RATIONALE FOR REPLACEMENT OF THE DESTRUCTIVE TEST BY NON-DESTRUCTIVE ONE IN MEDICAL DEVICES MANUFACTURING

Oldřich Beneš[1], David Hampel[1]

[1] Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

Abstract

Due to expanding demand for the level of testing on one side and reduction of costs on the other side, the question how to replace expensive destructive testing of medical devices without compromising the quality of final product arising urgently. This situation is common within all highly regulated industries – in this article is addressed the problem from medical device manufacturing industry. Based on real data containing testing and validation datasets, logit model and classification tree model are estimated for establishing the relationship between result of destructive test and measurements of explored device. Results point to possibility of replacing destructive test by non-destructive one in our case.

Keywords: classification tree model, destructive test, logit model, medical devices manufacturing, non-destructive test

INTRODUCTION

Every day business in highly regulated industries, like medical devices manufacturing, is simply the conflict. On the one side there are two main requirements in terms of manufacturing process definition. Maybe it would be better said duties. During development phase there should be done the process validation activities where the evidence of process capability is given. After the market launch during product life cycle the process control is required to have the evidence that the processes are within the same conditions as during process validation testing were.

But in the same time the strong demand for reducing of time and costs needed for any new product development or even the change implementation should be reflected, and the manufacturing costs are of high interest because of competition in the field. The economic side of the problem is covered for example in Marešová *et al.* (2018). Authors analyse and evaluate the role that economic indicators and economic effectiveness during the development of a medical device. Study further expanded on how other industries are different in terms of development as compared the medical industry. Real life is rather well described in detail.

Actual situation is even more complicated for manufacturers active under EU laws, where new MDR 2017/745 is coming (details can be found in Clemens, 2018). It can be stated that there is nothing wrong with MDR itself. But the way how it going to be implemented and readiness of all players on medical devices manufacturing and marketing creates the doubts and worries. One of the possibilities how to fulfil these requirements is to replace the expensive destructive testing from validation phase by much cheaper and faster

non-destructive testing during production phase. To implement this without compromising the quality and reliability of manufacturing processes the strong relation between the result of direct, typically destructive, test and the non-destructive testing of products during standard production could give the solution. From the practical point of view the way how to provide system with evidence like this should be described and rationalized in advance what means during the planning phase.

When company plan to introduce a non-destructive test instead of a destructive test, it has information coming from a destructive test typically. From this test coming the direct answer if the requirement was fulfilled. This value could be called structural value. In terms of target described above, depending value which could be measured by non-destructive test must be found. In the case of success so-called response variable have to be selected. All of this is typically job of development and process engineers. Now is time for quality assurance engineers and the first choice is to find and set up the relation between those two values. The simplest way can be the standard linear model (e.g. a simple regression model) and can be thought of as having two 'parts'. These are the structural variable and the response variable.

In common there are several tools well described in ASTM standards like ASTM (2018), which covers methods and equations for computing and presenting basic statistics and ASTM (2017) which deals with regression analysis. The problem comes because structural component is typically more or less normally distributed, covered a normally distributed error and rather often is the random component. The response variable could not be normally distributed. The idea can be to find other ways than these depending on normality assumption how to support and quantify relationship between two values and describe the way how to set up strong evidence of adequacy of such non-destructive testing during routine production. In the specific situation the goal is to verify possibility of replacement of destructive testing based on tensile test of the component by the non-destructive testing containing only measurement of the component part.

Statistical or artificial intelligence methods from the field of classification can be a correct solution. In Beneš and Hampel (2016) is explored destructive test of the part, where two pipes of the same diameter are connected. In the first process step one pipe is expanded by the tool. This tool is pressed against the pipe by defined, constant, force therefore the deepness of widening is depending on the same characteristics of material like the firmness of their connection later on. Next process step is to connect both pipes by the pressure. Destructive test during design verification is tensile strains test and the force is measured. The device passed the test when it is not broken after the test. For the purpose of non-destructive test evaluation, overlap length of pipes is measured. Tensile strength was modelled by pipes overlap length. The relationship was identified as unsymmetrical and was well described by cubic polynomial. Finally, logit and probit models were used to describe results of destructive test.

