows a resulting 'control chart', monitoring the end-of-year performance for a set of students. The chart shows the probability of passing the academic year and, depending on a chosen threshold, signals when this probability is too low. The system works well, for example identifying a quarter of the students that will fail the year with 100% precision after just 10% of the year has passed (see Table 1).

In conclusion, the increase in computing power and the explosion of available data is enabling improvements in monitoring the current state of the process and in monitoring the future state of that process. This predictive process monitoring has great potential to enable preventive action in services and industry, as demonstrated with examples in mental healthcare and education (see *Huberts, 2021,* for more details). At the Amsterdam Business School, we will continue working on these procedures, combining predictive modeling with process monitoring methods. Signaling as early as possible can be imperative in taking preventive measures in sectors such as healthcare, education, manufacturing, maintenance, and more. It can improve the quality of products and services.

References

Huberts, L.C.E. (2021). *Statistical and Predictive Process Monitoring; Monitoring Processes in the Age of Big Data (Doctoral Thesis)*. University of Amsterdam. https://pure.uva.nl/ws/files/59618725/Thesis.pdf.

Huberts, L.C.E., Does, R.J.M.M., Ravesteijn, B., & Lokkerbol, J. (2021). Predictive monitoring using machine learning algorithms and a real-life example on schizophrenia. Early view in *Quality and Reliability Engineering International*.

Huberts, L.C.E., Schoonhoven, M., & Does, R.J.M.M (2020). Monitoring student progress: A case study to predict student success or failure. Early view in the *Journal of Quality Technology*.

Leo C.E. Huberts is assistant professor in Business Analytics at the University of Amsterdam. He studies and teaches applied statistics and machine learning with an emphasis on process monitoring and analytics for a better world. This article is based on the work for his recent PhD thesis *Statistical and Predictive Process Monitoring; Monitoring Processes in the Age of Big Data* (2021).
E-mail: L.C.E.Huberts@uva.nl