Using data science to find predictors of adverse regulatory inspections - Enhancing quality risk management in the pharmaceutical industry

D I S S E R T A T I O N
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Management

submitted by

Steffen Eich

from

Germany

Approved on the application of

Prof. Dr. Thomas Friedli

and

Prof. Dr. Dietmar Grichnik

Dissertation No. 5171

Difo-Druck GmbH, Untersiemau 2021

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science, hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, October 29, 2021

The President:

Prof. Dr. Bernhard Ehrenzeller

Danksagung

Die vorliegende Dissertation entstand während meiner Zeit als wissenschaftlicher Mitarbeiter im Bereich Produktionsmanagement am Institut für Technologiemanagement an der Universität St.Gallen. In den drei Jahren als wissenschaftlicher Mitarbeiter sammelte ich in der Gruppe Operational Excellence Projekterfahrung mit Branchenführern der globalen Pharma-industrie. In dissertationsbezogenen Industrieprojekten und Industrie-Austauschplattformen konnte ich einerseits die Empirie für diese Arbeit sammeln und zudem die Zukunft der Branche mitgestalten.

Die internationale Ausrichtung der Projektarbeit war besonders gewinnbringend. Ein Netzwerk von Kooperationen mit einem Dutzend Unternehmen, mehreren Forschungsinstituten und der US Food and Drug Administration ermöglichte mir seltenen Einblick, der diese Dissertation bereichert hat.

Mein erster Dank geht an meinen Doktorvater Prof. Dr. Thomas Friedli. Sein Zuspruch für meine Idee, innovative Datenanalyse in einem neuen Anwendungsfeld zu versuchen, war die Grundlage für mein Dissertations-projekt. Für die Bereitschaft, das Korreferat meiner Dissertation zu übernehmen, danke ich Prof. Dr. Dietmar Grichnik. Seine Aufgeschlossenheit gegenüber meinem Forschungsvorhaben motivierte mich während des gesamten Forschungsprozesses.

Meinen Kollegen im Bereich für Produktionsmanagement danke ich für die hervorragende Zusammenarbeit, die wertvollen Diskussionen und die freundschaftliche Atmosphäre. Gemeinschaftlich konnten wir viele Herausforderungen meistern. Besonders danke ich dem Praxispartner für die Begleitung der Fallstudie und das Vertrauen in dieses Projekt.

Ein herzlicher Dank gilt meiner Familie und besonders meinen Eltern für die Ermöglichung und freie Wahl meiner Ausbildung und ihrem Rückhalt in schwierigen Zeiten. Zu guter Letzt möchte ich meiner Partnerin Kim danken, die mich stets unterstützt und viel Verständnis für dieses Projekt aufgebracht hat. Die gemeinsame Zeit und tollen Erlebnisse neben der Dissertation werde ich immer in bester Erinnerung behalten.

St. Gallen, Januar 2021 Steffen Eich

Summary

Quality risk in the domain of drug production is a key factor contributing to patient safety and is critical for business success. Today, there is a lack of guidelines for its management at the industrial level. In recent years the limelight has been on the quality metrics draft guidance created by the United States Food and Drug Administration (FDA), as well as the industry-wide efforts on gathering, and analyzing operational data. Both trends have incited increased interest in the quantitative assessment of quality risk.

Several scholars contributed to the discussion on two levels of abstraction. Guidance on Quality Risk Management in the pharmaceutical industry is typically focused on the micro-level of designing products and processes. On the other hand, Operations Management scholars provided research on the econometrical level. To the best of the author's knowledge, there is no published research yet that examines site internal operational characteristics, and their relationship to quality risk in terms of quality compliance evaluations.

The main goal of this research is to develop a data-driven management approach for Quality Risk Management at the level of the manufacturing site. In order to achieve the research objective, four major steps were followed. First, a literature review of the known interlinks of inspection outcomes and characteristics of manufacturing operations was performed. Secondly, the relationship between Operational Excellence and quality risk is evaluated. Thirdly, the operational performance indicators were analyzed with regard to their possible link to inspection outcomes. And lastly, the applicability of predictive modeling in this Quality Risk Management context is tested and discussed.

This research project is designed to follow a mixed-method approach combining qualitative and quantitative research, as well as the solution prototyping phase. An iterative process enables establishing a first-hand understanding, preparing credible analysis results, and improving upon it to get a thorough understanding of the issue. Employing a systems theory view enables the assessment of quality risk in manufacturing sites as an analytically measured system's state. Following the design science research methodology allows developing analytics methodologies, which will in turn facilitate and provide support to the decision-making process of the industrial managers.

Zusammenfassung

Qualitätsrisiken in der Arzneimittelproduktion sind ein Schlüsselfaktor für die Patientensicherheit und entscheidend für den Geschäftserfolg. Aktuell gibt es kaum Handreichungen für den Umgang mit diesem Risiko auf Werksebene. Jüngst erweckten neu entworfene Richtlinien der United States Food and Drug Administration zu Qualitätskennzahlen und ein branchenweiter Trend zur Sammlung und Analyse von Betriebsdaten das Interesse an einer quantitativen Bewertung des Qualitätsrisikos.

Der wissenschaftliche Diskurs findet aktuell auf zwei Abstraktionsebenen statt. Das Management von Qualitätsrisiken in der pharmazeutischen Industrie konzentriert sich in der Regel auf die Mikroebene der Gestaltung von Produkten und Prozessen. Die andere Ebene bilden Studien zum Qualitätsrisiko von Produktionsstandorten ab, die ökonometrisch geprägt sind. Nach bestem Wissen gibt es bisher keine Veröffentlichung, die Betriebsmerkmale in ihrem Zusammenhang mit behördlichen Inspektionen von Qualitätsstandards untersucht.

Das Hauptziel dieser Forschungsarbeit ist daher die Entwicklung eines datengesteuerten Managementansatzes für Qualitätsrisiken auf Ebene der Produktionsstandorte. Zunächst befasst sich der Literaturüberblick mit den bekannten Zusammenhängen von Inspektionsergebnissen und Merkmalen von Herstellungsbetrieben. Zweitens werden organisatorische und systemische Merkmale des Operational Excellence Reifegrades von Produktionsstätten und Inspektionsergebnisse quantitativ untersuchen. Drittens werden operative Kennzahlen der Produktion im Hinblick auf ihren möglichen Zusammenhang mit Inspektionsergebnissen analysiert. Schließlich werden die genannten Daten und darauf basierende prädiktive Algorithmen hinsichtlich ihrer Verwendbarkeit für die Entscheidungsfindung im Qualitätsrisikomanagement diskutiert.

Dieses Forschungsprojekt folgt einem «mixed methods» Ansatz, der qualitative und quantitative Forschung kombiniert. Die Anwendung einer systemtheoretischen Sichtweise ermöglicht die Erfassung des Qualitätsrisikos in Produktionsstätten als analytisch gemessener Systemzustand. In einem iterativen Prozess wird iterativ Verständnis aufgebaut, getestet und in Praxisnutzen umgemünzt. Letztlich bietet die Arbeit konkrete Handlungsanleitung für die Praxis zur Weiterentwicklung der analytischen Fähigkeiten in der Industrie.

Table of Contents

Table of Contents I
List of Figures V
List of Tables VII
List of Abbreviations IX
1. Introduction 1
1.1 Practical Relevance 1
1.2 Theoretical Relevance 5
1.3 Research Design 6
1.3.1 Conceptual Background 7
1.3.2 Research Theory 8
1.3.3 Research Questions 9
1.3.4 Research Methodology & Process 10
1.4 Thesis Structure 15
2. State of Research 17
2.1 Quality Management 17
2.2 Operational Excellence 20
2.2.1 World Class Manufacturing 20
2.2.2 Lean Production 21
2.2.3 Operational Excellence as the cumulation of an improvement journey 22
2.2.4 Operational Excellence in the Pharmaceutical Industry 23
2.3 Quality Risk Management 24
2.3.1 Quality Risk Management in the Pharmaceutical Industry 25
2.3.2 Quality Risk research using inspection outcomes 29
2.4 Data Analytics in Operations Management 34
2.5 Research Gaps 36
3. Quality Risk Predictors 37
3.1 Manufacturing site OPEX maturity analyses 39
3.1.1 Hypotheses 39
3.1.2 Empirical Set-up 40
3.1.3 Methodology 45
3.1.4 Results 45
3.1.5 Conclusion 47
3.2 Manufacturing site performance metrics 48
3.2.1 Hypotheses 48
3.2.2 Empirical Set-up 50
3.2.3 Methodology 52

3.2.4 Results 52
3.2.5 Conclusion 58
3.3 QC Lab OPEX maturity analyses 58
3.3.1 Hypotheses 58
3.3.2 Empirical Set-up 59
3.3.3 Methodology 62
3.3.4 Results 62
3.3.5 Conclusion 65
3.4 QC Lab performance metrics analyses 65
3.4.1 Hypotheses 66
3.4.2 Empirical Set-up 66
3.4.3 Methodology 68
3.4.4 Results 68
3.4.5 Conclusion 69
3.5 Consolidated indicator metrics for regulatory inspection outcomes 69
4. Industry project case study with predictive analyses 71
4.1 Introduction to case study industry partner 71
4.2 Project setting 72
4.3 Dataset 74
4.3.1 Regulatory inspections and internal GMP audits 74
4.3.2 Operational performance data 80
4.4 Descriptive analyses 81
4.4.1 Methodology 81
4.4.2 Results 82
4.4.3 Conclusion and Discussion 86
4.5 Predictive Modelling 87
4.5.1 Introduction 87
4.5.2 Methodology 87
4.5.3 Prediction of numeric inspection scores 91
4.5.4 Prediction of binary inspection outcomes 98
4.5.5 Conclusion and Discussion 106
4.6 Project impact and future quality audit planning at the industry partner 109
4.7 Validation of findings with industry experts 111
4.7.1 Discussion with industry representatives 111
4.7.2 The perspective of the regulators 112
5. Contributions to Theory and Practice 115
5.1 Implications for practitioners in Quality Management and regulators 115
5.1.1 Establishing a quality compliance score for tracking inspection outcomes 115
5.1.2 A good state of quality compliance can deteriorate over time 116
5.1.3 The importance of Quality Management practices for GMP compliance 116
5.1.4 Stable, in-control manufacturing processes benefit quality compliance 117

5.1.5 Creating a central hub for quality data analytics is the key 117
5.2 Theoretical Contribution 117
5.3 Practical Contribution 118
5.4 Limitations and outlook 119
6. Declaration of funding and conflict of interest 121
7. References 123
A Appendix 137
A.1 Case interview guidelines 137
A.2 R Code: Analyses for Chapter 3 146
A.2.1 Operational Excellence benchmarking database 146
A.2.2 QC Lab benchmarking database 156
A.3 R Code: Analyses for Chapter 4 169
A.3.1 Prediction models with numerical target score 183
A.3.2 Prediction models with binary target 194
Curriculum Vitae 214

List of Figures

Figure 1: Three lines of defense for Quality Risk Management.....3
Figure 2: Criteria for conducting Design Science Research 11
Figure 3: Stepwise approach to technological research 12
Figure 4: Iterative research process by *Kubicek, 1976*.....13
Figure 5: Research questions and proposed resources 14
Figure 6: Conceptual framework.....14
Figure 7: Thesis structure 16
Figure 8: Quality concepts over time 18
Figure 9: Structure of Analyses based on OPEX benchmarking program.....38
Figure 10: Structure of analyses based on QC lab benchmarking program....38
Figure 11: Geographical distribution of the manufacturing site datasets.....41
Figure 12: St. Gallen Operational Excellence Model 42
Figure 13: Enabler groups for manufacturing sites.....43
Figure 14: Matched FDA inspection outcomes for manufacturing database ...44
Figure 15: St.Gallen Pharmaceutical Production System Model 51
Figure 16: Overview and inter-correlation of identified manufacturing KPIs....56
Figure 17: Geographical distribution of the available QC lab datasets.....59
Figure 18: Enabler groups in QC labs 60
Figure 19: Matched FDA inspection outcomes for QC lab database.....60
Figure 20: Analysis of inspection citations and their connection to QC labs ...61
Figure 21: Distribution of lab-related citations in FDA inspections 62
Figure 22: Cumulative number of inspections of different agencies by site.....75
Figure 23: Number of findings per inspection by agency.....76
Figure 24: Distribution of inspection scores in the sample 79
Figure 25: Distribution of inspection outcomes as binary measure 80
Figure 26: Product types of site data records by inspection score 82
Figure 27: Box plot distribution of prediction target current inspection score ..88
Figure 28: Illustration of ROC and AUC.....90
Figure 29: Regression tree model visualization.....96
Figure 30: Neural network model visualization 97
Figure 31: Root-mean-square-error summary chart 98
Figure 32: ROC charts for logistic regression models 102
Figure 33: ROC charts for decision tree models.....103
Figure 34: Variable importance in random forest prediction 104
Figure 35: Area-under-the-curve summary chart.....106

Figure 36: Benchmark of achieved prediction performance AUC.....107
Figure 37: Adverse inspection risk ranking prediction 109
Figure 38: Maturity model for the integration of Data Analytics in QRM.....110

List of Tables

Table 1: Research questions 9
Table 2: Recent publications on pharmaceutical QRM 27
Table 3: Related literature focusing on Quality Risk on the plant level 32
Table 4: Logit regression output for overall OPEX implementation 46
Table 5: Logit regression output for OPEX practice categories 46
Table 6: Logit regression output for Total Quality Management 47
Table 7: Operational Excellence performance scores and assigned metrics. .50
Table 8: Logit regression output for OTIF delivery performance 53
Table 9: Logit regression output for Lot Acceptance Rate 53
Table 10: Logit regression output for Customer Complaint Rate 54
Table 11: Logit regression output for Customer Complaint Rate with control .55
Table 12: Logit regression output for combined variables 56
Table 13: Logit regression output for combined variables with control 57
Table 14: Logit regression output for lab-specific enablers 63
Table 15: Linear regression output for lab quality enablers 64
Table 16: QC lab operational excellence performance metrics and scores 67
Table 17: Linear regression output for IOOS proportions 68
Table 18: T-Tests for Lot Acceptance Rate and current inspection score 83
Table 19: T-Tests for Right First-Time Rate and current inspection score 83
Table 20: T-Tests for Customer Complaint Rate and current inspection 84
Table 21: T-Tests for previous quality audit and current inspection score 85
Table 22: T-Test for previous and current inspection score 85
Table 23: Strengths and weaknesses of Linear Regression models 92
Table 24: Strengths and weaknesses of decision tree models for prediction ..93
Table 25: Strengths and weaknesses of neural networks for prediction 94
Table 26: Tested configurations of neural networks 94
Table 27: Numeric prediction errors for linear regression 95
Table 28: Numeric prediction errors for regression trees 95
Table 29: Numeric prediction errors for neural network models 96
Table 30: Strengths and weaknesses of k-NN models 99
Table 31: Strengths and weaknesses of Random Forest models 100
Table 32: Binary prediction performance for logistic regression models 101
Table 33: Binary prediction performance for decision tree models 102
Table 34: Binary prediction performance for random forest models 103
Table 35 Binary prediction performance for k-nearest neighbor models 105

Table 36: Binary prediction performance for neural network models 105

List of Abbreviations

*AI	Any Action Indicated (Aggregation of OAI and VAI inspection results)
6σ	Six Sigma
BE	Basic Elements (of Operational Excellence)
BM	Benchmarking
CDER	Center for Drug Evaluation and Research (of the FDA)
cf.	confer
CM(O)	Contract Manufacturer / Contract Manufacturing Operator
DP	Drug Product
DSR	Design Science Research
DV	Dependent variable
e.g.	exempli gratia
EMA	European Medicines Agency
EMS	Effective Management System
FDA	U.S. Food and Drug Administration
FMEA	Failure-Mode-and-Effects-Analysis
FTE	Full-time equivalent
GMP	Good Manufacturing Practices
ICH	International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use
i.e.	id est
IOOS	Invalidated-out-of-specification(-rate)
IV	Independent variable
JIT	Just in Time
k-NN	k-nearest neighbor (algorithm for prediction)
KPI	Key Performance Indicator
LAR	Lot Acceptance Rate

MAE	Mean Average Error
MRQ	Main research question
NAI	No Action Indicated (FDA inspection outcome)
OAI	Official Action Indicated (FDA inspection outcome)
OM	Operations Management
OPEX	Operational Excellence
OTIF	On-Time In Full delivery rate
PAC	Post-approval change
PPSM	Pharmaceutical Production System Model
PQS	Pharmaceutical Quality System
QM	Quality Management
QRM	Quality Risk Management
Q9	The Statute on Quality Risk Management released by ICH
RMSE	Root Mean Square Error
SME	Subject matter expert
SPC	Statistical Process Control
SRQ	Sub-research-question
TPM	Total Productive Maintenance
TQM	Total Quality Management
VAI	Voluntary Action Indicated (FDA inspection outcome)
WIP	Work in progress

1. Introduction

This chapter highlights the practical need within industry and regulators for research in the Quality Risk Management domain, as well as the relevant literature in Section 1.1. Next, recent trends in research publications and key scholars are emphasized (Section 1.2). Based on this, Section 1.3 provides details on the research design. The chapter concludes with the outline of the objectives of this research project.

1.1 Practical Relevance

The human race has evolved and pushed the boundaries of science and knowledge tremendously in the last century. The development of modern medicine and pharmaceutics had a most significant impact on peoples' lives. In the first half of the 20th century, the invention of penicillin, aspirin, or the concept of vaccination alleviated human suffering improving the quality of human life significantly. These achievements resulted in reduced mortality and considerably increased life expectancies. Today, pharmacotherapy is the most frequently used, versatile, and often price-effective form of therapy. This has been enabled particularly by the industrialized production of drugs (Banker and Rhodes, 2002).

The quality requirements for pharmaceuticals are distinctly different from other consumer-based goods owing to three factors. First, consumers or patients can often find themselves in situations where their well-being or their life depends on the quality of the product. Secondly, a layperson does not possess the acumen to assess the quality of a therapeutic drug before or during consumption. Lastly, medicine is generally prescribed by doctors or pharmacists to patients providing the consumer with no or little choice to choose between product alternatives. These aspects put greater responsibility on drug manufacturers and government regulators to ensure minimal risk to the patients. (Banker and Rhodes, 2002)

Numerous incidents or events can negatively affect product quality during the manufacturing process. In practice, industry and regulators are generally quite effective in preventing faulty products from reaching the market to negatively impact the patients. However, the industry often struggles to supply a replacement in place of such disqualified pharmaceuticals. One example of this is the case of antiviral "Viracept" manufactured by Roche to treat HIV infections.

“Viracept” represents a breakthrough in therapeutics against disease progression of HIV, by virtue of its effectiveness and global affordability. However, in 2007, an error in the manufacturing procedure led to the production of an impure product unsuitable for human use. Large amounts of the drug had to be recalled and destroyed. This led to a global shortage of the drug that had fatal clinical outcomes (Johnston and Holt, 2014). Similarly, in 2009 a viral contamination in a Genzyme production facility affected the production of two drugs used in the treatment of rare genetic lysosomal storage disorders. With no alternative medication in the market, the resulting drug shortage caused several deaths (Pollack, 2010). After the anomaly was discovered, significant time was required to resolve the issue, and the drug shortage lasted for as long as seven years after the original incident (Waldron, 2017).

Today, drug shortages and product recalls occur frequently and affect companies, markets, and patients continuously. Quality issues account for 62% of these shortages, as documented recently by the FDA’s Center for Drug Evaluation and Research (FDA - CDER, 2019a). In recent years, the pharmaceutical market in the U.S. has also faced rising numbers of drug recalls (Hall *et al.*, 2016). This development explicitly also applies to Class 1 recalls, which categorizes violative drugs with the most severe adverse health consequences (O’Connor, Yu, and Lee, 2016). In addition to the risk to patients, such recalls also have an extensive negative economic impact on the affected companies (Shah, Ball, and Netessine, 2017).