The aim of this paper is to evaluate the possibility of replacing destructive testing based on tearing of the component with non-destructive testing involving several measurements of the component. More specifically, without empirical analysis, it is unable to assess whether the measurements taken in the context of non-destructive test have any relation to the result of the destructive test. If it will be possible to find such a relationship by several models, a model that better identifies the cases where the product did not pass the destructive test will be preferred.

MATERIALS AND METHODS

The problem concerns production of the relatively simple but important part of advanced medical device illustrated in Fig. 1. The standard process control tool is destructive tensile test on the samples with given force on the limit. The acceptance criterion is that part should pass this test without breakage. Finally, binomial variable indicates that device passed the test (1) or failed in the test (0) is analysed.

When suggesting replacement of this destructive test by non-destructive one, following variables were considered (based on the opinion of expert in manufacturing of this type of medical devices): Fork distance (*Fdist*), Parallelism (*Par*) and Depth of crimping (*Dcri*), see Fig. 1. Fork distance is designed to be 5.3 mm, both lower and higher values means an unwanted condition that can cause device damage in real use. Ideal Parallelism value is equal to zero, high values are unwanted. Correct depth of crimping is set to 0.45 mm; lower values can cause disconnecting of the device and higher values may cause breakage of the component.
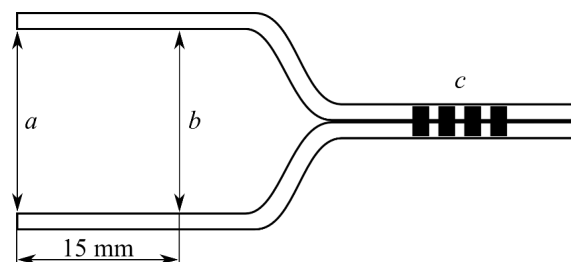
1: *Illustration of product under investigation with associated characteristics. Measure a is called Fork distance. Together with analogous distance b, so-called Parallelism characteristic is defined as absolute value of a – b. Finally, c represents Depth of crimping.*

I: *Characteristics of investigated datasets*

Dataset	Count		Mean			Standard deviation		
	Total	Failed	*Fdist*	*Par*	*Dcri*	*Fdist*	*Par*	*Dcri*
Training	120	29	5.361	0.192	0.438	0.033	0.046	0.014
Validation 1	60	1	5.340	0.179	0.449	0.038	0.050	0.006
Validation 2	60	9	5.383	0.153	0.444	0.027	0.039	0.016
Validation 3	60	9	5.352	0.156	0.433	0.039	0.041	0.011

This is a real case of manufacturing and testing within a medium-sized medical manufacturing company that has provided data and other relevant information. To assess the possibility of replacing the destructive test by non-destructive one, results of 6 typical destructive test series, where failed pieces appear, 60 devices were tested in each, are considered. The randomly selected three test series are used as testing dataset (batches are merged with idea to reach results robust in the terms of the ability to describe the possible variability of individual batches), following three as validation datasets (which corresponds to the practice of destructive testing). Variables *Fdist*, *Par* and *Dcri* are included together with result of destructive test. Characteristics of these datasets are given in Tab. I.

The question arises whether results of destructive test can be explained by measures *Fdist*, *Par* and *Dcri*. For this purpose, logit model (see for example Agresti, 2007) was employed as a baseline tool for binary classification. From experience with analysis presented in Beneš and Hampel (2016), cubic polynomials for *Fdist* and *Dcri* and with linear form of *Par* is considered as a starting model. Further, classification tree model (see for example Breiman *et al.*, 2017) is employed, which can reveal complicated relations in data. This model belongs to artificial intelligence methods. In comparison to majority of these methods, classification tree calculations can be explained in details and outputs are presented clearly. This makes the resulting decision trees a very suitable basis for managerial decisions. Binary classification characteristics are elaborated in accordance with Klepáč and Hampel (2017). Techniques related to ROC curve are described in Staňková and Hampel (2018, 2019). All calculations were performed in computational system Matlab R2019b. Level of significance was set to 0.01.

RESULTS

Logit model with cubic polynomials was estimated, but the third powers of variables were not statistically significant. The third powers were removed from the model gradually, and the resulting model with all variables significant is presented in Tab. II. Used variables have power to describe results of destructive test, but it is necessary to investigate quality of this description.

II: *Estimated logit model*

Parameter	Estimate	SE	T	*p*-value
Intercept	-52191	12854	-4.06	< 0.001
Fdist	18621	4627.1	4.02	< 0.001
Fdist^2	-1740	432.21	-4.03	< 0.001
Par	-42.15	13.448	-3.13	0.002
Dcri	10819	2917	3.71	< 0.001
Dcri^2	-12260	3320.2	-3.69	< 0.001

III: *Prediction accuracy (in %) of estimated logit model*

Dataset	Total accuracy	Failed accuracy	Passed accuracy
Training	95.0	82.8	97.4
Validation 1	91.7	100.0	91.5
Validation 2	96.7	88.9	98.0
Validation 3	96.7	77.8	100.0

Quality measures of logit model predictions are summarised in Tab. III. Relatively high total accuracy was reached for testing dataset as well as for validation datasets. This is given mainly by high accuracy of passed cases predictions. Although the absolute number of wrongly classified failed cases is low, it is hardly acceptable not to catch just one potentially defective product. Note that logit model with cubic polynomials resulted in slightly higher total accuracy and passed accuracy, but there was about 60% and 30% drop in failed accuracy for validation sets 2 and 3.

Results described above are related to logit model with "standard" threshold 0.5. Because such a setting may not be the best choice, the ROC curve has been estimated to find the optimal threshold. Quality of classification was measured by the area under ROC curve characteristics, which resulted 0.985 – this can be interpreted as "perfect classification". Both "distance to corner" and Youden index (equal to 0.885) indicated optimal threshold of 0.821. Within this setting, prediction accuracy characteristics changed slightly, see Tab. IV. For the training dataset, total accuracy remained the same, failed accuracy increased for about 10 percentage points and passed accuracy
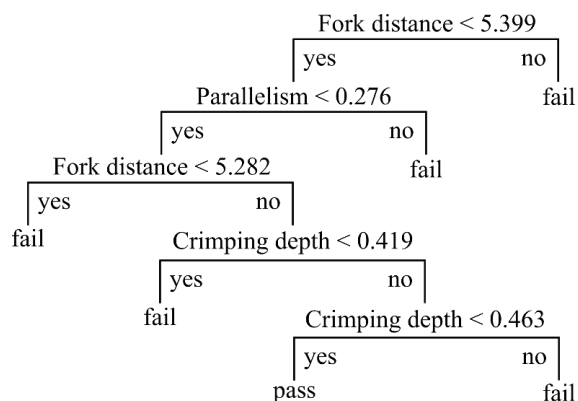
IV: *Prediction accuracy (in %) of estimated logit model with optimal threshold*

Dataset	Total accuracy	Failed accuracy	Passed accuracy
Training	95.0	93.1	95.4
Validation 1	88.3	100.0	88.1
Validation 2	93.3	88.9	94.1
Validation 3	95.0	88.9	96.1

decreased for 2 percentage points. When failed accuracy is primary characteristics, it is possible to declare that the new threshold improves results of classification. The result is not the same for validation datasets: total accuracy decreased as well as passed accuracy, but failed accuracy increased only for the third validation dataset.