Drug manufacturers and regulators are on an ongoing quest to improve stability and control of production, and supply safety. To achieve high standards of pharmaceutical supply safety, industry and government agencies act on different aspects and levels of abstraction of quality risk which is represented as a framework in Figure 1. In re-imagining the established model from financial compliance (The Institute of Internal Auditors, 2013), one can perceive three lines of defense designed to preserve the pharmaceutical manufacturing quality. The first line of defense is the product and process design. The goal of this design is to maintain the quality of drugs and their delivery system, while also ensuring that their production meets the quality standards. The above-mentioned points can be potentially achieved through efficient quality control units, which operate independently of the manufacturing unit and typically test all the products. The second line of defense is the company’s own internal quality assurance unit and quality audit function. Their task is to maintain and

review adherence to the set quality standards. Unlike the financial compliance model, the third line of defense lies outside the company. National authorities approve drugs for launch into their respective markets and routinely check the manufacturing and distribution sites for their adherence to the quality standards. The industry commonly refers to such standards as *Good Manufacturing Practices GMP* (Nally, 2007).

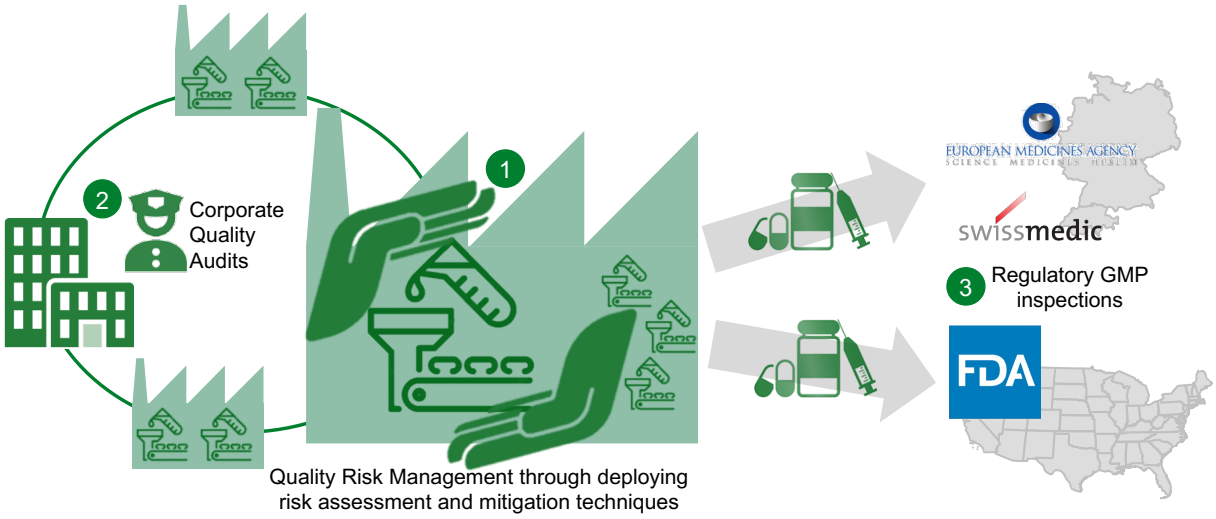


Figure 1: Three lines of defense for Quality Risk Management shown as a flow chart for the pharmaceutical industry

The United States Food and Drug Administration is a leader in applying innovative methods to fulfill its duty of surveilling manufacturing facilities supplying pharmaceutical products to the United States. The globalization of pharmaceutical manufacturing during the last years and decades, as well as segmentation of value chains across several facilities and companies, has increased the complexity of this task (Gottlieb, 2018). Therefore, the FDA has deviated from a regular inspection routine that required inspectors to visit every site in a specified time interval, and rather adopted a risk-based inspection planning approach. The rationale for that decision was that *"resource limitations prevent uniformly intensive coverage of all pharmaceutical products and production"* (FDA - CDER, 2002). The factors in the risk model include product, type of plant and its capacity, on-site process complexity, and past quality compliance data, which includes product recall history, as well as prior records of quality compliance of the manufacturing plant (FDA - CDER, 2004). In 2015, the FDA announced its plan to gather operational performance indicators from

the manufacturers to include them in and improve the risk-based inspection planning method through the "*Quality Metrics*" program (FDA - CDER, 2015). The Quality Metrics could facilitate a strategic change from solely monitoring structural and historical information to transform it into a more proactive approach. Analogously, Medina & Richmond (2015, p. 6) added that "*leading metrics such as process performance measures [might] provide an early warning of quality problems*".

The FDA initiative was met with strong criticism by the pharmaceutical industry. The important critique by various companies and industry-based associations as recorded in their statements (Mezher, 2015; Stephen Barlas, 2017; van Arnum, 2017) center around the following:

- Doubts regarding the link between the proposed metrics and the risk level of the sites.
- The work burden required for reporting the data could be rather used for quality improvements in the companies.
- Doubts regarding the legitimation of the program if it is made de-facto mandatory and its compatibility with current legislation.

The required effort is likely to decrease with advancements in the digitization of production, and homogenization of company data and system structures. The legitimation or necessary authority of the agency may be evaluated by professional legal scholars and clarified by the legislator. The relationship between such metrics and quality risk certainly appealed to the regulatory decision-makers but has yet been explored scientifically.

Similarly, managers in the corporate quality functions within the company are tasked to deploy quality auditors to sites and to focus corporate attention and resources on the most relevant quality risk issues. As first shown by Mänder, there is some evidence of the quality compliance status of production sites being related to the measurable state of its quality system (Mänder and Friedli, 2018). At the same time, quality executives from the pharmaceutical industry are less confident that the quality indicators will help steer operations in the right direction (Torkko *et al.*, 2013). O'Connor et al. concluded their study stating that predictive models can efficiently improve quality risk management (O'Connor *et al.*, 2017). Therefore, insight around the relationship of operational characteristics, performance, and quality risk has to be broadened. This thesis aims to develop

a proof of concept to show that such data can be used to establish a more actionable basis for the decision-making process.

1.2 Theoretical Relevance

The risk perspective on manufacturing quality is a relatively new research area that is mainly applied in the domain of pharmaceutical drug manufacturing. Since then, scholars from academia, authorities, and industrial associations have published their research on quality risk and its management.

Researchers started to apply a risk perspective to supply chain management in the early 2000s. Jüttner & Christopher highlighted, *"developing risk management approaches for specific supply chains and/or industries"* and *"approaches helping managers to track the vulnerabilities"* as demands for normative research (Jüttner & Christopher, 2003, p. 23). In a more recent literature review, Ho et al. found that *"the area of manufacturing or process risk has not received much attention"* and therefore shall be a research priority. Furthermore, they recommend *"developing an early warning monitoring system with adaptive risk indicators for various types of supply chains and validating the system empirically"* (Ho et al., p. 5061).

A majority of the current research on Quality Risk Management in the pharmaceutical industry focuses on the first line of defense as shown in Figure 1. It is mainly characterized by ICH's[1] Q9 guidelines and their described methods. Few examples of Quality Risk Management in the pharmaceutical industry existed before the release of these guidelines by an international group of experts in 2005. The ICH Q9 statute addresses all aspects of quality management, ranging from development to manufacturing and distribution, including audits and inspections (ICH, 2005). However, most of the published academic literature focusses on eliminating risks on the shop-floor, designing quality into processes and guidance for their application, and advancement of the associated methods such as Failure Mode and Effect Analyses (FMEA), etc. (O'Donnell and Greene, 2007; Woodcock *et al.*, 2012; Claycamp, 2015; Ramnarine, 2015; Ramnarine *et al.*, 2017; Greene, O'Donnell and Calnan, 2018). Besides these qualitative or quasi quantitative methods, ICH Q9 also

[1] ICH stands for the *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*

suggests the use of *"statistical tools"* (ICH, 2005, p. 14). However, academic research literature sparingly follows the above-mentioned recommendation

FDA inspection outcomes have often been used as a measure of quality risk in a manufacturing site or analyzed in various other contexts at the industrial level. Significant effects on quality risk have been shown for offshore and outsourced manufacturing (Gray, Roth, and Leiblein, 2011; Gray, Roth and Tomlin, 2016), cultural and language differences (Gray and Massimino, 2014), co-location with R&D facilities (Gray, Siemsen, and Vasudeva, 2015), ISO 9000 certification, and financial compliance patterns (Gray, Anand, and Roth, 2015; Altamuro, Gray, and Zhang, 2017). While these are valuable perspectives, they offer little insight into the sites' operations on a day-to-day basis.

The medical device industry is similar to the pharmaceutical industry based on the fact that their manufacturing facilities are also inspected by the US FDA throughout the lifecycle of their products. In their examination of quality compliance inspections and product recalls, Ball et al. concluded that *"leading indicators of […] recalls are not well understood"* (Ball, Siemsen, and Shah, 2017, p. 549). This also applies not only to product recalls but also to adverse quality compliance events in general and the pharmaceutical industry. In another study regarding adverse quality events, it is especially "*critical to understand operational causes because managers face significant and often contradictory challenges*" (Shah, Ball, & Netessine, 2017, p. 2442).

In conclusion, quality risk in pharmaceutical production so far has been studied on two levels. There is a detailed perspective on methods and tools for handling and reducing quality risks in the process. Secondly, there are econometric studies that evaluate various factors concerning quality risk at the manufacturing plant level. However, there is a lack of literature addressing operational characteristics and performance in the context of plant inspection outcomes. A detailed breakdown of existing related research and gap in the research database is provided in Chapters 2.3.2 and 2.5.

1.3 Research Design

This chapter outlines the research approach followed in this dissertation project. Specifically, the conceptual background is introduced here. and the ties to management research theory are discussed. The chapter then introduces the research questions and the intended research plan.

1.3.1 Conceptual Background

As characterized by Ulrich, business studies are the teaching of the leadership of social institutions (Ulrich, 1984, p. 168f.), including industrial companies. Further, research in this domain is understood as an applied social science (Ulrich, 1984, p. 169, 199). In contrast to fundamental sciences, applied science tackles problems that stem from practice. As such, it begins with the investigation of issues whereby the actors possess insufficient knowledge about the situation (Ulrich, 1984, p. 172), and its mission is to provide them with the knowledge that allows for the implementation of science-based decision-making (Ulrich, 1984, p. 179). After all, the goal of applied science, contrary to fundamental sciences, is to implement and introduce change rather than simply observing it (Ulrich, 1982, p. 4). Practical problems never stem from a singular discipline, therefore applied research by definition must take an interdisciplinary approach to solve them (Ulrich, 1982, p. 4, 1984, p. 185). Finally, although practical relevance is the main scientific objective, business studies should not be "theory-less" teaching (Ulrich, 1984, p. 168f.), and its practical use should be highlighted (Ulrich, 1982, p. 8).

Based on these paradigms, Ulrich designed a research process with seven phases to generate knowledge relevant to practice and solutions which include the following: 1. recording and typification of practice-relevant problems, 2. recording and interpretation of problem-relevant theories and hypotheses of the empirical basic sciences, 3. recording and specification of problem-relevant procedures of the formal sciences, 4. recording and investigation of the relevant application context, 5. derivation of assessment criteria, design rules, and models, 6. checking the rules and models in the application context, 7. practical advice. The proximity of the research process to the practice of application required is not only apparent at the starting point and final stages of the research process (phases 1 and 7), but also in intermediate alignments (phases 4 and 6) which ensures the application of the knowledge generated in tackling practical problems with viable and feasible solutions (Ulrich, 1984, p. 192ff.).

This research follows Ulrich’s understanding of business studies as an applied social science in several ways. Its starting point is a practical problem that currently lacks a solid scientific understanding of the matter and aims to improve current practices. It is interdisciplinary as the topic touches different domains,

such as operations management, pharmaceutical engineering, and data science[2].

As a much more recent concept, Design Science Research (DSR) also embraces a practically oriented approach. It specifically offers the application of strategies to ameliorate problems in the domain of operations management and information systems (van Aken, Chandrasekaran, and Halman, 2016). Specifically, the focus on a practically value-generating research output characterizes a typical DSR approach, which aims to develop purposeful artifacts (Dresch, Lacerda, and Antunes Jr, 2015). The application of the Design Science Research approach applied in this work is probed in greater detail in Chapter 1.3.4.

1.3.2 Research Theory

Research theory is used to guide new research and serves as a reference for the interpretation and understanding of new findings throughout this dissertation project. This thesis aims to quantitatively assess the quality state of manufacturing sites and later quantitatively determine the managerial implications of the findings. Systems theory lays the appropriate foundation for this project since it allows for a comprehensive understanding of complex socio-technical constructs as a system with a set of properties, such as a manufacturing site. The aim is to achieve an integrative structuring abstraction of the essential influencing variables that allow to control and steer the system. This directly leads to the design of a model-like, system-technical reference framework. (Von Bertalanffy, 1968; Ulrich, 1984; Gassmann, Frankenberger and Sauer, 2016)

Systems theory bases on several concepts that apply to its original dimension in biology, but also technical, social, and socio-technical systems. A system is composed of multiple interrelated parts or components which can be assigned to different hierarchy levels. At the same time, the system is more than the sum of its parts, which means that it can only be described comprehensibly in totality. Systems have defined boundaries that can be physical and rigid or permeable. An often-used aspect in the systems theory is that systems are controlled in a

[2] This thesis adopts Cao's multidisciplinary view on data science as a "a cross-disciplinary body of knowledge" including "informatics, computing, communication, management, and decision-making" (Cao, 2018, p.35). It applies data oriented thinking and methodology to various domains outside traditional statistics.

feedback loop as typically used in cybernetics. Finally, systems seek multiple goals, which can be contradictory to one another such as the iron triangle of cost reduction, shorter times, and higher quality. The above-mentioned aspects lend themselves quite suitably to the application in management (Johnson, Kast and Rosenzweig, 1964).

The pharmaceutical manufacturing plant is a relatively closed system that processes incoming materials into intermediates or drug products. Its boundaries typically are defined by the physical dimensions of the facility. A pharmaceutical manufacturing site uses several control mechanisms that provide input to the feedback loops which include quality control, financial controlling, and quality compliance audits.

1.3.3 Research Questions

The main research questions to be answered within this thesis are driven by managerial perspective, and the practical need for organizations to improve on their capabilities in proactive Quality Risk Management. The specific demands for having a comprehensive understanding of this topic within the pharmaceutical industry, and the regulatory authorities have is outlined in Chapter 1.1. The systematic approach adopted to answer these questions has been divided into four sub-research-questions (SRQ) as presented in Table 1.

Table 1: Research questions

How can data analytics be used for predictive Quality Risk Management on a manufacturing plant level?		
Sub-research-questions	1	What are published indicators for plant operations that are related to quality compliance issues?
	2	What is the effect of Operational Excellence implementation on inspection outcomes?
	3	What operational performance indicators link to quality compliance issues?
	4	How can the identified indicators improve regulatory and/or corporate decision making in the context of quality compliance?

The first question shall summarize existing academic literature and the documented discussions between industry and regulators. It will enable placing the results of the other research questions in the greater context and evaluate the degree of innovativeness of the latter. Building on that, the second and third research questions aim at identifying new indicators by assessing company data possibly linked to quality compliance evaluations.

The second question focuses on the organizational context or specific production system characteristics of the manufacturing sites that are relatively static. In contrast, question three emphasizes the operational performance metrics that directly result from the day-to-day business. The stark contrast between the two questions is based on three factors. First, it allows maintenance of consistency along with the prominent distinction between enabler and performance measures established in the St. Gallen Operational Excellence benchmarking model (Friedli *et al.*, 2013). Secondly, performance metrics are widely available and used in the industry, but have not yet been analyzed in the context of quality compliance (cf. Chapter 1.2). Thirdly, such performance indicators have received special attention within the FDA Quality Metrics draft guidance (FDA - CDER, 2015, 2016), which sparked discussion among practitioners. Therefore, it is justified to examine these measures specifically through a designated research question.

The fourth research question is designed to turn the newly developed knowledge into practically feasible and viable advice for industrial application. Other than strictly consulting managerial decision-makers, this research aims to develop mathematical models that will help practitioners in the pharmaceutical industry and regulatory authorities to derive insights from rather complex abstruse data.

1.3.4 Research Methodology & Process

Finding new scientific knowledge requires the newly produced information to be credible, plausible, as well as relevant to a problem (Lee and Lings, 2008). The research must be designed in a non-ambiguous way that allows the readers to confidently accept the results (Greener, 2008). This thesis builds on an established methodology and process from the domain of information systems and business management. The selected approaches are introduced in the subsequent sections.

The research methodology followed in this thesis is Design Science Research (DSR). The successful application of the Design Science Research approach requires the fulfillment of several criteria (cf. Figure 2). Overall, the main goal is to address real-world problems and challenges, to rigorously address the research questions in the field, and develop viable practical solutions or *artifacts* (Hevner *et al.*, 2004; Dresch, Lacerda, and Antunes Jr, 2015). The methodology is also particularly suited for research in Operations Management which justifies its increasing popularity in the domain (Dresch, Pacheco Lacerda, and Cauchick-Miguel, 2019).

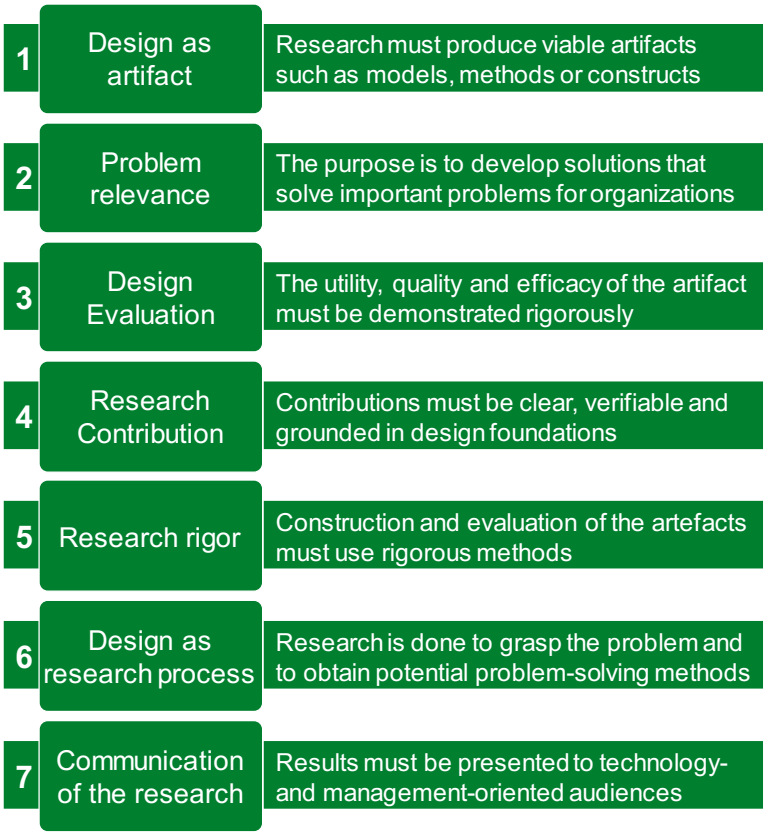


Figure 2: Criteria for conducting Design Science Research. Illustration adapted from Dresch et al., 2015.

The desired artifacts designed or produced in this research represent a decision support mechanism for quality risk assessment at the plant level which is based on data analysis. The study is based on verifiable analytics results that subsequently were used to solve the existing problem of measuring a manufacturing plant's state of quality risk. Any intermediary results were shared and discussed with an audience interested in the inter-disciplinary areas of

technology and management, such as fellow research scholars and industry partners.

The step-wise research process first described in connection with Design Science Research is presented in Figure 3. It begins with the researcher developing an understanding of the problem followed by attempting to solve the research problem. In case of the attempt fails, a new hypothesis or technique is employed for a subsequent draft solution. Once an exact or approximate solution is obtained, the solution is evaluated, and necessary corrective measures are taken. Such corrections can be accomplished through restarting the whole process, adapting hypotheses, or re-testing (Bunge, 1980).

It thereby offers several potential ways of iterating between the different research steps, which connects to the requirement of the artifacts being rigorously tested. Within this thesis, such tests were performed in two ways. First, the quality of the indicators found and the predictive power of algorithms was evaluated quantitatively. Secondly, the potential benefit for practical implication was challenged by discussions with industrial managers and regulatory policymakers. The chosen methodology combines quantitative and qualitative research methods.

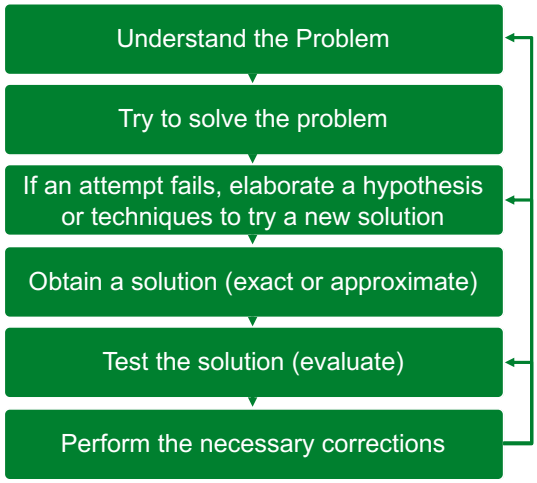


Figure 3: Stepwise approach to technological research according to Bunge, 1980. Illustration adapted from Dresch et al., 2015.

Described as a research methodology for technological research, there are several similarities of the research process associated with DSR to the pre-established approaches in management research, such as the iterative

research process described by Kubicek (1976). Applying an iteration-based research process through theoretical and empirical research phases is a well-proven concept in management research. This allows turning an explorative view of a new phenomenon into achieving a sound understanding of the phenomenon (Kubicek, 1976; Gassmann, 1999; Tomczack, 1992, cf. Figure 4).

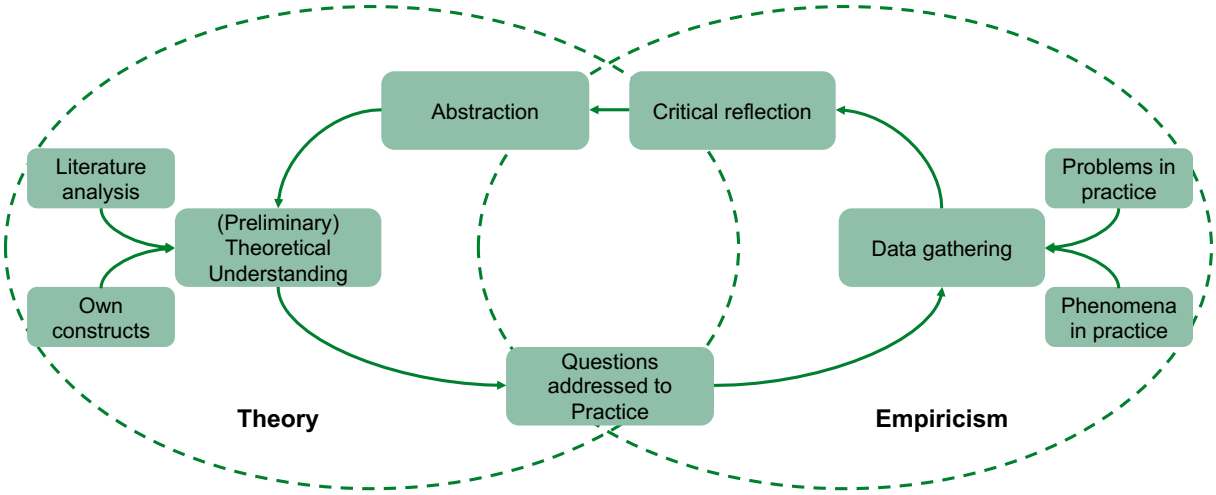


Figure 4: Iterative research process by *Kubicek, 1976*. Illustration adapted from Gassmann, 1999.

The phase of critical reflection and subsequent abstraction of a renewed theoretical understanding aligns well with the DSR process steps of evaluation and postulating new hypotheses. The application of data analytics in the domain of operations management is rarely studied and requires an exploratory research approach, with adequate opportunity for learning iterations. This research therefore will follow an iterative problem-solving approach based on an initial phase of gaining a more comprehensive understanding based on the pre-existing theory. It thereby fulfills the requirements of Design Science Research and the iterative research process.

The presented approach also hints at the different sources that inform the research. Figure 5 complements the proposed research questions with the methodologies that also implicate corresponding sources. The methods highlighted are literature review or analysis, quantitative analyses, and case studies that make use of expert interviews. It also highlights the inter-connectedness of the (sub-)research questions. SRQ 1 will inform the analyses for SRQ 2 and 3. These, in turn, will serve as the foundation for building decision support mechanisms for SRQ 4.

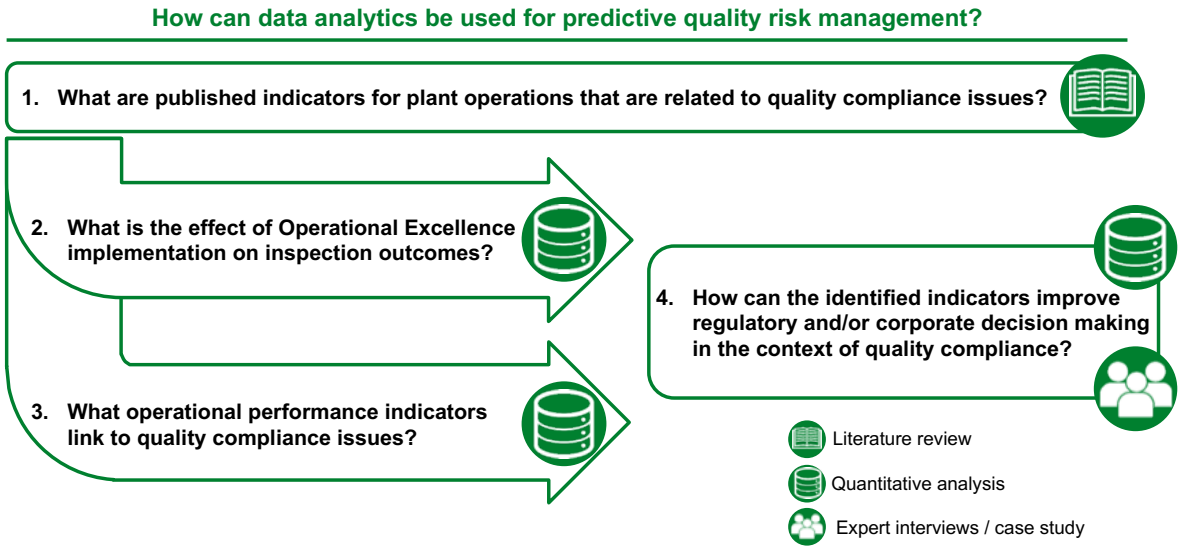


Figure 5: Research questions and proposed resources

As discussed, several iterations between sub-research question 4, and the sub-research questions 2 and 3 are intended. Factors tested for investigating SRQ 2 and 3 will be tested independently. The statistical relationships to be evaluated are represented in a conceptual framework as shown in Figure 6.

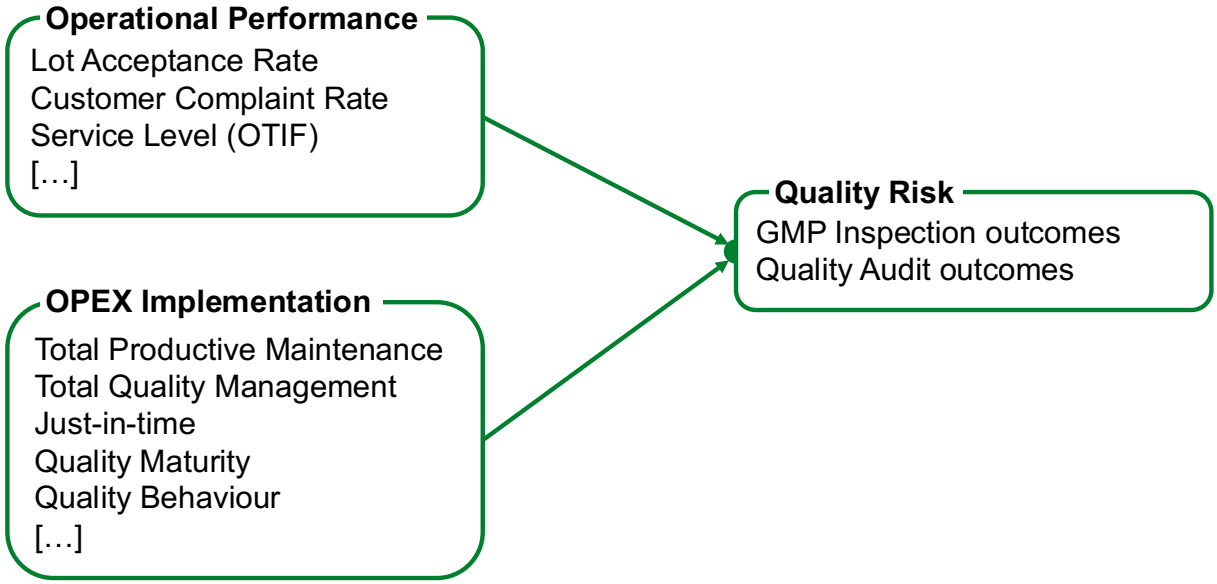


Figure 6: Conceptual framework

Decision-making mechanisms developed in SRQ 4 will serve as the artifact or product for industry practice. They and have to be verified and refined to fulfill

the requirements for successful DSR (Dresch, Lacerda and Antunes Jr, 2015). That verification will be done using a case-study based on a joint discussion of analytics results, and their practical usability with subject matter experts. This aligns well with scholars recommending “companies should look for combinations of qualitative and quantitative tools” for Quality Risk Management (Martins, Eduardo, and da Silva, 2019, p. 266).

1.4 Thesis Structure

The structure of this monograph follows the previously introduced methodology. The specific investigation of quality risk predictors lays the foundation for subsequent modeling in the thesis. First, a literature review compiles existing knowledge based on SRQ 1 in Chapter 2.3.2. The analyses of predictors from Operational Excellence implementation and operational performance metrics, each for manufacturing sites, and QC labs happens in distinct sub-chapters of Chapter 3 (SRQ 2 &3). Within the fourth chapter, the introduction to the project setting, and available data is described followed by an overview of the analysis methodology. Model selection and performance evaluation are prepared separately for two different operationalizations of inspection outcomes and thereby reflect the iterative approach (cf. Chapter 1.3.4 Research Methodology & Process). The discussion and validation of the analyses consolidate the findings of the chapter. This dissertation concludes by addressing the practical implications and contributions for scholars and industry practitioners. It also discusses the limitations of the performed research and enumerates future research opportunities. Figure 7 sketches a visualization of the described thesis structure.

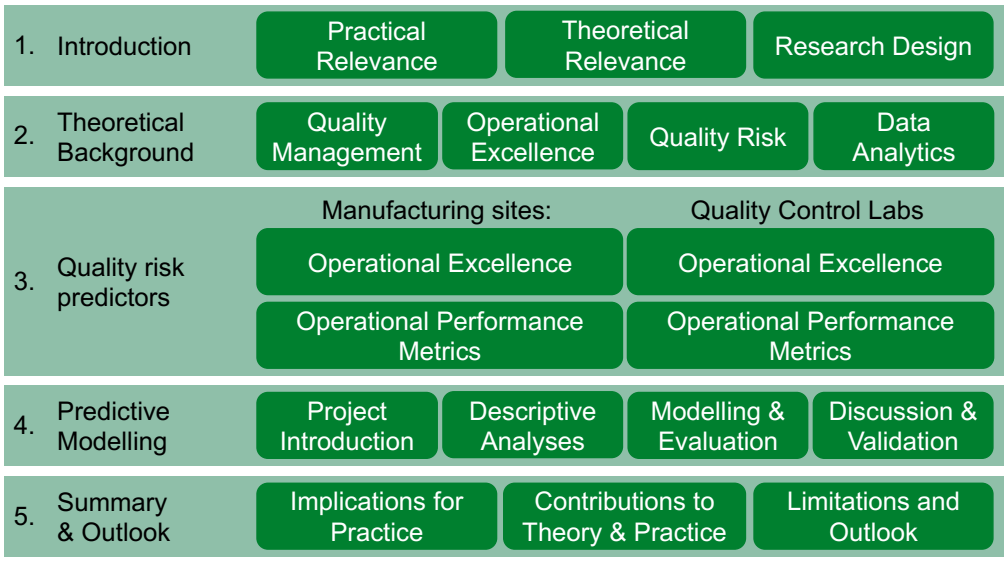


Figure 7: Thesis structure

2. State of Research

The framework of new research should be based on a comprehensive understanding of existing research and publications. This makes the exploration of the research gaps worthwhile, providing the researcher an in-depth understanding of the relevance of the potential research questions (Cooper, 1988; Webster and Watson, 2002). The following section addresses the theoretical background of the research project. Methodologies for assessing, managing, and assuring product quality have been at the heart of industry initiatives, and academic research since the early years of the last century. That development and evaluation of the state-of-the-art practices are introduced in Chapter 2.1. Since the 1980s, Lean Production and Operational Excellence (OPEX) has popularized a holistic approach to customer value creation that builds on continuous improvement. OPEX as a term and concept was broadly adopted in the pharmaceutical industry, and its evolution is documented in Chapter 2.2. The inclusion of the term risk in quality management practices is a rather recent development in the pharmaceutical industry. Chapter 2.3 provides details on general and industry-specific concepts. Derived from scientific literature and practical insights, this chapter concludes with the presentation of the identified research gap.

2.1 Quality Management

Quality Management (QM) has been investigated as part of management studies for more than a hundred years. Throughout that time, many academic researchers, and industry scholars published and advocated the need for innovative methods, strategies, and processes. An overview summarizing the development of QM, its concept maturity, and scope is represented in Figure 8.

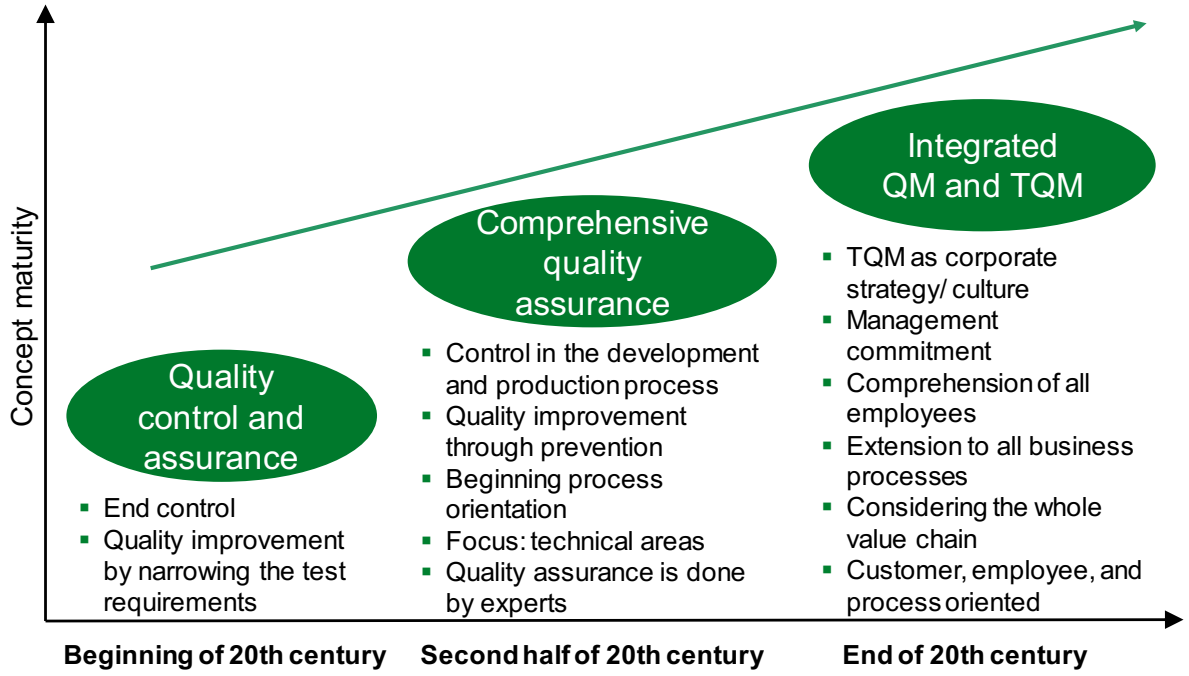


Figure 8: Quality concepts over time. Illustration adapted from Pfeifer, 2002.

Quality Control or Quality Management functions have been implemented in most manufacturing companies starting with specifications testing after the implementation of mass production. Such inspection of products is designed to detect defective products and prevent them from getting shipped to customers. This aims to maintain product quality and ensure delivery of acceptable products only. Identifying sub-standard products, reworking the same to meet the specifications, or discarding them uses a significant amount of resources. Therefore, the paradigm of quality control and assurance shifted towards the identification of the sources of errors (Weckenmann, Akkasoglu, and Werner, 2015).

Tracking defect types and error frequencies effectively requires capabilities other than metrology. The shift towards modern quality control meant that the technical quality discipline adopted a more mathematical statistics approach. Shewhart pioneered this development by implementing sampling procedures, and control charts grounded in statistics. The key element of these charts is the visualization of manufacturing process variance against the defined specification limits. Shifting such control mechanisms from separate quality departments after the fact to production and individual manufacturing steps resulted in the establishment of Statistical Process Control (SPC). Detecting discrepancies early on allowed companies to achieve substantial savings in

reducing rejected product and reaction time (Crathorne and Shewhart, 1933; Weckenmann, Akkasoglu, and Werner, 2015; Dale, Dehe, and Bamford, 2016). According to Pfeifer, the combination of these concepts forms Quality Assurance (Pfeifer, 2002).

The attention paid to Quality Management, and the speed of further development of the discipline culminated in the 1980s. Several well-known scholars contributed to the new methods and perspectives in the academic and practitioner discussions. Crosby and Deming again expanded the scope of Quality Management by enhancing the focus to the overall organization, and its environment. They paid special attention to the interactions and interfaces between all business processes, the product, and the customer, thereby establishing a systems perspective. This system-oriented view constitutes the establishment of Total Quality Management (TQM) (Crosby, 1979; Deming, 1982; Weckenmann, Akkasoglu, and Werner, 2015).

The holistic management approach is one of the common denominators of the various interpretations and understandings of TQM across the different scholar communities and quality organizations. TQM integrates all areas of the value chain and organization. It requires the top management's commitment to quality and involved providing necessary training to the employees at all levels. Horizontally, it requires the involvement of all steps from research and development to supplier quality management, operations, and customer service. Academics have shown that the implementation of TQM also results in significant performance gains on the product level and market performance (Saraph, Benson and Schroeder, 1989; Kaynak, 2003; Dahlgaard, Kristensen and Khanji, 2005).

Since the 1980s, TQM has been integrated with other quality methods such as Kaizen, Six Sigma, as well as the quality-related aspects of Lean Production. The latter and its demarcation from TQM will be illustrated in the following chapter. Another Quality concept that relates specifically to the emphasis of statistical methods is Six Sigma (6σ). In essence, it describes the goal of controlling a (manufacturing) process to produce specifications within several standard deviations of the specification limits. To do so, a selection of quality methods are utilized that draw from existing concepts, such as the Deming's cycle (Tennant, 2001).

2.2 Operational Excellence

In this thesis, OPEX shall serve as an umbrella term for productivity and quality improvement programs that originated from western interpretations of the Toyota Productions Systems. The definitions and limitations of Lean Production and World Class Manufacturing are discussed in the subsequent sub-sections.

2.2.1 World Class Manufacturing

Not explicitly referring to the Toyota Production System, Wheelwright and Hayes introduced the term ‘world class manufacturers’ in a study comparing the competitiveness of manufacturing organizations in the US, Europe, and Japan. The essence of that study is that manufacturing can significantly contribute to companies’ overall competitive advantage. This is in contrast to American companies that traditionally relied on highlighting their product offerings and their marketing. However, building a competitive advantage through manufacturing requires companies to make the right choices, and follow through with the implementation of the resulting changes. This can “take a tremendous effort” and requires the organization to bear “organizational strain” that comes with a risk of failure (Hayes and Wheelwright, 1984; Wheelwright, Hayes and Wheelwright, 1985). Similarly, Voss et al. concluded that starting the process of change to overcome the organization’s intrinsic inertia is the critical obstacle for organizations attempting to achieve excellence in manufacturing (Voss *et al.*, 1995).