Finally, classification tree was estimated with default setting for medium-size trees. Resulting decision tree is presented in Fig. 2. As expected, variables *Fdist* and *Dcri* are presented twice, *Par* once. Decision tree gives algorithm for getting prediction for arbitrary measures of *Fdist*, *Par* and *Dcri*. Limiting values of these values are given, but it is necessary to use them simultaneously.

Quality of classification tree model predictions can be assessed by Tab. V. Total accuracy is slightly lower than for logit model, but failed cases accuracy is higher in average. Especially, failed cases accuracy is 100 % for testing dataset.

2: *Decision tree based on estimated classification tree model*

V: *Prediction accuracy (in %) of estimated classification tree model*

| Dataset	Total accuracy	Failed accuracy	Passed accuracy
Training	91.1	100.0	94.7
Validation 1	91.7	100.0	91.5
Validation 2	88.3	88.9	88.2
Validation 3	95.0	88.9	96.1

DISCUSSION

Results are of similar quality like in Beneš and Hampel (2016). In that paper, total accuracy of prediction based on testing dataset is 88.3% and ranges from 96.7% to 100.0% for validation datasets. The main difference lies in the fact that in compared paper there is only one measured variable and in logit model this variable occurs in the cubic polynomial function form. In present paper is estimated multidimensional model with quadratic polynomial maximally. Theoretically, asymmetries requiring cubic polynomial can be present also in our case, but in multivariate model there is no space for the third powers of variables.

Predictions based on decision tree estimated by classification tree model can be compared to expert approach, where device pass the test when 5.15 < *Fdist* < 5.45, *Par* < 0.3 and 0.4 < *Dcri* < 0.5. With these settings predictions are provided and their accuracy is evaluated: for entire dataset, all passed cases are correctly estimated, but almost all failed cases are predicted as passed. It is visible, that expert approach underestimates effect of deviations from ideal measurements on damage of the device. Ranges of parameters given by classification tree model can warn experts about this.

Although the performed analysis shows the possibility to replace destructive test by non-destructive one, it is necessary to point out the limits of this analysis. The most serious limit is the fact, that because medical manufacture is strongly regulated industry, it is necessary to check whether the resulting procedure is in accordance with the regulations. Probably some "safety adjustment" given by the expert way will be provided to eliminate the production of defective pieces. Although representative batches of products have been selected for analysis, it will be necessary from time to time to verify the parameters found for a non-destructive test by performing a destructive test. The reason may be wearing of production machines, change in the composition of the material for production and other minor influences.

Performed analysis has important economic aspects. In general, such consequences will be very case specific. However, it is possible to evaluate the data of the real business case as is provided in the following text. The main reason for this alternative testing is to reduce test costs necessary for process control during regular production without any compromise in quality of the product. The situation where the test results are coming from is rather typical: high volume of annual production with rather regular price for one piece. The annual production of such product is approximately 1 million parts with the price €1.70 for each.

The current practice is based on the process validation results where the process was under the control with acceptable process capability Cpk of 1.55, but still with rather significant variability

Payback calculation
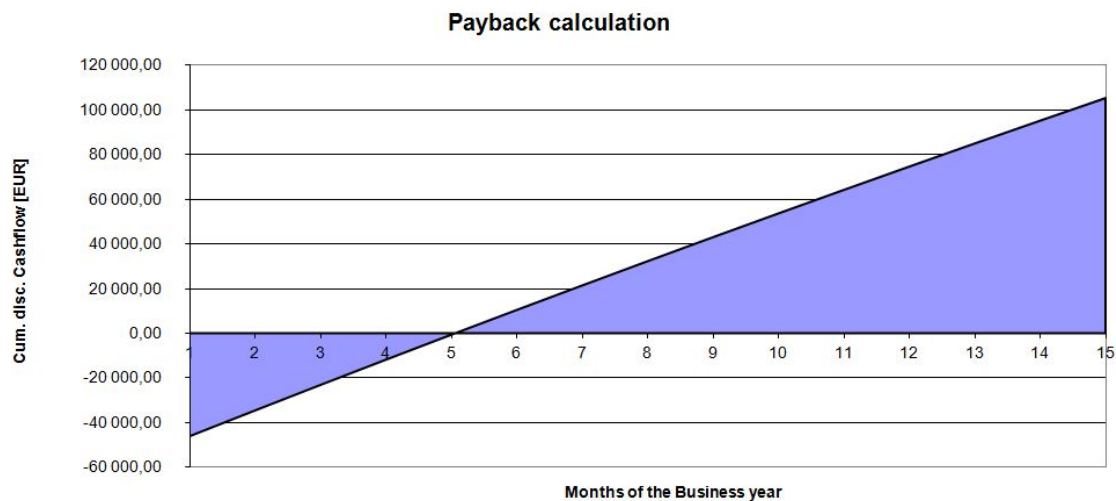
3: *Calculated return of investment curve*