In order to reach the status of a world class manufacturer, companies shall foster several capabilities and methodologies:

- Employee-empowering workforce management (learning, indirect control, management of attention, and problem-solving)
- Pro-active scouting and usage of new technology in manufacturing (alignment of manufacturing development according to enterprise strategy, and creation of long-range plans, the autonomy of manufacturing units to pursue the exploitation of production-related opportunities)
- Continual improvement (Wheelwright, Hayes and Wheelwright, 1985).

Since then, many scholars have used their work as inspiration, and incorporated the proposed capabilities to base their research on this framework (cf. Ferdows

& De Meyer, 1990; Friedli et al., 2013; Friedli, Basu, Gronauer, & Werani, 2010; Shah & Ward, 2003, 2007).

2.2.2 Lean Production

Nowadays, Lean Production is an integral part of most industrial manufacturing operations all around the globe. It is credited with enabling superior performance that allows companies to gain invaluable competitive advantages through manufacturing. There are numerous interpretations of Lean Production that range from a philosophy to a tool-kit of specific methods which has created ambiguity and confusion. The philosophical interpretation of Lean Production is to continuously adapt processes and products to eliminate all kinds of waste. The tools and methods placed within Lean Production vary tremendously across different scholars. Some of the most prevalent practices across different interpretations involve the continuous introduction of improvement programs, flow production, and the pull system (Shah and Ward, 2003).

The overall goal of creating an organization that follows the lean production principle is to become more flexible and focus all activities on generating value for the customer. Flexibility is established by organizing the workforce in a flat hierarchy, training employees to perform various tasks, and empowering them to adapt and improve performance. The procedural part of achieving flexibility lies in reducing intermediary inventories, engineering a just-in-time process, and fostering adaptive supply networks. (Krafcik, 1998; Womack and Jones, 2003)

For Womack and Jones, value must be defined from the customers' perspective, and all activities or attributes not contributing to that goal are waste (Womack, Jones and Roos., 1990; Womack and Jones, 2003). Ohno, the visionary behind the Toyota Production System, classified the possible types of waste in processes into seven groups of waste or *muda* (Ohno, 1988):

- **Transportation**: Unnecessary transport of parts under production.
- **Inventory**: Stacks of parts waiting to be completed or finished products waiting to be shipped.
- **Motion**: Unnecessary movement of people working on products.
- **Waiting**: Unnecessary wait by people to begin the next step.
- **Over-Processing** the product with extra redundant steps.
- **Over-Production** of products exceeding the requirement or demand.
- **Defects** in the product.

Lean Production is also tightly connected to technical changes to products and processes that improve quality. *Jidoka*, the 'automation with a human touch' describes a contrivance that automatically interrupts the manufacturing process once a deviation occurs. Subsequently, the issue is analyzed and corrected by the workers before the production resumes. Similarly, the concept of *Poka-Yoke* was conceived to fool-proof manual processes. Products and processes are designed in a way such that it automatically points workers to executing the correct next steps or makes incorrect handling impossible. (Ohno, 1988; Shingo and Dillon, 1989)

Based on the implementation of a lean production system, Womack and Jones observed fundamental gains in manufacturing effectiveness and efficiency. An optimized pull value stream allows cutting down throughput times and inventories to a tenth of the previous batch system. In parallel, labor productivity can be doubled with an error rate reduced to half. (Womack and Jones, 2003)

Since then, Lean has been applied to various managerial functions, such as general administration, research and development, project management, and services (Schuh, 2013; Sirkia & Laanti, 2015). Womack and Jones themselves promoted a broader conception of Lean as *Lean Thinking* that can benefit "*any industry in any country*" (Womack and Jones, 2003). However, Lean remains the most prominent in manufacturing. Within this thesis, the terms Lean Production and Lean Manufacturing are used synonymously as previously coined in several publications (Krafcik, 1998; Pettersen, 2009; Womack & Jones, 2003).

2.2.3 Operational Excellence as the cumulation of an improvement journey

As highlighted in Chapter 2.2.1, excellence can only be achieved over a longer period of building and sustaining improvement capabilities (Hayes and Wheelwright, 1984). Operational Excellence, therefore, is the organizations' sustained adoption of Lean Production, which specifically requires stable, unceasing dedication (Pettersen, 2009). This includes the employees' absorption of the Lean philosophy, and their steady empowerment to becoming change agents in the implementation of lean practices (Womack, Jones and Roos., 1990). Only then can they maintain an unbroken motivation to continuously improve their workplace, processes, and products (Zhang, L.;

Narkhede, B. E.; Chaple, 2017). This process demands the top management to devote their time to Lean and show their dedication (Jackson, 1996; Zhang, L.; Narkhede, B. E.; Chaple, 2017).

Research suggests that the accumulation of skills and sophisticated methods in the manufacturing environment increases performance output in various aspects. Such factory fitness goes beyond the individual progress in implementing specific lean methods. Instead, it enables cumulative effects that start with quality at the base of a sand cone model of cumulative performance. Establishing an advanced state of factory fitness enables a plant to cope with challenges, and met the requirements in various areas (Ferdows and Thurnheer, 2011; Bortolotti *et al.*, 2015).

2.2.4 Operational Excellence in the Pharmaceutical Industry

While OPEX and Lean Production became a major part of the industrial guiding principle in the 80s and 90s, similar initiatives in the pharmaceutical industry only commenced in the early 2000s (Friedli *et al.*, 2013). Since then, it has received strong attention within the industry, and most companies have developed their own OPEX initiatives. Many of them documented these which includes industry leaders like Abbott, Amgen, Novartis, Novo Nordisk, Pfizer, and several others (Friedli *et al.*, 2010, 2013; Friedli, Basu, *et al.*, 2018). In one survey, it was even found that every responding pharmaceutical company has implemented OPEX at least partly (Porsche Consulting, 2013). In addition to the practitioners' domain, there is a limited amount of literature that provides a more academic perspective on the industry, and its adoption of Operational Excellence (Coffey, 2008; Pavlović and Božanić, 2010; Nenni, Giustiniano and Pirolo, 2014; Haleem *et al.*, 2015; Muse, Njeru and Waiganjo, 2016; Yu and Kopcha, 2017; Birkie *et al.*, 2018; Garza-Reyes *et al.*, 2018; Köhler, 2019; Köhler, Friedli and Basu, 2019).

Quality certainly is a key part of manufacturing excellence in its traditional sense (Voss *et al.*, 1995; Friedli, Basu, *et al.*, 2018). However, this concept of quality certainly is not well-defined in the pharmaceutical industry. On the one hand, Operational Excellence (OPEX) or Lean Production initiatives in the companies often evolved out of earlier 6σ programs that directly addressed quality (Friedli *et al.*, 2010; Nenni, Giustiniano and Pirolo, 2014). On the other hand, Quality and Operational Excellence are often separated organizationally and culturally,

which can potentially be attributed to the difference in specializations of scientists and industrialists who are in charge of the programs. There is also a regulatory requirement of having an independent quality function. (Colledani *et al.*, 2014; Eger *et al.*, 2018; Friedli, Basu, *et al.*, 2018)

2.3 Quality Risk Management

The desire to eliminate the chance of failures and non-quality in products and processes, rather than correcting them has been part of industrial engineering for a long. The catalogs of Quality Management techniques have turned to preventive approaches several decades ago (cf. Figure 8; Pfeifer, 2002). Similarly, in the Lean Production or Operational Excellence sphere, preventive methods such as *Poka Yoke* also are an important component (Ohno, 1988).

The combination of the terms ‘risk’ and ‘quality management’ however, is rather new. The evidence of adverse quality events persisting and increasing even though industries mature in Quality Management and Operational Excellence, could have motivated the specific risk perspective (Bates *et al.*, 2006). ISO recognized the need to foster risk-based thinking beyond selective preventive methods in their renewed 9001 standards (ISO, 2016). Since then, the topic has received increased attention in academic literature (Martins, Eduardo, and da Silva, 2019; Rampini, Berssaneti, and Saut, 2019).

The methodologies and tools used in association with Quality Risk Management (QRM) are not new, but well established like the Deming-inspired Plan-Do-Check-Act cycle (Deming, 1982), and the Failure-Mode-and-Effects-Analysis (FMEA). In their review, the authors not only identify general conceptual contributions, but cite examples of studies specific to industries, such as Banking, Engineering, and Woodwork (Martins, Eduardo, and da Silva, 2019).

Contributions to Quality Risk Management also exist outside the ISO 9001 context. While there are a few examples from diverse backgrounds, such as higher education, civil engineering, and agriculture (Raban and Turner, 2006; Borodin *et al.*, 2014; Qing *et al.*, 2014), the vast majority represents the domain of pharmaceuticals. The above-mentioned literature is reviewed in the following sections.

As discussed in Chapter 1.2, Supply Chain Management is another closely-related discipline, which also employs a risk management perspective (Jüttner and Christopher, 2003). While quality risk related to manufacturing is a key part

of possible threats to a functioning supply chain, it has not yet been explored extensively (Ho *et al.*, 2015). Quality risks often are viewed as immanent to global supply networks that lack transparency to the brand manufacturer (Tse *et al.*, 2011).

Moktadir *et al.* identified machine, equipment, or facility failure, and quality risks to be among the top three threats to pharmaceutical supply chains (Moktadir *et al.*, 2018). However, the literature does not guide evaluating the risk associated with individual manufacturing plants. Therefore, the companies that do employ a risk perspective in their supply chain, do not use operational metrics or quantitative assessments (Kalska-Zimnicka, Calnan and O'Sullivan, 2017).

2.3.1 Quality Risk Management in the Pharmaceutical Industry

Prior to 2005, there were very limited examples of Quality Risk Management being applied in the pharmaceutical industry (ICH, 2005). However, it became popular in the industry after the release of the ICH Q9 statute developed by an International Consortium of Regulators and Industry (Waldron, 2017). The core principles of QRM according to ICH Q9 are:

- *The evaluation of the risk to quality should be based on scientific knowledge and ultimately link to the protection of the patient; and*
- *The level of effort, formality, and documentation of the quality risk management process should be commensurate with the level of risk.*

The statute suggests the application of a quality risk perspective at all stages of a drug's lifecycle, and all steps of the value chain, as well as during inspections. Applying the second core principle to the field of inspections, ICH states that "*Risk ranking and filtering can be used to prioritize manufacturing sites for inspection/audit by regulators or industry*" ICH, 2005, p.14). It thereby endorses the use of quality risk perspective as also proposed in the case presented in this thesis. In this light, an industry convention suggested a model for risk-based inspection planning that was based on ICH Q9. It determines inspection frequency based on product complexity, criticality, and regulatory compliance

history. However, that model still relies purely on quantitative assessments (PIC/S[3], 2012).

So far, the research publications inspired by ICH Q9 have focus areas other than inspection urgency determination or plant-level risks. A representative overview of recent publications on QRM is shown in Table 2. There are several examples of research addressing the implementation of Quality Risk Management in the industry. These studies focus on the overall adoption of ICH Q9, the challenges associated with it (Mollah, Long, and Baseman, 2013), and the subsequent trends regarding overall quality (Waldron, 2017). One highlighted aspect is the risk assessor's subjectivity and its influence on the results, since small deviations can have a great impact on the results with the example of FMEA (O'Donnell, 2010; Claycamp, 2015; Mcfarland and Waldron, 2015; Ramnarine, 2015). Similarly, improving cultural awareness (Richter and Haddad, 2015), and knowledge in the organization can improve risk estimations, and there are numerous publications that consider knowledge management to be closely related to QRM (Greene, O'Donnell, and Calnan, 2018).

The regulatory burden of changing products or manufacturing operations through post-approval changes (PAC) is a big challenge for pharmaceutical companies, which makes them resistant to change. In consequence, companies might refrain from implementing incremental improvements of their process quality or efficiency, because the effort for approving them is too high. The recent Q12 guidelines by ICH suggest lowering these hurdles by providing less stringent PACs for companies that successfully demonstrate PQS effectiveness (ICH, 2019). To date, it is unclear by what measure the effectiveness of the PQS can be potentially measured. Successfully utilizing a predictive approach to quality failures, and mitigating weaknesses before they become relevant for the regulator could be a qualifying marker of an eligible Quality System. Therefore, this is a strong motivator for companies to further develop their capabilities and systems in this sector. (Ramnarine *et al.*, 2017; Greene, O'Donnell, and Calnan, 2018; Ramnarine and O'Donnell, 2018)

Computational technologies such as data analytics and simulations are recognized as being beneficial for risk assessments, but primarily for process design and validation (O'Connor, Yu, and Lee, 2016; Walters and Barneva,

[3] PIC/S stands for *Pharmaceutical Inspection Convention/Pharmaceutical Inspection Co-Operation Scheme*

2017; Deidda *et al.*, 2018). These generally are among the primary domains of QRM and therefore, several contributions that target QRM in development and validation exist (Charoo and Ali, 2013; Soto, 2018; Nagle, 2019). This research is closely related to earlier discussions of the application of Quality-by-Design, which aims to eliminate risks in processes before they are implemented (Fahmy *et al.*, 2012; Patil and Pethe, 2013; Milmo, 2014; Das, Mishra, and Nayak, 2017).

Table 2: Recent publications on pharmaceutical QRM

Article	**Research Focus**	**Object of Analysis**
Mollah, Long, & Baseman, 2013	The authors comment on different challenges of the adoption of ICH Q9. For instance, many companies invest resources to formally complete risk assessments but do not utilize the findings in manufacturing.	QRM implementation
Charoo & Ali, 2013	QRM shall also be applied to product development and a few specific use cases are discussed in this paper. The authors make the case for an integrated perspective on the life cycle, so that product development already informs manufacturing controls.	Product development
Richter & Haddad, 2015	Senior Leadership must accentuate the importance of Quality and responsibility. That way, creating a strong Quality Culture in the company will in turn foster QRM as part of the culture.	Quality Culture
Claycamp, 2015	Most QRM methods (such as FMEAs) rely on the subjective judgement of failure probability and severity. The author suggests that users should deepen their knowledge of decision theory to reduce bias and the impact of emotion.	QRM implementation
O'Connor et al., 2016	The authors evaluate the benefit of simulation models in the process design phase to inform the QRM evaluation and test mitigation strategies.	Process design / validation

Ramnarine et al., 2017a	Effectiveness of the Pharmaceutical Quality System is a prerequisite to enable easy post-approval changes. Quality Risk Management is a key part of that PQS.	Qualification procedures & investigations
Yu & Kopcha, 2017	The pharmaceutical industry is urged to pursue Six Sigma, Quality by Design, Operational Excellence, and other methodologies to improve drug supply. This includes a stronger focus on reporting quality-related metrics.	Quality Management Systems
Waldron, 2017	The study evaluates the industry's progress in implementing QRM and evaluates the results on a macro scale. Despite the rollout, the state of quality worsens. Proposed measures include focusing on risk-informed decision making and exploring new ways of QRM.	QRM implementation
Hata, 2017	The author links various activities from planning facilities in a flood area to covered climate chamber fans and possible risks.	Miscellaneous
Walters & Barneva, 2017	Risk assessments based on historical data analysis as an estimation for failure occurrence does improve accuracy. This helps with the prioritization of process improvement.	Process design / QRM implementation
Soto, 2018	Proposal for a risk-based re-qualification process of critical systems based on their characteristics and monitoring of risk indicating events such as deviations and CAPAs.	Production systems/processes
Greene et al., 2018	Several contributions, among others on the subjectivity in risk rating, QRM competencies as well as future visions of measurable and evidence-based QRM.	Qualification procedures & investigations
Ramnarine & O'Donnell, 2018	Suggestion on evidence-based methods to demonstrate the effectiveness of Pharmaceutical Quality Systems for risk reduction to become more flexible and allow freedom for change management.	Production systems/processes

Deidda, Orlandini, Hubert, & Hubert, 2018	Authors argue the application of statistical tools for quality control (testing) methods, such as analytical-quality-by-design.	QC method development
Nagle, 2019	Proposal for periodic reviews of equipment and computer systems' validation that is risk-based.	Production systems/processes
Haddad, 2019	Development of standard roles for QRM within the company and assignment of competencies assigned to roles. The competency model is intended for practical use e.g. in specifying training plans.	Organizational roles and competencies in QRM

Similar to the disconnect existing between Quality and Operational Excellence (cf. Chapter 2.2.4), the literature sparingly links OPEX to QRM. However, Yu & Kopcha state that QRM is a condition for enabling the successful implementation of a Quality Management System. Further, Operational Excellence will control instabilities that can trigger adverse quality events like drug shortages. (Yu and Kopcha, 2017)

2.3.2 Quality Risk research using inspection outcomes

As alluded to in Chapter 1.2, FDA inspection outcomes are relatively easily available, providing extensive data analyzed in several contexts by a group of authors. Starting in 2011, several articles were published using different sets of publicly available data that could be connected to the inspection outcomes. Table 3 summarizes the relevant literature analyzing the connection of inspection outcomes in FDA-regulated industries and selected related publications[4]. Additionally, it highlights that FDA inspection outcomes most often are used as dependent variables (DV). However, the use as an independent variable (IV) also exists. The literature was compiled through a Scopus and Google Scholar search for literature that connects Operational Excellence (also Lean, Production Systems, Quality Management), and regulatory inspection outcomes (also quality compliance, quality risk, process compliance). Since the implementation of Operational Excellence in the pharmaceutical industry is

[4] Please see appendix A.1 for full size table

relatively new, the search corridor was limited to the last 15 years. Individual sources were then selected based on whether they use inspection outcomes in the study.

The findings show that offshore manufacturing locations with language differences between the company headquarter and the plant increase quality risk. However, these two studies did not have suitable controls for other variables. Other risk factors include lower education levels and the absence of other manufacturing facilities in the vicinity. As reported by two other studies, longer periods between inspections also increase the risk of adverse inspection outcomes, which is in line with the FDA's approach. (Gray, Roth, and Leiblein, 2011; Anand, Gray, and Siemsen, 2012; Gray and Massimino, 2014)

Evidence pointing towards the deterioration of quality compliance can also be observed when analyzing the relationship between inspection outcomes and ISO 9000 certifications. Companies with recent certification show better inspection outcomes, which deteriorates over time. The positive effect is particularly strong for companies that pursued early certification, possibly demonstrating a pro-active approach to improving their Quality Management system. (Gray, Anand, and Roth, 2015)

Another thoroughly studied phenomenon is the relationship between brand manufacturers, and contract manufacturers (CM). While CMs typically operate at a higher risk, that risk particularly increases if the CM manufactures with high process and/or product heterogeneity. The risk can be reduced by closer collaboration between the buying and supplying company, as well as more stringent monitoring by the regulator (Handley and Gray, 2015; Gray, Roth and Tomlin, 2016).

Lastly, an unpublished study by Seiss (2018), which was carried out as a part of an FDA research contract applied predictive models to structural data from manufacturing plants and used past regulatory compliance to predict subsequent inspection outcomes. The data produced models with some predictive power on whether or not actions were indicated in the conclusion of the inspections. The predictive performance increased slightly when additional financial information was added to the pre-existing dataset.

Interestingly, several different operationalizations of the inspection outcomes have been used over time. The most prominent version developed by Gray *et al.* (2011) is a scale that uses the final decision following the inspection, as well

as the issuance of the FDA form 483. This form is the formal way of communicating deficiencies in the plant. Through a Delphi process with an expert panel, these inspection outcomes were transformed onto a scale from 0 to 3.5. However, only three of these outcomes occur frequently. Typically, without a form 483, there are no actions indicated (NAI) and otherwise, either voluntary (VAI) or official actions (OAI) are expressed (Gray, Roth, and Leiblein, 2011). Operationalizations directly based on the final decision (NAI / VAI / OAI) are also popular. In addition to that, there is also a binary distinction between NAI and VAI or OAI inspection outcomes in Seiss (2018).

Table 3: Related literature focusing on Quality Risk on the plant level

Article	Type of data / sample size	Industry	Independent variables	IV/DV	Dependent variables	Control variables	Terminology	Methods	Results
Gray et al., 2011	672 inspections of 60 plants	Pharma	Regional high school education share; regional industry cluster, site distance from company HQ, US mainland / offshore	DV	Delphi-process based score of inspection final decision and form 483 issuance	Plant age (years), plant size (number of employees), and company size (sales)	Quality risk	Paired sample; Ordinary least squares regression	Used individually in regression models, regional high school completion quota, industry agglomeration, proximity to headquarter, and mainland location reduce quality risk
Anand, Gray, & Siemsen, 2012	8080 inspections of 1493 plants	Pharma	Time since last inspection; merger/acquisition prior to inspection	DV	Factor operationalization: „OAI" / „VAI" / „NAI"	Inspection type, previous inspection outcome, year, foreign location	Process entropy	Random Coefficients Ordered-Probit Analysis	Good states of process entropy deteriorate over time; mergers negatively affect entropy whereas acquisitions do positively
Gray & Massimino, 2014	7696 inspections	Pharma	Language and cultural differences between site and head quarter	DV	Factor operationalization: „OAI" / „VAI" / „NAI"	Previous inspection outcome, time since previous inspection, Ownership change, head quarter-plant distance	Process compliance	Ordered probit regression, polynomial regression	Time since last inspection and language difference clearly harm process compliance; no clear picture for cultural differences, but similar power distance at HQ and site has positive impact
Gray, Siemsen, et al., 2015	8832 inspections	Pharma	Firm size, site (co-)location with R&D, information and communication technologies (ICT) availability, Tacit process knowledge	DV	Binary operationalization: „OAI" vs. rest	Plant type, age, population density, R&D intensity	Conformance quality	T-test, (two-stage) probit regression	Colocation improves quality conformance regardless of ICT access or time; colocation especially important for large companies and complex products
Gray, Anand, et al., 2015	Overall, 552 plants	Medical devices	(Early) ISO 9000 certification and time since certification	DV	See Gray et al., 2011	Public companies, standalone sites, plant size	Process compliance	Pearson Bivariate Correlation, T-Test	ISO 9000 certification has a positive effect on process compliance if obtained early; the effect decays over time
Handley & Gray, 2015	106 supplier – customer pairs	Food, pharma, and medical device	CM's Manufacturing heterogeneity, coordination mechanisms eg. Cooperative relationship (moderator), attention mechanisms eg. Formal performance assessments (moderator)	-	Four level scale survey	Regulatory intensity, CM size, plant age, lenght of relationship, process type	Conformance Quality Performance	Regression	Negative link between heterogeneity and CM's conformance quality performance; contractual coordination provisions and formalized performance assessments can minimize the adverse effect of heterogeneity on the CM's conformance quality performance

Gray et al., 2016	1253 inspections	Pharma	Contract manufacturing (CM) vs. Internal plant, manufacturing experience (moderator), inspection intensity (moderator)	DV	See Gray et al., 2011	Company sales	Quality risk	Ordinary least squares regression	CMs operate at higher risk, but more experience reduces risk; higher inspection intensity reduces risk particularly for CMs
Altamuro et al., 2017	216 companies	Pharma	Internal control weakness	DV	See Gray et al., 2011; Financial non-compliance (accounting restatements)	Management ability, ROA, sales growth & volatility, loss, leverage ratio (total debt / assets), firm size, complexity (number of geographic & business segments), Foreign income / Foreign plant, 'Big Four' auditor, free cash, new debt / equity	Operational non-compliance	Probit regression	Weak compliance culture harms operational and financial non-compliance; contemporary financial and operational non-compliance correlate
Ball et al., 2017	2231 recalls	Medical device	Inspection outcome, inspector (site specific) experience	IV	Recall classification (class I vs. II)	Sales, Firm, Public companies, local college completion ratio, Time since last recall, previous recalls, product scope, number of products, year, products' technological level, product novelty, inspection type	-	Cox regression analysis	Inspection outcomes does predict future recall severity; site-specific experience of the investigator does harm that relationship
Shah, Ball, & Netessine, 2017	67 recalls	Auto-motive	Product and plant variety, utilization	-	Recall	Plant size (volume, employees), year, manufacturer, plant age, years since launch, segment	-	Correlation matrix, generalized estimating equation GEE regression	High variety and utilization increase the likeliness of recalls, particularly when combined
Seiss, 2018 (non-public project report)	3429 inspections	Pharma	Previous inspection outcomes, time since last inspection, financial data	DV	Binary operationalization: „NAI" vs. rest	-	-	Logistic regression & scorecard prediction models	Quality compliance history does somewhat predict future inspection outcomes; financial data slightly improves the predictions

The more recent study by Shah *et al.* (2017) is included here, even though it does not make use of FDA inspection outcomes based on three underlying reasons. Other than the closely related topic, and the similar set of authors, it is one of the scarce examples that connect plant internal operational characteristics with quality failures as an outcome. Thirdly, the research remarkably demonstrates the possibility of generating valuable insight out of a smaller empirical dataset (product recalls, n = 67).

As shown in Table 3, inspection outcomes have been used to measure several concepts with "*Process compliance*" and "*Quality risk*" being used repeatedly. Generally, inspection outcomes document a manufacturing site's conformance with the Good Manufacturing Practices (GMP). These guidelines document "*minimum requirements for the methods, facilities, and controls used in manufacturing, processing, and packing of a drug product*" (FDA - CDER, 2018). With inspections being an assessment of compliance with the guidelines, any time the FDA observes compliance failures, the minimum requirements can be deemed violated. In turn, this poses an immediate risk to the quality of goods manufactured. The study by Ball *et al.* (2017), supports the assumption that inspection outcomes can predict the severity of future product recalls. In conclusion, the term "*Quality risk*" is a fitting representation of the studied phenomenon, and connects well to adjunct research around QRM (cf. Chapter 2.3.1), as well as inspection outcomes.

2.4 Data Analytics in Operations Management

Operations Management (OM) as a research practice is deeply rooted in the early works of Frederick Taylor and his approach to scientific management in optimizing manufacturing processes, which classifies as empirical research (Smiddy and Naum, 1954). While the amount of empirical research lagged in other disciplines for a long time, empirical research has gained traction in Operations Management in recent years. This can for example be seen in an increasing number of such publications in leading journals such as *Manufacturing and Service Operations Management* (Fisher, Olivares, and

Staats, 2020)[5]. *Management Science* today remains connected to its origins in Smiddy and Naum (1954) and publishes a similar share of empirical research. *Production and Operations Management* from 2010 to 2015 has also published articles that have an empirical focus in less than 15% of cases (cf. Ho *et al.*, 2017). Roth and Rosenzweig (2020) contributed to this discussion as well and concluded that a sample of 91 empirical articles assessed mostly employed interviews and survey data. These research articles did not report any real company data. They make a strong point encouraging researchers to strengthen this domain by pursuing empirical data in the realm of OM.

Even though it was not widespread until recently, empirical research in itself is not new, inside or outside the domain of Operations Management. Data analytics, as a further development of that, is now enabled by access to *"machine learning, which exists at the intersection of computer science and statistics".* This technique *"has created methods that allow one to obtain high-quality predictive models"* (Mišić and Perakis, 2020, p. 158). In their review of data analytics in Operations Management, they describe data analytics *"as using data to create models leading to decisions that create value"*. They proceed to identify supply chain, revenue management, and healthcare as three areas that so far have received notable contributions based on data analytics. Further, they highlight small datasets as a challenge in doing such research with small meaning the number of observations being relatively smaller compared to the number of variables. Closer related to manufacturing and specifically for *"system-level optimization within and across networks of factories"*, Olsen and Tomlin (2020, p. 117) propose the use of data analytics to enable research that capitalizes on opportunities present in times of Industry 4.0 and increased digitalization of manufacturing. Belhadi *et al.* (2019) confirmed the importance of these techniques for manufacturing organizations to achieve Operational Excellence.

When focusing on the aspect of risk in combination with predictive models, (see also Chapters 1.2 and 2.3) the research domain encompassing supply chain risk is more popular, with several published articles regarding applications of data analytics. Baryannis *et al.* (2019) reviewed related research in that domain

[5] In addition to the assessment by Fisher, Olivares and Staats (2020), (Ho *et al.*, 2017) found that 25% of publications from 2010 to 2015 are empirical. This is in stark contrast to 14.3% of papers being considered empirical, while taking into consideration earlier years (2000 – 2018), as reported by Roth and Rosenzweig (2020). Dai *et al.* (2020, p.2) also observed the contribution of empirical research *"on the rise"*.

and concluded that data analytics still is at a very early stage. They found that only a fractional minority of published research offers practical decision support for practitioners. Similarly, an even smaller share of publications utilizes predictive algorithms. Regarding the data that was analyzed, most research projects (93%) did not have access to large datasets and most analyzed very few parameters at a time. However, the potential of data science and predictive analytics in this domain is to provide opportunities for the advancement of knowledge, and relevant research (Waller and Fawcett, 2013).

2.5 Research Gaps

Concluding the literature review of existing Quality Risk Management literature and empirical quality risk studies, the following research gap can be inferred:

Existing empirical studies fail to investigate the connection between pharmaceutical manufacturing plants' system attributes, operational performance, and their regulatory inspection outcomes. Further, there is no guidance on using such data for quantitatively assessing quality risk at a plant level as advised by the ICH Q9 QRM guidelines. Regarding the used methodologies, there is a scarcity of research applying data analytics in the production management domain.

3. Quality Risk Predictors

Quantitative research projects and their successful conclusion depend on available data access and convincing analytics results, which bears uncertainties. Access to sufficiently big datasets is a prerequisite to succeed in quantitative research and to progress in the field of predictive modeling or machine learning projects. Therefore, the existing and utilized access to data is discussed in this section. While the accumulation of all data sources into one database could grant the biggest statistical power, several practical limitations do not permit this approach. The data attributes in question for this thesis are descriptions of technical advances in production, management characteristics, and operational performance metrics. There is no universal definition for these measures and metrics. Therefore, only data that was gathered on identical scales along with definitions can be compared and the different data sources were not integrated into joint datasets.

The research in this thesis utilizes several different sources of data. For this chapter, the regulatory quality surveillance data from the US Food & Drug Administration was compiled by the research team in St. Gallen, along with the databases mentioned in the subsequent paragraphs.

The key to researching Operational Excellence implementation and performance lies in the St. Gallen benchmarking programs (OPEX BM). These two survey programs consist of production site-specific questionnaires designed specifically for the pharmaceutical industry. At the time of writing this thesis, the St. Gallen benchmarking databases consists of 381 manufacturing sites and 95 quality control labs. Details on the manufacturing-specific database are provided in sub-section 3.1. It serves as the basis of the analyses on Operational Excellence practices (Chapter 3.1) and operational performance metrics (Chapter 3.2), respectively. Section 3.3 introduces the database on pharmaceutical quality control labs in more detail. The subsequent analyses on the implementation of Operational Excellence practices in Quality Control labs operational metrics of robustness and productivity are introduced in Sections 3.3 and 3.4, respectively.

As a surrogate for Quality Risk, this thesis utilizes quality compliance evaluations of the manufacturing sites by the regulatory authority FDA. These FDA inspection outcomes as well as their warning letters to deficient companies are publicly available under the ‘Freedom of Information Act’. Inspection

outcomes, which are used as an important dependent variable and indicator for quality compliance failures, were downloaded from the agency's website (FDA - CDER, 2020c).

The analyses are structured by the domain and corresponding database they are built upon. The focus area of Operational Excellence enabler implementation (SRQ 2) or performance metrics (SRQ 3) are grouped in separate sections. The following figures depict the structure of the remaining chapter and outline the types of analyses and dependent variables used. Figure 9 describes the analyses based on the overall manufacturing sites, and the two streams for OPEX practices (Section 3.1), as well as metrics (Section 3.2). Figure 10 outlines the QC lab-related Sections 3.3 and 3.4.

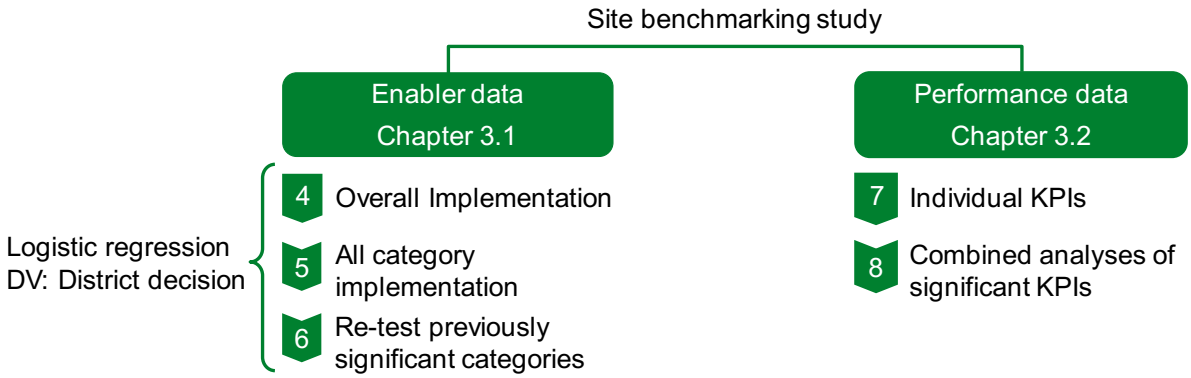


Figure 9: Structure of Analyses based on OPEX benchmarking program

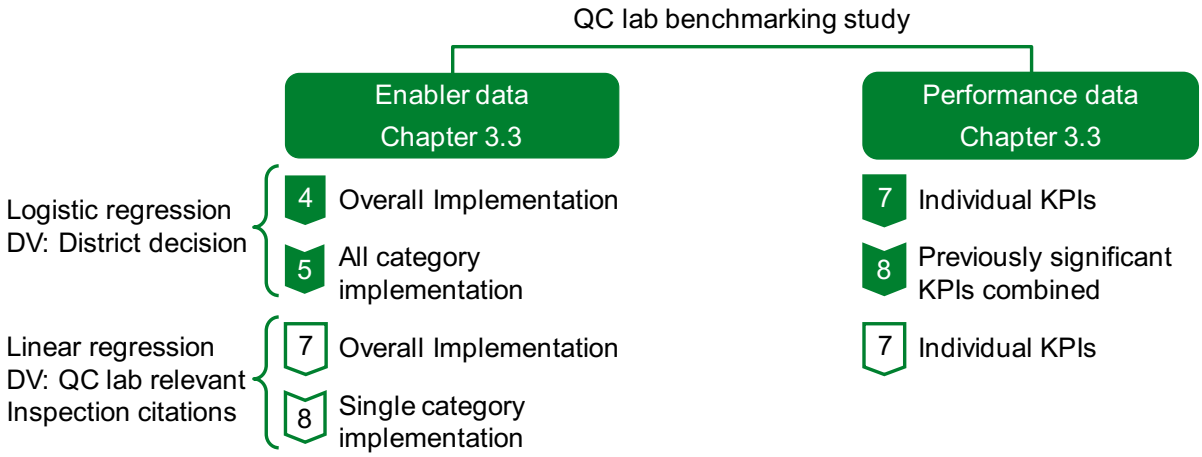


Figure 10: Structure of analyses based on QC lab benchmarking program

Parts of this chapter, particularly sub-sections 3.1 and 3.3, have in part been adapted from my publications as the lead author:

Eich and Friedli, (2021) ‘Analysis of the effects of Operational Excellence implementation on Inspection Outcomes in the Pharmaceutical Industry: An Empirical Study’, *Brazilian Journal of Operations & Production Management*, 18 (Special Issue-IJCIEOM)

Eich *et al.* (2020) ‘Progress of Operational Excellence implementation and performance in pharmaceutical manufacturing sites - descriptive study from 2003 to 2018’, in *Proceedings of the 27th EurOMA*. Warwick, UK: European Operations Management Association.

3.1 Manufacturing site OPEX maturity analyses

Section 2.2.4 introduced the complicated relationship of quality and OPEX in the pharmaceutical industry, which is analyzed in the following. This chapter aims to identify the characteristics of production systems or OPEX programs that can be directly linked to deficient inspection outcomes. This could help industry practitioners to better direct their inspection readiness programs or internal quality audits. At the same time, it can help regulatory authorities perform risk assessments of manufacturing facilities they surveil.

3.1.1 Hypotheses

As highlighted in Chapter 2.3, there is a close connection between Quality Management practices and the concept of Quality Risk, particularly in the pharmaceutical industry. The advancement of Quality Management techniques and their contribution to the overall state of Operational Excellence has been well researched (cf. Chapter 2.2.4). However, the connection of these concepts and Quality (Compliance) Risk, i.e., regulatory inspection outcomes has not been explored yet.

Since the manufacturing sites improve their factory fitness by implementing Lean, this set of cumulative performance advantages should also enable them to achieve better results in quality compliance (cf. Chapter 2.2.3). With special regard to SRQ 2, the first hypothesis states that *plants that are further advanced in their OPEX implementation and therefore reached a high maturity in their improvement journey, are more likely to receive favorable inspection outcomes (I)*.

The Quality System can evolve independently of the overall Operational Excellence, which is particularly true for the pharmaceutical industry (cf. Section 2.2.4). Accordingly, the author expects a *stronger relationship for Total Quality Management implementation (II)*.

3.1.2 Empirical Set-up

This research utilizes a database of pharmaceutical operations compiled in a benchmarking program. Pharmaceutical manufacturing sites are the object of analysis. At the time of writing this thesis, the database consists of 381 site-specific datasets provided by the companies. In return, the participants are provided with an extensive report that compares their Operational Excellence implementation levels, and performance to industry peers. The motives for participation in this study vary: While some sites seek to validate the success of longstanding Operational Excellence initiatives, others use it as a baseline assessment before initiating such an improvement program. Large companies that participate with several sites often aim to cover a broad spectrum from lighthouse or flagship sites to others that are lagging on their Operational Excellence journey. The team, therefore, assumes the sites are not more advanced than the industry average and there is no notable selection bias.
The datasets have been captured over a period of sixteen years starting in 2004. For comparability, the participants were asked to provide their information and metrics for the previous calendar year. The data labels start in 2003 with the geographical distribution of the surveyed manufacturing sites being biased towards European sites (details in Figure 11).

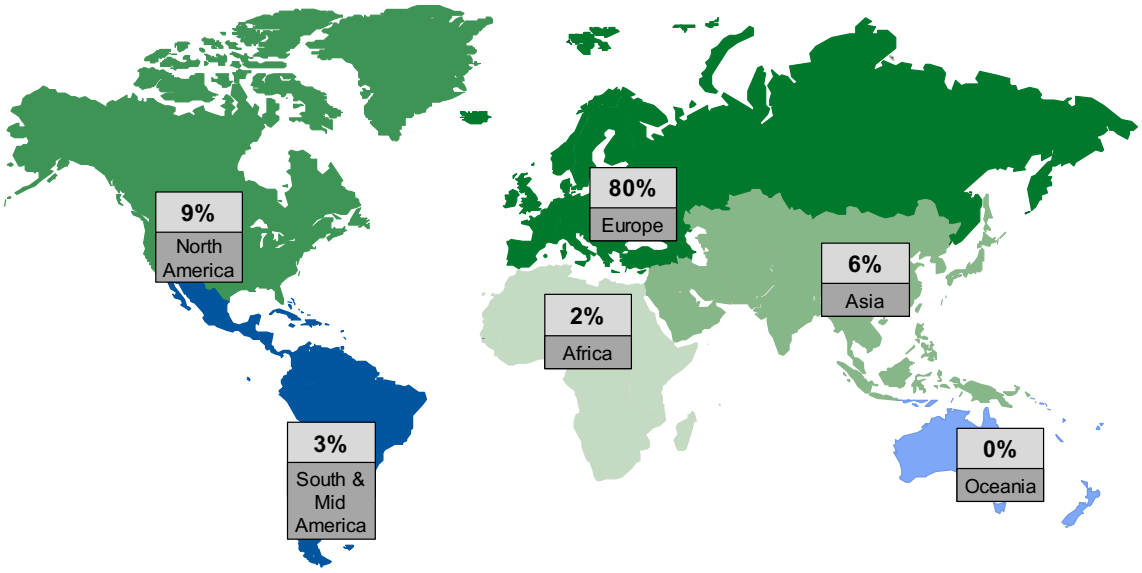


Figure 11: Geographical distribution of the manufacturing site datasets

The research team closely collaborates with the corresponding manufacturing sites and labs to gather these datasets remotely. The data gathering process involves several steps ranging from data validation to feedback to ensure data quality.