expressed by standard deviation. Therefore the process control is settled as destructive testing and the sample size is determined using the AQL tool as described by the ANSI/ASQ standard Z1.9 (ANSI/ASQ, 2018). The typical batch size contains 250 pieces therefore based on a potential risk, 8 pieces should be overproduced to have 250 acceptable parts in the end. Overproduction of 8 pieces on every 250 pieces means extra 3.2% for each manufacturing batch.

It means that to have 1 million of acceptable parts ready, it is necessary to produce 32,000 pieces above. If the price of each is €1.70 it means existence of scrap material in price of €54,400 per year because of the required destructive testing. But this is the direct price only. Mentioned facility is producing complete working elements of three different types where the price of the cheapest one is approximately €3, the middle one is about €5 and the most expensive one cost as much as €11. The cheapest one is the most common and makes about 60% of annual turnover, while the middle one is about 30% and the most expensive about 10%. Those numbers are only rough estimation based on the data from last 3 years and they are changing. The changes show that the portion of more expensive types is growing, so the data can be used from the past as a worst case because if the trending will stay as it was documented, the impact of discussed superseding destructive testing by non-destructive ones supported by statistical evaluation will be higher.

The 32,000 pieces mentioned above represent 19,200 of the cheapest working elements, 9,600 of the middle priced ones and 3,200 of the most expensive ones. Expressed in its value it means €35,200 on the most expensive working elements, €48,000 on the middle-priced ones and €57,600 on the cheapest ones. All together means €140,000 per year. Considering this number, it is possible to calculate the Return of Investment (ROI) curve to see the possible impact on facility Profits and Losses calculation, see Fig. 3. As a cost of mentioned testing

it is considered €1,836 direct cost of used material and estimated €44,320 indirect cost of labour etc. This results in €46,156 as the worst case.

The important thing is that in this case it is adequate to speak about the business situation "sales of what can produce" not "produce what can sell". The manufacturing site is the member of the corporation and the working elements are the bottleneck of complete HF resection electrodes line. The more working elements will be produced, the more complete electrodes will be sold out. The average price of a complete electrode is about €615 for each so for a corporation those 19.200 electrodes mean increase of turnover of €11,808,000. The margin is not available in this case so it is not possible to calculate difference in the profit. Still, beside of the additional turnover it definitely means additional satisfied physicians and patients, in the end.

Currently it is possible to evaluate the usability of this tool during another business case: manufacturing of worldwide unique GIT biodegradable stents where the situation is completely different. The price is much more stable, and forecast is about annual decrease of 2% only. In the same time this product is manufactured on individual order therefore the standard process validation is not possible. It is necessary to go for manufacturing of two identical pieces, where one of them will be tested by destructive testing and the second will be sent to customer. In such situation it makes even better sense despite small annual production.

However, the total amount of pieces produced during estimated life cycle five years will be very probably below 2,500 pieces only, the economic effect is still high due to high price of sole product. With safety factor mentioned above this situation shows the payback moment after little bit more than two years and the complete saving before the end of life cycle is nearly 50% of calculated turnover. Similarly, without safety factor it would be even more promising.