The benchmarking study consists of more than 130 survey items related to the implementation of OPEX practices ("enablers"). The data is gathered on a standardized five-point Likert scale and answered by a group of operations and quality representatives of the respective plants. The items are categorized in technical (Total Productive Maintenance TPM, Total Quality Management TQM, Just-in-Time JIT, Standard Work, and Visual Management as basic elements BE), and social question items (Effective management System EMS). The cumulative effect of these competencies shall enable companies to gradually stabilize equipment availability and process quality to reduce inventories *(Friedli* et al.*, 2006, 2010, 2013)*.

In addition to the implementation of OPEX practices, participants are also asked to provide structural information on their establishment, and various performance metrics (cf. Chapter 3.2).

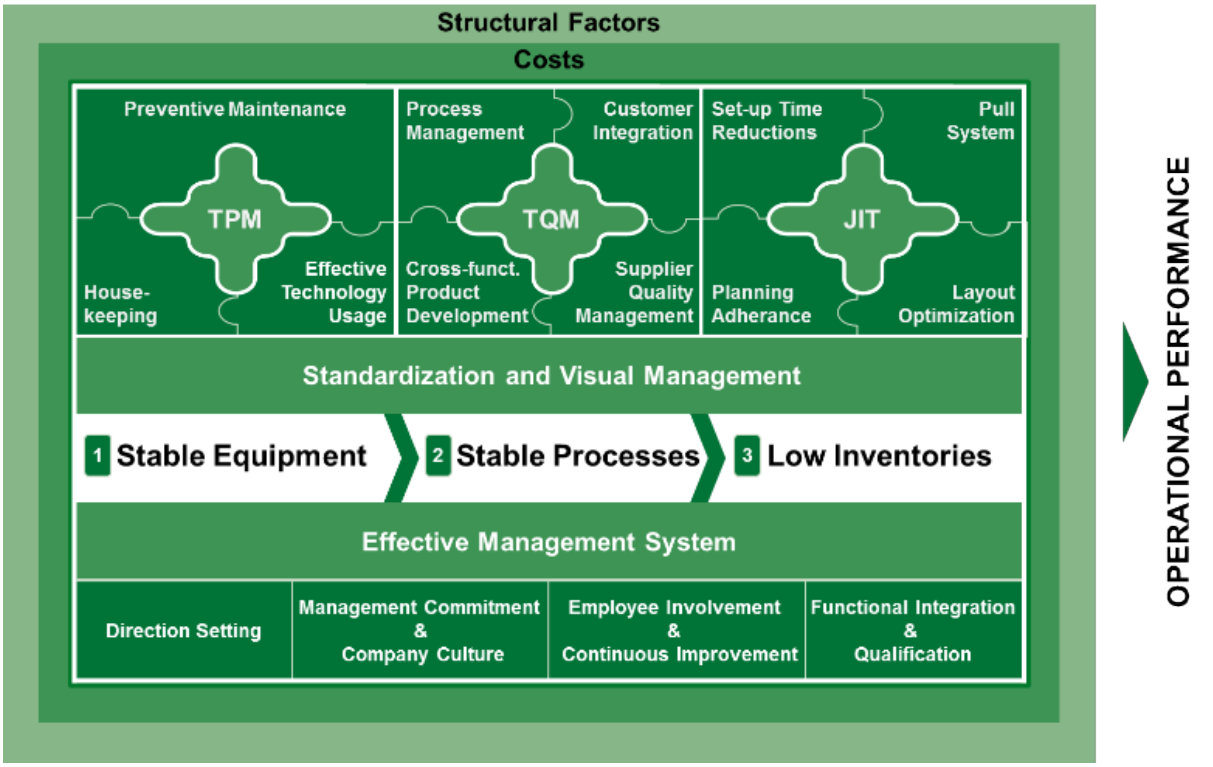


Figure 12: St. Gallen Operational Excellence Model (*Friedli* et al. *2013)*

This company data was paired with publicly available data on FDA inspection outcomes from routine inspections that target compliance with Good Manufacturing Practices (GMP). In the study, 64 site benchmarking datasets were matched with inspection results. After the elimination of datasets with insufficient enabler data, 57 pairs remained. Every of the five enabler categories consists of about 20 questions in three to four sub-categories. While the investigation of such sub-categories could enable much more detailed findings, it was not viable given the current database and the applied "rule of ten"[6].

The comparison between the surveys and FDA databases was done using company name, site address, and year as key criteria for a match. The data reported into the benchmarking surveys typically refers to a site's operational performance related to the full calendar year immediately before when the benchmark was undertaken. Therefore, the criterion "year" seeks to compare the benchmarking survey with the last available FDA inspection conducted before the benchmark for that given organization. Inspection outcomes for more

[6] The "rule of ten" was coined after a number of publications which established the requirement of observations to be about ten times of the analyzed variables in logistic regression models for obtaining reliable results (Harrell *et al.*, 1984; Harrell, Lee and Mark, 1996; Peduzzi *et al.*, 1996). To avoid "extremely large parameter estimates and standard errors, and, possibly, failure of convergence", the sub-categories therefore must remain collapsed (Tabachnick and Fidell, 2014).

than three years prior to the benchmark were not considered. The cut off is a compromise between timeliness and data availability. The authors acknowledge that this study bears certain limitations since it includes data pairs with a considerable time gap of three years. Also, the included inspection could be followed by changes in the site, particularly based on CAPAs (Corrective and Preventive Actions) after adverse inspection outcomes. However, CAPAs following an inspection typically specifically target compliance and may not necessarily have an impact on Operational Excellence programs or enabler implementation. Furthermore, sites might reassign resources from OPEX and continuous improvement to CAPA, and remediation activities. Therefore, the analysis is approach is deemed reasonable with the limitations being kept in mind.

The interrelations between the different groups of enablers are shown in Figure 13. The distribution of each variable is shown on the diagonal. On the bottom of the diagonal, the bivariate scatter plots with a fitted line are displayed on the top of the diagonal. The figure depicts the correlation coefficients, as well as their significance level. The enabler groups are notably correlated with each other. While this hinders a causal inference between a potential change in implementation level and resulting variance in ‘goodness’ of inspection result, it allows predictive statements on the current implementation state.

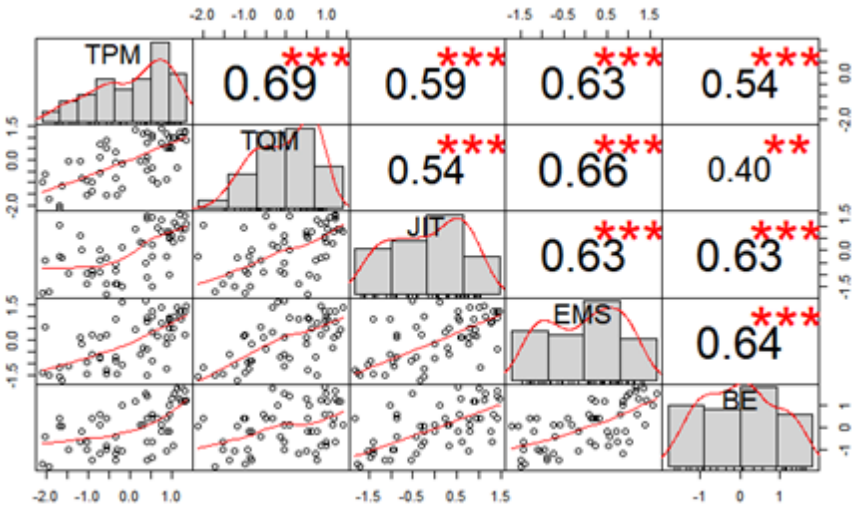


Figure 13: Enabler groups for manufacturing sites: Exploratory data visualization (Significance levels: *** 0.001, ** 0.01).

As for the dependent variable, this research uses the inspection outcome or more specifically the “district decision” or inspection classification. Within the inspection outcomes, ‘NAI’ means that the inspection concluded favorably with

'No Actions Indicated', meaning that no violations were documented. Inspections concluded as 'Voluntary Actions Indicated VAI' are intermediate outcomes with some violations found. Highly adverse inspections result in with the agency assigning 'Official Actions Indicated OAI'. The combined dataset shows fewer OAI final decisions than the overall population (Figure 3), but the distribution frequency is consistent with the overall population of all FDA inspection outcomes when evaluated per Pearson's Chi-squared ($\chi 2 = 2.73$) test.

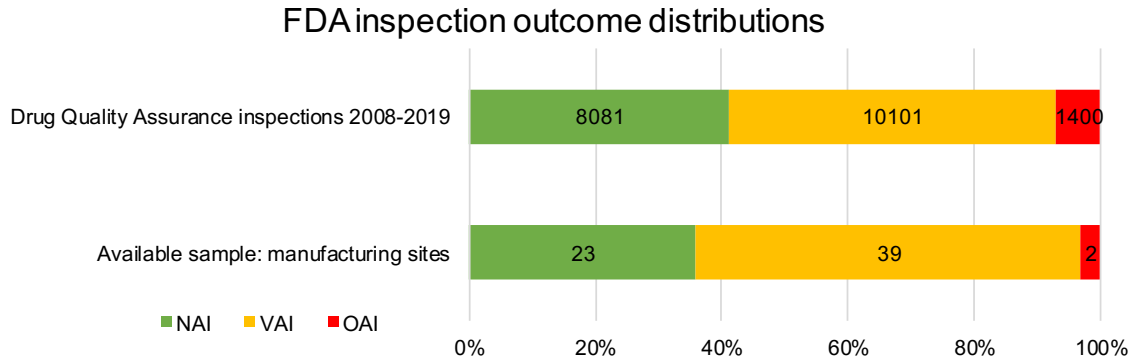


Figure 14: Matched FDA inspection outcomes for manufacturing database and overall population from 2008 –2019 (n=19582).

As discussed earlier, different operationalizations of inspection outcomes have been published in the academic literature (cf. Table 3). The given dataset is limited to inspection final decisions (OAI, VAI, NAI), which rules out operationalizations that require additional information like the issuance of a form 483. To overcome the challenge of small samples not meeting the required conditions for the planned analyses, the dependent variable is simplified to the binary categories where 1 represents NAI and 0 represents OAI or VAI. This means that VAI and OAI are combined into one target variable even though there are distinct. Whereas OAI certainly is a highly adverse inspection outcome and rather rare, VAI is quite common and less critical. Being aware of this limitation, OAI and VAI both nevertheless are notably more different from the ideal inspection outcome NAI. Also, this approach was chosen in alignment with a recent study describing predictive models for risk-based inspection planning (Seiss, 2018).

3.1.3 Methodology

In order to test the connection between enablers and inspection outcomes, a logistic regression model was chosen. The first model incorporated the overall OPEX enabler implementation as the only independent variable (IV). The second hypothesis was tested using a model with five IVs. These IVs are chosen in alignment with the five groups of enablers in the study that are: TPM, TQM, JIT, EMS, and BE.

Inspection outcomes are a complex measure and can depend on various factors that are not part of the scope of this research. To improve the reliability of the achieved results, the models include control variables for the size of the manufacturing sites and their location. To best represent the size of the manufacturing plant, the number of full-time equivalent employees was selected (Variable name: size). It is less sensitive to the type of production than floor area or production output. It is, however, dependent on the degree of automation. The site location is operationalized as a binary indicator of low or high-cost countries (LC). The countries China and India are the most important producers of pharmaceuticals with low-cost labor worldwide. The FDA stated that sites in these countries generally have a poorer history of quality compliance evaluations, and hence the analysis design controls for that factor (FDA - CDER, 2020b).

3.1.4 Results

The output of the first regression model is shown in Table 4. It is found that overall OPEX implementation is positively associated with favorable inspection results, such that the coefficient shows up to be statistically significant at the 5% level. The location in a country with low labor cost appears to have a strong negative effect on compliance evaluation, which was but significant at a 5% threshold. Offshore or low-cost locations have already been found to be of higher risk in previous research (Gray, Roth, and Leiblein, 2011). The model fit is evaluated using the change in deviance between the Null-model and the hypothesis model using the overall enabler implementation. Using the Chi-squared-characteristic, the deviance change (74.095-64.189) at three reduced degrees of freedom was found to be significant at a 5% level ($p = 0.019$). The model fit, therefore, is sufficiently good and provides support for the first hypothesis, whereas the control variables do not contribute at a significant confidence level.

Table 4: Logit regression output for overall OPEX implementation (n=57).

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-0.3947	0.3184	-1.24	0.2151	
Enabler Overall	0.709	0.3256	2.177	0.0295	*
LC	-1.899	1.143	-1.662	0.0965	•
Size	3.7E-05	9.03E-05	0.41	0.6822	
Significance		'*' 0.05		'•' 0.1	
Null deviance		74.095	55	Degrees of freedom	
Residual deviance		64.189	52	Degrees of freedom	

The model output for the individual OPEX practice category implementation is displayed in Table 5. It shows that TQM implementation is linked strongly positive to favorable inspection outcome with a significance level of 5%. The negative link between the 'Basic Elements'-enablers and inspection outcomes could be an effect of the collinearity but remains undetermined.

Table 5: Logit regression output for OPEX practice categories implementation (n=57).

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-4.75E-01	3.59E-01	-1.324	0.1853	
Enabler TPM	0.005197	0.5142	0.01	0.9919	
Enabler TQM	1.077	0.5361	2.009	0.0445	*
Enabler JIT	0.6262	0.5201	1.204	0.2286	
Enabler EMS	0.1323	0.561	0.236	0.8135	
Enabler BE	-0.9209	0.5161	-1.784	0.0744	•
LC	-2.267	1.289	-1.759	0.0786	•
Size	4.34E-05	8.41E-05	0.516	0.606	
Significance		'*' 0.05		'•' 0.1	
Null deviance		76.028	56	Degrees of freedom	
Residual deviance		62.136	51	Degrees of freedom	

The model fit is again assumed good with the Chi-squared-characteristic of the deviance change resulting in a p-value <1% (p = 0.0068). The estimate of the coefficient for TQM implementation is about 50% higher than for overall Operational Excellence implementation above.

Since the model is burdened with a relatively high number of coefficients compared to the available sample size, the non-significant coefficients were removed for an adapted model. Tested individually, the effect of Total Quality Management implementation remained approximately the same but showed up on an even higher level of significance with p < 1% (Table 6). Analogously, the evaluation of model fit increased as well (p = 0.0017). The control variables did not produce significant coefficients. The analyses thereby support hypothesis II and the positive effect particularly of TQM implementation on inspection outcomes.

Table 6: Logit regression output for Total Quality Management implementation (n=57).

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-0.4756	0.3418	-1.392	0.16403	
Enabler_TQM	1.044	0.3613	2.89	0.00386	**
LC	-1.838	1.158	-1.588	0.1123	
Size	3.88E-05	7.74E-05	0.501	0.61635	
Significance		‘*’ 0.05		‘**’ 0.01	
Null deviance		74.095	55	Degrees of freedom	
Residual deviance		58.925	52	Degrees of freedom	

In conclusion, the logistic regression analyses show links between bundles of OPEX practices and the manufacturing sites’ quality compliance in their FDA inspection outcomes with an emphasis on Total Quality Management practices. The control variables do not show any contributions above significance levels of 5%.

3.1.5 Conclusion

Sites that have a greater level of implementation of overall Operational Excellence programs have better inspection outcomes. This relationship is

stronger and highly significant for enablers specifically associated with the quality systems as measured in the enabler category Total Quality Management TQM. These results shall serve as motivation to industry and regulators to take OPEX and TQM maturity into account in inspection planning and preparation. From an academic perspective, this offers an additional perspective into the firm and its relationship with the regulators.

3.2 Manufacturing site performance metrics

Companies in all sectors have long recognized the importance of tracking the performance of their operations for more than 100 years. In the second half of the past century, the focus of that practice shifted from cost and efficiency towards effectiveness and quality measures. Nowadays, companies typically track performance indicators in several differentiated categories to monitor their enterprise or sub-systems, such as manufacturing sites. (Radnor and Barnes, 2007)

In the given context, such manufacturing and quality performance indicators are an obvious source of information for the analysis of quality risk on a manufacturing plant level. This chapter presents the motivation for the use of individual metrics, as well as performance aggregation, and outlines the research findings achieved based on the St. Gallen OPEX benchmarking program.

3.2.1 Hypotheses

As established in Chapter 2.2.3, the failure of a manufacturing site to comply with Good Manufacturing Practices is deemed Quality Risk. The specific quality risk for a given manufacturing site at a given time, in turn, is established in GMP inspections by the regulating agency. The anti-thesis of quality risk is the robustness of a production and quality system. As framed by Stricker and Lanza (2014, p.90), "*the robustness shall stabilize the systems' performance in case of varying conditions (risk)*".

While research publications often focus on maturity attributes and qualitative assessments of performance, the robustness construct is less studied. To make the robustness of production systems measurable, the afore-mentioned authors

argue that *"the KPIs have to be taken into account. Mostly, one KPI is not sufficient to describe the performance of a whole production system"* (Stricker and Lanza, 2014, p.91).

Regarding the practitioners' perspective, the regulators highlighted the importance of operational performance metrics. In their effort to better inform the risk-based inspection planning process (cf. Chapter 1.1), the FDA proposed two indicators related to manufacturing as part of their quality risk scores. The selected metrics for manufacturing and their rationale according to FDA - CDER[7] (2016, p.7) are:

- ***Lot Acceptance Rate*** *(LAR) as an indicator of manufacturing process performance.*
  LAR = the number of accepted lots in a timeframe divided by the number of lots started by the same covered establishment in the current reporting timeframe.
- ***Product Quality Complaint Rate*** *(PQCR) as an indicator of patient or customer feedback.*
  PQCR = the number of product quality complaints received for the product divided by the total number of dosage units distributed in the current reporting timeframe.

Aiming to answer SRQ 3 and testing a link between production system robustness and quality compliance risks, two hypotheses were proposed. Firstly, *manufacturing plants with higher performance levels in their aggregated performance are likely to receive favorable regulatory inspection results (III).*

The analyses of single Key Performance Indicators (KPI) allow a much more detailed perspective into the relationship between operational performance, and quality compliance. It also allows several different options for fitting statistical models. If successful, such models could then be used by regulators and corporate functions to estimate risk levels, and take appropriate action. Therefore, the second hypothesis claims that *individual indicators of manufacturing and quality performance can be predictive of GMP inspection outcomes (IV).*

[7] CDER stands for the FDA's *Center for Drug Evaluation and Research*

3.2.2 Empirical Set-up

This research is based on the same underlying benchmarking study as presented in chapter 3.1, with a more detailed description in paragraph 3.1.2. The database contains numerous performance metrics that describe the effectiveness and efficiency of the manufacturing process. Accordingly, these metrics are assigned to the three categories: Total Productive Maintenance, Total Quality Management, and Just in Time. To identify potential indicators for quality compliance failures, this research relies on earlier work that assessed the performance and robustness of pharmaceutical production. Particularly, the model for Operational Excellence in pharmaceutical manufacturing (Figure 12) in Friedli, Goetzfried and Basu (2010), Friedli *et al.* (2013), and Friedli, Köhler, *et al.* (2018) shall be referred to. The metrics included in the tested Operational Excellence performance constructs are displayed in Table 7.

Table 7: Operational Excellence performance scores and assigned metrics.

Technology Performance Score		
Total Productive Maintenance Score	**Total Quality Management Score**	**Just-in-Time Score**
Maintenance Cost/Conversion Cost	Quality Cost per Conversion Cost	Production Schedule Accuracy
Overall Equipment Effectiveness	Deviation per Batch	On-Time In Full delivery rate
Setup and Cleaning	Deviation Closure Time	WIP Turns
Unplanned Maintenance	Complaint Rate Supplier	Raw Material Turns
	Scrap Rate	Finished Goods Turns
	Rejected Batches	Order Lead Time
	Right First Time	Forecast Accuracy
	Yield	Replacement Time to Customer
	Complaint Rate Customer	Changeover Time
		Production Lead Time

In addition to the performance focus of the above systematics, similar operational performance indicators can also be rearranged to represent the effectiveness of the Pharmaceutical Quality System (PQS) as introduced by

Friedli *et al.* (2017). Such effectiveness in turn represents robustness against risk. Therefore, it is employed as a supplementary perspective on the aggregated performance levels. Again, this research builds on an established construct, the Pharmaceutical Production System Model (PPSM) shown in Figure 15.

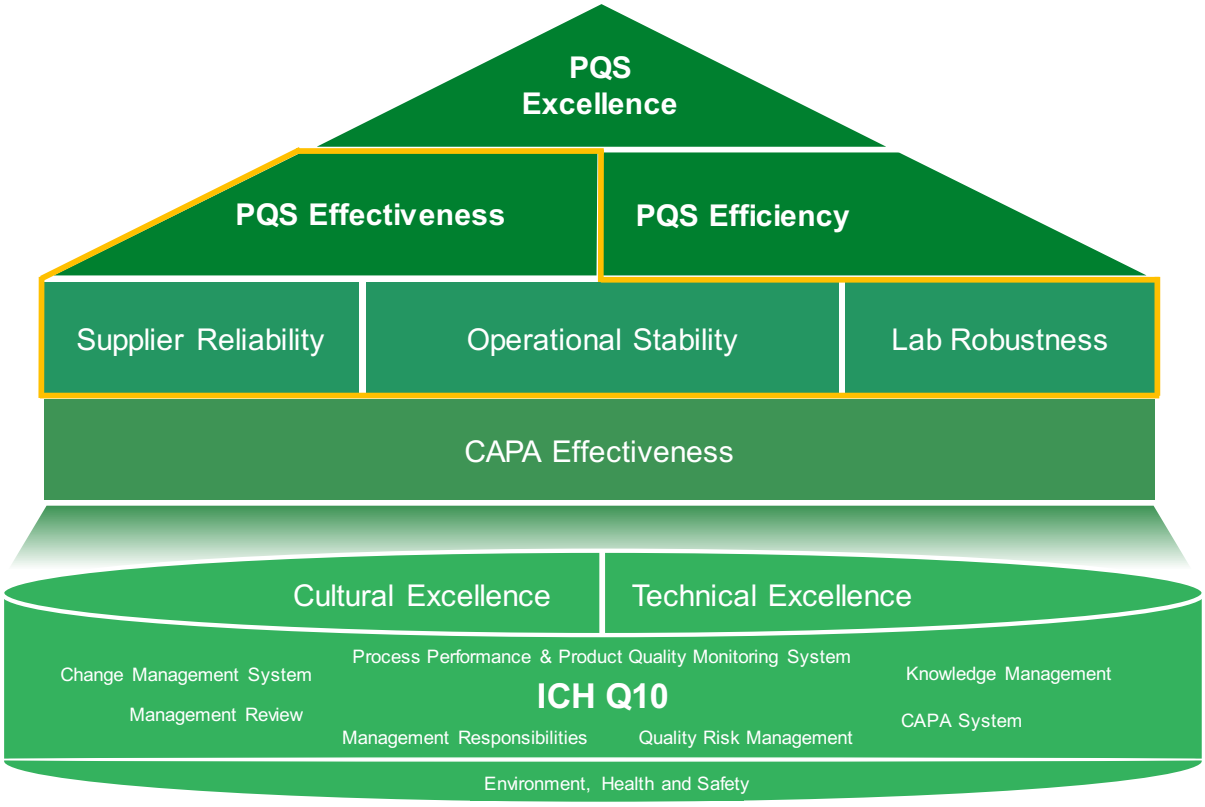


Figure 15: St.Gallen Pharmaceutical Production System Model. Illustration adapted from *Friedli* et al.*, 2019.*

Of the different parts that make up the PPSM, all items above “CAPA Effectiveness” are operationalized as aggregations of performance indicators. “Supplier Reliability”, “Operational Stability”, and “Lab Robustness” are aggregations of performance indicators of the manufacturing site or the QC lab, respectively. PQS Effectiveness is the aggregation of the former two elements. “PQS Efficiency” is based on costs per batch produced. It is excluded in these analyses due to lacking data availability. Similarly, “PQS Excellence” is excluded, as it is an aggregation of the two elements mentioned below with low data availability. The remaining constructs that are included in the analyses have been highlighted in yellow.

The process of mapping the two databases as well as the binary operationalization of the FDA inspection outcomes are identical to Chapter 3.1. The data availability varies for individual KPIs as there is a challenge of incomplete datasets in the performance section. The number of data pairs has is provided along with the individual analyses in Chapter 3.2.4.

3.2.3 Methodology

KPIs and performance scores were tested in individual logistic regression models with the inspection outcome as the target variable and jointly with the control variables. Metrics identified as significant predictors for inspection outcomes then were evaluated in combined models. This aims to examine their redundancy or the cumulative effects of the coefficients.

3.2.4 Results

Of all the tested coefficients and logistic regression models, only the outputs of models with significant links are presented below. Significant results for three distinct metrics in terms of the coefficient strength and model fit were found, while the performance aggregations were not statistically significant. These results are highlighted in the following paragraphs and tables, sorted by the model fit. In cases of notable differences between the models with and without control variables, both variants are included.

The strongest coefficient found belongs to the *On Time In Full delivery rate* (OTIF). Manufacturing sites that continuously deliver based on the exact specifications of their customers, receive fewer inspections with regulatory actions indicated. The metric showed a highly significant link ($p<1\%$) to the target and the model also showed a very good fit ($p = 0.0026 <1\%$). The full model summary is represented in

Table 8.

Table 8: Logit regression output for OTIF delivery performance (n=48).

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-0.79268	0.63958	-1.239	0.21521	•
OTIF	1.14024	0.391541	2.912	0.00359	**
LC	-1.23107	1.271491	-0.968	0.33294	
Size	0.000273	0.000788	0.346	0.72918	
Significance		'*' 0.05		'**' 0.01	
Null deviance		61.513	46	Degrees of freedom	
Residual deviance		47.296	43	Degrees of freedom	

The robustness of manufacturing processes, which reflects their ability to contain any variances well within specification limits, can be assessed through the *Lot Acceptance Rate.* In a state of control, the production process will deliver product lots consistently according to their specifications. This metric therefore is very closely linked to the 6σ goals set by companies, minimizing process variance against the defined accepted tolerances of production process parameters. As shown by the analyses summarized in Table 9, production establishments, higher *Lot Acceptance Rates* tend to receive more inspections without objections. The coefficient shows a positive impact significant at the p<5% level in the logistic regression model at the same confidence level for its fit (p = 0.03 <5%).

Table 9: Logit regression output for Lot Acceptance Rate (n=53).

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-0.4756	0.3418	-1.392	0.16403	
Lot Acceptance Rate	1.044	0.3613	2.89	0.00386	**
LC	-1.838	1.158	-1.588	0.1123	
Size	3.88E-05	7.74E-05	0.501	0.61635	
Significance		'*' 0.05		'**' 0.01	
Null deviance		72.087	52	Degrees of freedom	
Residual deviance		61.364	48	Degrees of freedom	

The *Customer Complaint Rate* captures the satisfaction of the direct downstream stakeholders with the production output. Depending on the position in the value chain and specific manufacturing output of the given facility, this can include different constellations. The customers of a drug substance production site can be processing facilities at the same or different locations owned by the same company or external organizations. Similarly, finished drug product manufacturing can have customers that are warehouses, wholesale, or brand pharmaceutical companies in case of CMO production. Regardless of the individual setting, the metric captures the share of orders that the customer was not satisfied with and had filed a complaint.

This ability to deliver products according to the customers' demand does also link to the quality compliance evaluations. This means that manufacturing sites that deliver with lower *Customer Complaint Rates* receive fewer unsatisfactory inspections from the FDA. The impact of the coefficient nears the established confidence interval and stands at about p<7% (cf. Table 10). Again, the fit of the overall model similarly stands at <6% (p = 0.058). While this result does not quite meet the established 95% confidence threshold, it is notably close. Therefore, it is presented here and included in the next analysis step.

Table 10: Logit regression output for Customer Complaint Rate without control variables (n=50)

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-0.616	0.3085	-1.996	0.0459	*
Customer Complaint Rate	0.5779	0.314	1.841	0.0657	•
Significance		'*' 0.05		'•' 0.1	
Null deviance		65.342	49	Degrees of freedom	
Residual deviance		61.736	48	Degrees of freedom	

Tested in combination with the control variables, the hypothesized effect approximately remains at the same strength, but the significance drops notably as well as the model fit. The full model summary is included in the following

Table 11.

Table 11: Logit regression output for Customer Complaint Rate with control variables (n=50)

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-0.5713	0.3422	-1.669	0.095	•
Customer Complaint Rate	0.5372	0.328	1.638	0.101	
LC	-1.032	1.183	-0.872	0.383	
Size	4.1E-05	0.000209	0.196	0.844	
Significance		'*' 0.05		'•' 0.1	
Null deviance		65.342	49	Degrees of freedom	
Residual deviance		59.828	46	Degrees of freedom	

With the individual links between quality compliance, and the three metrics *Lot Acceptance Rate*, *Customer Complaint Rate*, and *On-Time In Full delivery rate* (OTIF) being established, their inter-dependency is of further interest. In theory, the shared predictive power of the metrics could be attributed to all of them being surrogates for the same phenomenon. In the given case, all of them could be a proxy for overall process stability and therefore their predictive power redundant. In order to assess this, they were further examined in two steps.

First, the correlations between the individual variables were evaluated. The correlation coefficients, histograms, and scatter plots are shown in Figure 16. There are no significant correlations present, which rules out the existence of collinearity between the independent variables of the previous analyses. Therefore, it may be assumed, that they each possess predictive power on the inspection outcomes, and do not serve as a proxy for the same underlying phenomenon.

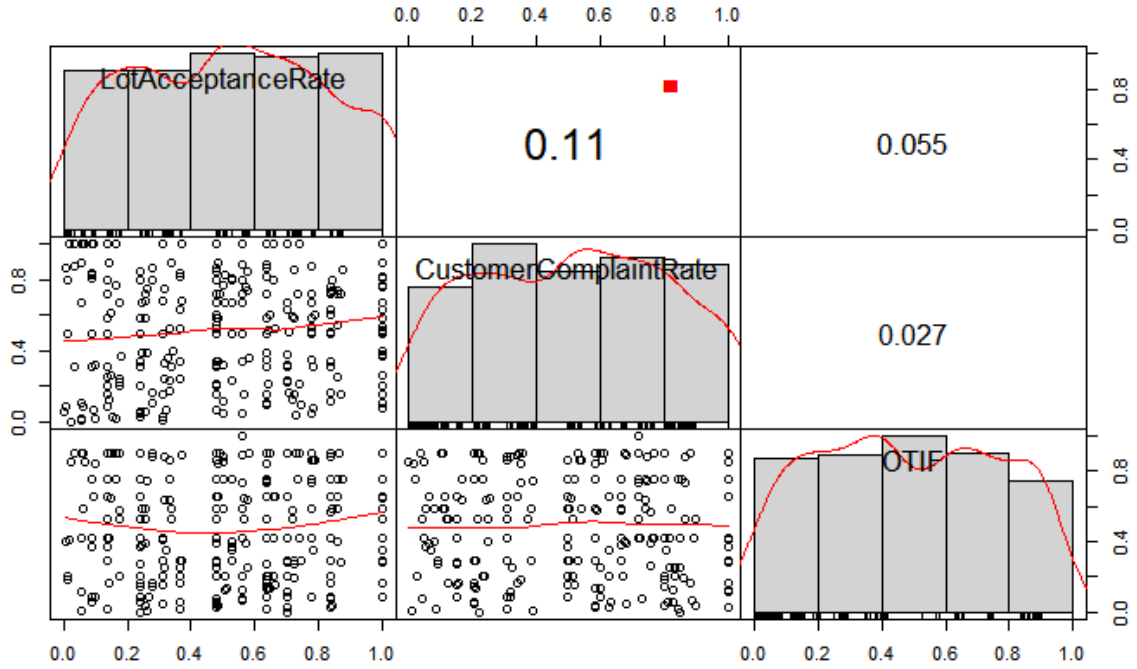


Figure 16: Overview and inter-correlation of identified manufacturing KPIs

As a second step, a combined model tested whether the combination of the predictor variables achieves a better representation of inspection outcomes as the target variable. The model was set up as logistic regression, analogous to the previous iterations, but with three independent variables. Table 12 presents the model output.

Table 12: Logit regression output for combined variables: OTIF, Lot Acceptance Rate, and Customer Complaint Rate (n=36)

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-0.58	0.5296	-1.095	0.27344	
Customer Complaint Rate	1.9787	0.7747	2.554	0.01064	*
Lot Acceptance Rate	0.705	0.6003	1.174	0.24022	
OTIF	2.4054	0.8475	2.838	0.00454	**
Significance		'*' 0.05		'**' 0.01	
Null deviance		49.461	35	Degrees of freedom	
Residual deviance		25.166	32	Degrees of freedom	

The model summary shows OTIF and Customer Complaint Rate as significant or highly significant coefficients, respectively. All coefficients link positively to favorable inspection outcomes. The most notable aspect is the strong reduction of deviance in the residual model. This means, that the model enables a much better representation of the target value when incorporating the combined coefficients. Computed into a model fit evaluation for the regression model, the p-value is smaller than 1‰ (p=2.17e-05). Thereby, the model shows a much better fit than the individual models using only one of the variables, and the combination of coefficients appears to be valuable for the prediction.

When tested in combination with the control variables, the model produces somewhat different results. Customer Complaint Rates showed up as a highly significant coefficient. In contrast to the previous model, OTIF is not a significant coefficient anymore, while Lot Acceptance Rate (LAR) becomes a statistically significant variable. The full model summary is presented in Table 13.

Table 13: Logit regression output for combined variables with control: OTIF, LAR, Customer Complaint Rate and control variables (n=36)

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-0.4983	0.5535	-0.9	0.36797	
Customer Complaint Rate	2.362	0.8377	2.819	0.00482	**
Lot Acceptance Rate	2.008	0.8107	2.477	0.01326	*
OTIF	0.7606	0.6148	1.237	0.21606	
LC	-2.196	2.618	-0.839	0.40152	
Size	2.68E-05	7.98E-05	0.336	0.73695	
Significance		‘*’ 0.05		‘**’ 0.01	
Null deviance		49.461	35	Degrees of freedom	
Residual deviance		25.166	32	Degrees of freedom	

The changes in the significance of coefficients between the different models suggest an interplay between Customer Complaint Rate and other test and control variables. The strong changes in results could also, at least in part, be

attributed to the relatively low number of observations in contrast to included coefficients. This effect is due to missing and overlapping data between the indicators, which led to several case exclusions.

3.2.5 Conclusion

Manufacturing sites that show better performance in some operational KPIs are likely to receive more favorable inspection outcomes. More specifically, sites that provide more reliable delivery service to their customers in terms of quantity, quality, and timeliness, appear to be more compliant with GMP regulations. A similar trend, yet not as highly significant, exists for manufacturing sites that receive fewer complaints from their customers. Thirdly, manufacturing sites that show better control of their manufacturing processes and therefore have fewer faulty production lots are evaluated better in their FDA inspections. The combination of all three indicators does deliver optimum model fit. Therefore, all three of them add valuable information relevant to the state of quality compliance risk of a manufacturing site.

3.3 QC Lab OPEX maturity analyses

Quality Control (QC) Labs are an essential part of pharmaceutical value chains and may sometimes represent a system bottleneck (Köhler, 2019). The test results of drug products, substances, and components are the basis for the release of only high-quality drugs into the market. Timeliness and reliability of such tests, as well as diligent handling of their results, therefore, are a critical prerequisite for the compliant supply of pharmaceuticals. This fundamental importance and the problems linked to it are well known to the FDA. The QC Labs were the target of a large share of the warning letters within the industry in recent years (The FDA Group, 2019). At the same time, companies strive to strengthen Operational Excellence practices in their QC labs (Köhler, 2019).

3.3.1 Hypotheses

Analogously to Chapter 3.1.1, the author hypothesizes that *plants with QC labs that are further advanced in their OPEX implementation are more likely to receive favorable inspection outcomes* (V). Again, a *stronger relationship for (Maintenance and) Quality System implementation* (VI) is expected. Furthermore, the next *hypothesis proposes that there is a negative link between*

the labs' implementation of 'Maintenance and Quality Systems' practices and the number of lab-related citations issued in an inspection (VII). Therefore, more advanced quality practices in QC labs should result in fever GMP violations documented by the regulator.

3.3.2 Empirical Set-up

This research utilizes a database of pharmaceutical QC lab operations data. It currently consists of data on 95 QC labs. The data is gathered in a process that is identical to the study underlying Chapter 3.1. The study has been ongoing in the calendar years 2015 to 2019. The distribution of participating establishments (see Figure 17) is similar to the one for manufacturing sites, but slightly less biased towards European locations.

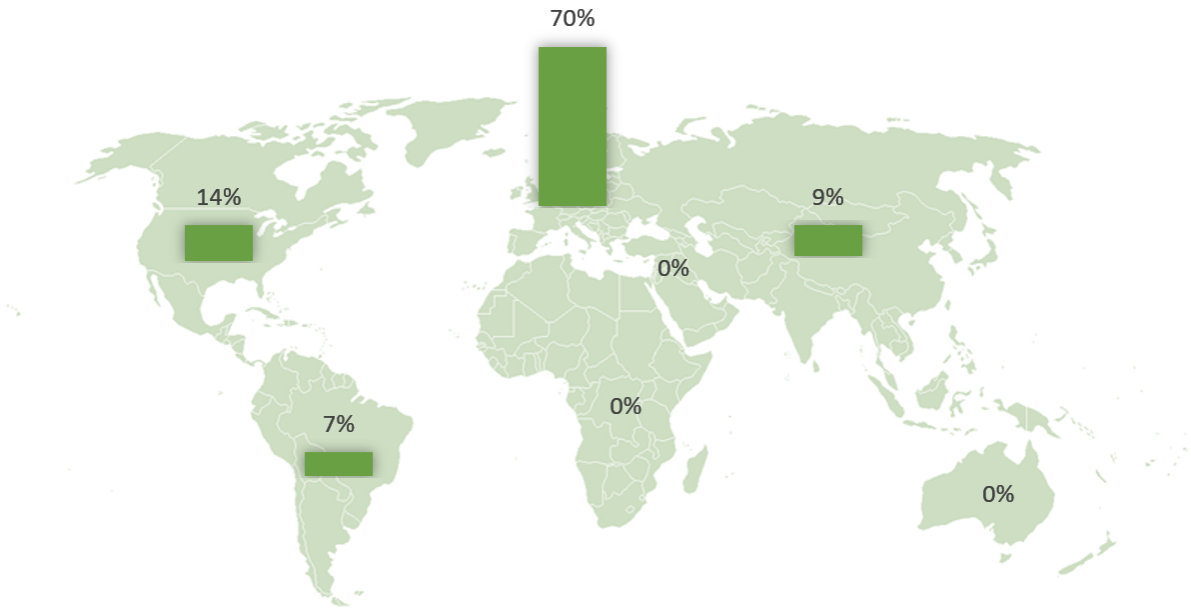


Figure 17: Geographical distribution of the available QC lab datasets

Similar to the manufacturing database, each dataset for the QC labs consists of structural information, enabler implementation, and performance metrics. The enabler data is clustered into three categories: Maintenance & Quality System, Planning & Steering System, and Management System. Analogously to Figure 13, Figure 18 shows the descriptive statistics for the enabler groups to describe Operational Excellence implementation in QC labs.