In this case, a company is currently in the stage of trying to find the proper characteristic to set up adequate function. There is common understanding that for degradation process of such material the residual water content is critical and should be above the limit. Actually, several engineering studies is running to find the characteristic, easy to be measured by non-destructive testing.

The economic aspect should be evaluated very carefully for each situation specifically starting with ROI calculation. Whether outlined approach will be appropriate for manufacturers or not, depends on the specific situation of each other. Suggested way of non-destructive testing cannot be the universal solution for any kind of testing during the production.

CONCLUSION

It is possible to conclude that relationship between binomial variable the component is destructed, and measured variables Fork distance, Parallelism and Depth of crimping was found based on the dataset dedicated for estimation. This means that it is possible to set up these measurements as adequate for described process control. Adequate description is given by quadratic polynomial maximally. Stability of this relationship was judged by validation datasets. Performed analysis shows that it is possible to replace destructive test by non-destructive one on such confidence level. The acceptability of this confidence level should come from risk analysis of such product and discussed characteristic.

Finally, proposed analysis can be seen as an effective and viable way how to avoid destructive testing. From the economic point of view, attempts to find alternative test method make sense in the case of high-volume production or with high price of one unit together with verification or high-volume process control sampling. Calculation of effectiveness should be executed any time but in these situations, the initial cost of alternative test method validation will be paid back in acceptable time and increase the competitiveness on the market.

REFERENCES

AGRESTI, A. 2007. *An Introduction to Categorical Data Analysis.* John Wiley & Sons Inc.

ANSI/ASQ. 2018. ANSI/ASQ Z1.9-2003 (R2018): Sampling Procedures and Tables for Inspection by Variables for Percent Nonconforming. *ASQ.* [Online]. Available at: https://asq.org/quality-resources/z14-z19 [Accessed: 2020, June 11].

ASTM. 2017. *ASTM E3080-17, Standard Practice for Regression Analysis*. West Conshohocken, PA: ASTM International.

ASTM. 2018. *ASTM E2586-18, Standard Practice for Calculating and Using Basic Statistics*. West Conshohocken, PA: ASTM International.

BENEŠ, O. and HAMPEL, D. 2016. Evaluation of the destructive test in medical devices manufacturing. In: *Proceedings of Mathematical Methods in Economics*. September 6th–9th, 2016, Liberec, Czech Republic, pp. 37–42.

BREIMAN, L., FRIEDMAN, J., OLSHEN, R. and STONE, C. 2017. *Classification and Regression Trees*. Chapman and Hall/CRC Press.

CLEMENS, N. 2018. The European Medical Device Regulation 2017/745/EU: Changes and Impact on Stakeholders. *Journal of Clinical Research Best Practices,* 14(9): 1–7.

KLEPÁČ, V. and HAMPEL, D. 2017. Prediction of financial distress of agriculture companies in EU. *Agricultural Economics*, 63(8): 347–355.

MAREŠOVÁ, P., MOHELSKÁ, H., HONEGR, J. and ČERNÝ, M. 2018. Phases of Product Life Cycle in the Context of Economic Effectiveness and Medical Device Development. In: *Proceedings of IIER International Conference*. Paris, pp. 53–56.

STAŇKOVÁ, M. and HAMPEL, D. 2018. Bankruptcy Prediction of Engineering Companies in the EU Using Classification Methods. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 66(5): 1347–1356.

STAŇKOVÁ, M. and HAMPEL, D. 2019. Bankruptcy Prediction Based on Data Envelopment Analysis. In: *Mathematical Methods in Economics 2019: Conference Proceedings*. September 11–13, 2019, České Budějovice, pp. 31–36.

Contact information

Oldřich Beneš: xbenes19@mendelu.cz
David Hampel: david.hampel.uso@mendelu.cz