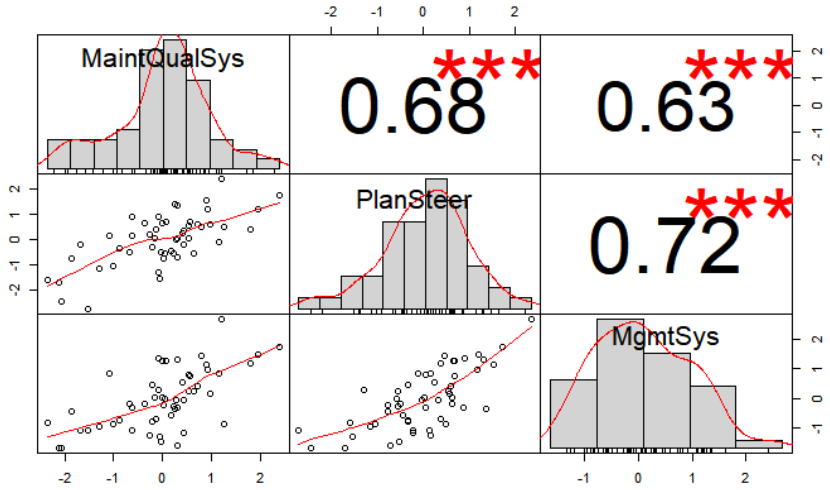


Figure 18: Enabler groups in QC labs: Exploratory data visualization (Significance levels: *** 0.001, ** 0.01).

The datasets then were matched to FDA inspection outcomes. In this database, 57 pairs of lab and inspection data were identified. Figure 19 depicts the resulting distribution of inspection outcomes and the comparison to the overall sample. The given sample does not contain any OAI inspection result but nevertheless is similar in its distribution.

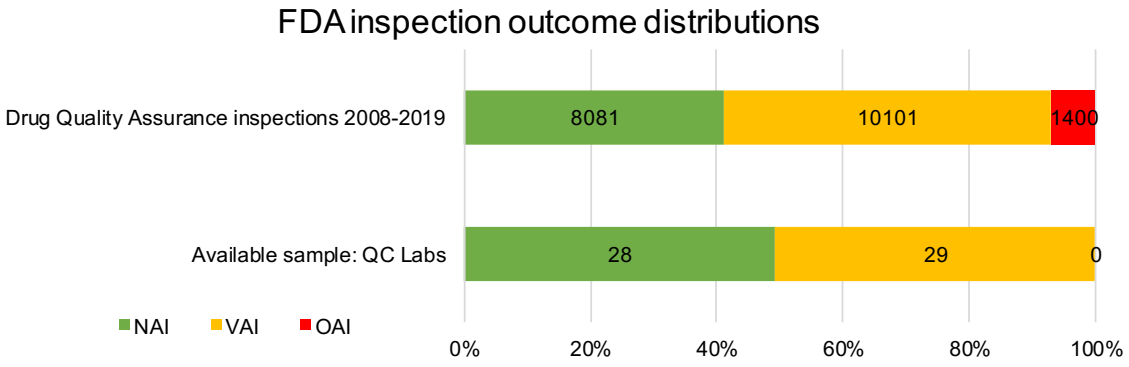


Figure 19: Matched FDA inspection outcomes for QC lab database and overall population from 2008 – 2019 (n=19582).

The FDA inspections target the overall manufacturing establishment. Therefore, a possible link between lab-specific attributes and inspection outcomes might is not certain. To further investigate the extent to which labs influence site inspection outcomes, the individual citations may be used as a viable source of information. Citations are the observations found during an inspection and later

cited in forms FDA 483. Thus, they provide additional details on an establishment's quality compliance. The citations can be found in the FDA's Inspections Citations Details Database, where a short and a long description of the citation as well as the corresponding policy sections are listed. (FDA - CDER, 2020a)

By combining the two databases, not only the inspection outcome but also the deficient principles related to the QC lab were considered. For all the matched inspection outcomes, the citations were available for only 15 of them. For these inspections, 72 citations were available in total. These were examined for whether or not individual citations are directed at violations in QC labs. Directly lab-related citations and weaker, partial relations were distinguished. The evaluation was done by independent, blinded reviews of two researchers in juxtaposition. Any deviation between the evaluations was then examined closely, discussed, and then jointly assigned a weighting factor. In conclusion, only the minority of citations connect to QC labs (cf. Figure 20). In their review, the FDA links about 20% of total citations to the QC laboratories, wherein data integrity issues were assigned a separate category and therefore are not included in that figure (FDA - CDER, 2020b).

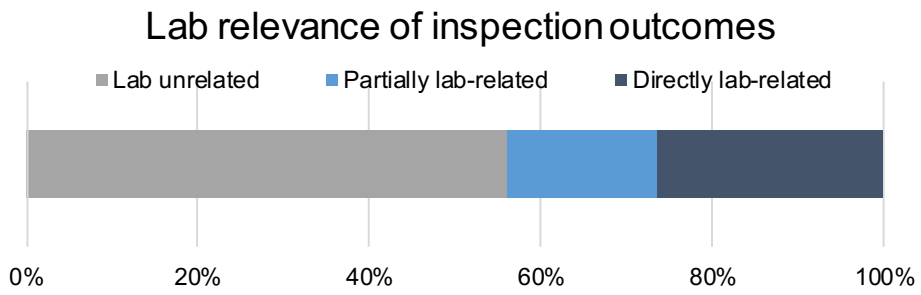


Figure 20: Analysis of inspection citations and their connection to QC labs

The weighted number of citations related to the lab was used as an alternative target variable for analyses with the QC data. To obtain the metric, directly lab-related citations were weighted at 1, and partially lab-related citations at 0.5. The resulting near-normal distribution of citations per inspection is shown in Figure 21. That visual impression was confirmed by a Shapiro-Wilk normality test p-value of 0.514.

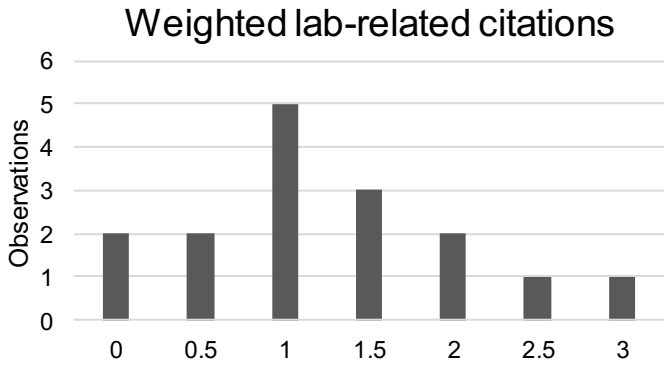


Figure 21: Distribution of lab-related citations in FDA inspections

3.3.3 Methodology

A logistic regression model was chosen to test the connection of the enablers on the inspection outcome. The first model uses the overall enabler implementation as only IV. The second hypothesis was tested using a model with three IVs. These IVs were chosen in alignment with the three groups of enablers in the study which are: Maintenance & Quality System (MaintQualSys), Planning & Steering System (PlanSteerSys), and Management System (MgmtSys), respectively. In addition to the control variables introduced in Chapter 3.1.3, the number of FTE employees in the QC labs was added (SizeQC).

In another attempt to test Hypothesis V, the possible connection between the quality enabler implementation and lab-specific violations documented by the FDA is considered. It is analyzed in context to possible predictors through linear regression. Due to the low number of observations, only a model with a single input at a time was feasible. Therefore, any predictor was analyzed in a separate model. For the sake of completeness, the same approach was taken for all control variables. Models with non-significant results were not included in the table.

3.3.4 Results

Table 14 summarizes the output of the two logistic regression models for QC labs. OPEX enabler implementation, as well as the three enabler categories, did not have an impact on observed inspection outcomes. Therefore, hypotheses V and VI were rejected. On the other hand, it was observed that establishments with larger QC labs received adverse inspection outcomes slightly more often.

The control variable was significant at the 5% level of confidence in both models. This relationship remains undetermined at this time. Potential hypotheses could be that larger QC lab organizations represent more complex testing operations or that the organizations grew as a response to process instability. The model fit was significant at the 5% level for the first model (p = 0.037), but slightly inferior for the second model (p = 0.076).

Table 14: Logit regression output for lab-specific enablers and inspection outcomes (n=56)

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-0.80838	0.550821	-1.468	0.1422	
Enabler Overall	-0.26715	0.311529	-0.858	0.3911	
LC	-1.92872	1.339323	-1.44	0.1498	
Size	-0.00138	0.000878	-1.575	0.1151	
SizeQC	0.023135	0.009499	2.436	0.0149	*
Significance		‘*’ 0.05			
Null deviance		76.082	54	Degrees of freedom	
Residual deviance		65.892	50	Degrees of freedom	

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	-0.85531	0.561454	-1.523	0.1277	
MaintQualSys	-0.31644	0.449636	-0.704	0.4816	
PlanSteerSys	0.387418	0.493178	0.786	0.4321	
MgmtSys	-0.44261	0.458441	-0.965	0.3343	
LC	-2.10589	1.363387	-1.545	0.1224	
Size	-0.00121	0.000877	-1.377	0.1685	
SizeQC	0.02228	0.009462	2.355	0.0185	*
Significance		‘*’ 0.05			
Null deviance		73.304	52	Degrees of freedom	
Residual deviance		71.846	49	Degrees of freedom	

The lack of a statistical link between the Operational Excellence implementation in QC labs specifically and overall site inspection is understandable. The inspections have a much larger scope than the QC labs themselves. Thus, hypothesis VII aims to analyze the connection between the 'Maintenance & Quality System' enablers in the lab and the inspection citations directed at it. Table 15 displays the corresponding analysis results.

Table 15: Linear regression output for lab quality enablers and lab-related inspection citations (n=15)

Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)	Significance
Intercept	1.2317	0.1586	7.764	3.1E-06	***
MaintQualSys	-0.6252	0.2324	-2.69	0.0185	*
Significance		'*' 0.05		'***' 0.001	
Multiple R-squared		0.3576	F-statistic	7.237	on 1 and 14 DF
Adjusted R-squared		0.3082	p-value	0.01854 *	

The analysis shows a notably significant fit of the overall model and negative impact of "Maintenance and Quality System" enablers on the weighted lab-related citations in FDA inspections. With the significance level of the coefficient and the overall model, it is reasonable to assume the existence of a link between the maturity of the labs and their evaluation by the FDA. At about 30%, the R-squared values are reasonably good in the context of complex socio-technical systems, which strongly depend on human behavior[8]. Nevertheless, a relatively large part of the variation in the compliance evaluation of QC Labs was not determined by the tested variable. Finally, the relatively low number of underlying observations is highlighted here as a limitation to the robustness of that result. In conclusion, the analysis provided limited support for hypothesis VII.

Tested in separate models, the other enabler categories, and control variables did not show a significant link to the number of lab-related violations cited by the FDA.

[8] Frost (2019) points out that studies involving human behavior generally have maximum R-squared values below 50% since it is harder to predict than e.g. physical processes.

3.3.5 Conclusion

No link between Operational Excellence implementation in quality control labs and site inspection outcomes could be established. However, there appears to be a link between lab-specific enabler implementation and the deficiencies detected by regulators. Recently FDA has dedicated a lot of attention to QC labs (The FDA Group, 2019). This finding shall motivate regulators to keep the overall scope of manufacturing sites in mind and not to narrow their focus. Industrial managers must be aware of the substantial influence of Operational Excellence implementation on the quality compliance evaluation. However, they are best advised to refrain from silo optimization.

3.4 QC Lab performance metrics analyses

As established in Chapter 3.3, quality control laboratories are of great importance for the delivery of high-quality drugs. Hence, it is the center of attention of regulatory agencies, which frequently finds deficiencies in these organizational units. Besides inspections, such deficiencies in laboratory processes can also show up in the operational performance metrics. The ability of a manufacturing lab to deliver accurate test results when needed can be captured in their quality and service or delivery metrics (Köhler, 2019). Alternatively, such metrics can also be designated to quality control laboratory robustness (Friedli *et al.*, 2019).

Many of the associated deficiencies highlighted in FDA inspections specifically concern the handling of test results (Curtis, 2020). Two prominent possible failure modes include the inability of companies to establish the method in which the positive test results were achieved in the first place, which were subsequently used to legitimize the release of the product. On the other hand, some companies at times invalidate negative test results without providing proper scientific reasoning (The FDA Group, 2019; Curtis, 2020). The general attention to labs, as well as the aforementioned issue have been reflected by the FDA with the inclusion of the “invalidated-out-of-specification-rate” (IOOS) in their three designated reporting metrics (FDA - CDER, 2016):

- ***Invalidated Out-of-Specification*** *(IOOS)* ***Rate*** *(IOOSR) as an indicator of the operation of a laboratory.*

IOOSR = the number of OOS test results for lot release and long-term stability testing invalidated by the covered establishment due to an aberration of the measurement process divided by the total number of lot release and long-term stability OOS test results in the current reporting timeframe.

IOOS is also included in the given dataset as a performance metric. The nominator for that ratio is the number of test results that are out of specification but ruled out for reasons elaborated and documented by the company. There are varying definitions for the denominator, which are described in the footnote[9].

3.4.1 Hypotheses

Similar to the discussion presented in Chapter 3.2 in the context of manufacturing sites and inspection outcomes, this chapter *hypothesizes that labs with better performance metrics also are evaluated more favorably by the regulator, i.e., receive fewer citations (VIII)*. The performance metrics that were established to be relevant for the assessment of the robustness of the processes in pharmaceutical QC labs (Friedli, Köhler, *et al.*, 2018), and are closely related to SRQ 3. Special emphasis has been placed on the "invalidated-out-of-specification-rate" (i.e. IOOS proportion[9]) due to the increasing interest from practitioners and regulators following the FDA draft guidance metrics (IX).

3.4.2 Empirical Set-up

This research is based on the same underlying benchmarking study as in Chapter 3.3 with a detailed description being presented in Paragraph 3.3.2. The database contains numerous performance metrics that describe the effectiveness and efficiency of the quality control procedures. Accordingly, these metrics are assigned to the four categories Quality, Service, Productivity, and Cost. To identify potential indicators for quality compliance failures, this research relies on earlier work that assessed the performance and robustness of quality

[9] The benchmarking study, industry and FDA use different denominators for the calculation of this metric. Dividing IOOS tests by overall number of tests is a measure for process stability, as such a test result means additional work for the lab. Using the share of IOOS of all initial OOS instead can be interpreted as an indicator of how seriously a company takes such results. This could potentially in some cases unveil a tendency of valuing efficiency over quality. To avoid ambiguity of the terms, the latter ratio from here on is referred to as "IOOS proportion".

control labs. Particularly, Friedli, Köhler, *et al.* (2018), Friedli *et al.* (2019), and Köhler (2019) worked on the mentioned issue and proposed corresponding performance indicators. Table 16 presents the metrics included in these analyses.

Table 16: QC lab operational excellence performance metrics and scores[10].

Lab Performance Score			
Quality Score	**Service Score**	**Productivity Score**	**Cost Score**
Lab Deviations/1'000 Tests	Adherence to Schedule	Handled Samples/QC FTEs	QC Costs/QC (D+I) FTEs
Reoccurring Deviations	Adherence to Lead Time	Batches processed/QC FTEs	QC Costs/Total No. of Tests
Product Re-Tests due to Complaints	Release Time	Tests/Direct QC FTEs	
Right First Time	Lead Time		
Investigations/1'000 Tests	Cycle Time		
IOOS/100'000 Tests			
CAPAs Overdue			
Lab Robustness Score[11]			

In addition to the individual metrics, scores for the various performance dimensions were examined, too. These scores are based on an equally weighted aggregation of the normalized metrics assigned to them[11]. For the development of these scores, the previously mentioned sources were used as a reference.

[10] Where applicable, several different versions of the metrics were tested. In addition to metrics using all tests, tests for releasing drug product (DP) and stability testing were specifically assessed separately. These are of particularly importance due to their immediate connection to drugs that reach the market and patients or already are in the market, respectively.

[11] The Lab Robustness Score consists of all metrics assigned to the Quality category, but only the two adherence-related figures in the Service category. Through the equal weighting of all metrics, this emphasizes the importance of Quality aspects.

3.4.3 Methodology

The statistical test methodologies performed are identical to the description in Sub-section 3.3.3.

3.4.4 Results

Similar to the previous analyses of the OPEX implementation in QC labs as presented in Sub-section 3.3.4, there is no statistical link between any of the analyzed metrics and the final outcome of an inspection. However, notable results for lab-specific citations were found. Sites that invalidate a higher share of their out-of-specification tests are likely to receive more citations against their lab operations. The relationship is particularly strong for tests performed for drug release which shows strongly significant positive coefficients. The p-values for the model fit are significant ($p < 5\%$) and even stronger for the second model. Table 17 lists the full results of these two configurations of linear regression models.

Table 17: Linear regression output for IOOS proportions and lab-related inspection citations (n=16).

Coefficients

	Estimate	Std. Error	t value	Pr(>|z|)	Significance
Intercept	1.54	0.1941	7.934	2.45E-06	***
IOOS Proportion	0.4721	0.1903	2.481	0.0275	*
Significance		'*' 0.05		'***' 0.001	
Multiple R-squared		0.3214	F-statistic	4.53	
Adjusted R-squared		0.2692	p-value	0.02754 *	

Coefficients

	Estimate	Std. Error	t value	Pr(>|z|)	Significance
Intercept	1.5438	0.1867	8.267	8.82E-06	***
IOOS Proportion DP	0.5699	0.1827	3.119	0.0109	*
Significance		'*' 0.05		'***' 0.001	
Multiple R-squared		0.4932	F-statistic	6.157	
Adjusted R-squared		0.4425	p-value	0.01089 *	

Of the other metrics that were evaluated, none showed up as a significant coefficient or produced a significant model fit.

3.4.5 Conclusion

As with enabler implementation in QC labs, there is no link between the performance of QC labs and the evaluation of the overall site in inspection outcomes. Of all the metrics and performance dimensions tested, the share of invalidated out-of-specification test results shows up as a predictor for negative evaluations of QC labs. This trend is particularly driven by tests related to drug product release, which shows the strongest statistical link. This finding supports the FDA in their identification of IOOS proportions (IOOS rate as defined in FDA - CDER, 2016) as a meaningful metric to inform risk rankings.

3.5 Consolidated indicator metrics for regulatory inspection outcomes

Based on the shown analyses and the literature review (cf. Table 3), several factors can be concluded to be influential on or at least indicative of adverse inspection outcomes. From the literature, it is evident that organizational factors such as outsourcing or offshore production link to a higher risk of adverse inspection outcomes. In contrast to that, manufacturing sites that are co-located with central company R&D perform better in this regard. Similarly, language differences between manufacturing sites and headquarters are harmful to quality compliance, as well as cultural differences in terms of power distance.

When focusing on the actual operation, on-site heterogeneity of products and processes increase the risk of adverse inspections for contract manufacturers. Close oversight of the customer company can mitigate that risk to some extent. Non-compliance in financial reporting does also indicates quality non-compliance in manufacturing sites. Manufacturing sites that are ISO 9000 certified show better results in regulatory inspections. The effect is even stronger for early adopters of the certification. However, the effect of the certification does deteriorate over time. Lastly, the time duration between inspections also links to the outcome of an inspection, with longer durations having more chances of producing a negative inspection outcome. Thereby, a phenomenon of decaying state of quality compliance can be established.

This thesis confirms findings in existing literature in all aspects of the analyses that overlapped. These include the localization of manufacturing sites in low-cost countries and longer periods yielding negative outcomes to inspections. Also, advanced Quality Management practices on the other hand are indicative of a better state of Quality Compliance. The research extends this finding by separating specific practices related to Total Quality Management under the overall umbrella of Operational Excellence, which is often used in pharmaceutical companies. Furthermore, there is a link between the implementation of Quality Management practices in Quality Control laboratories and Quality Compliance deficits. However, these do not necessarily reflect decisively on the inspection outcome of the overall manufacturing site.

The link between operational performance metrics and Quality Compliance forms the central crux of this thesis. It is a unique contribution to the academic discussion and will add substantially to practitioners' discourse. All anchored in a conceptual perspective, the delivery service level (OTIF), Lot Acceptance Rate, and Customer Complaint Rate are indicators of adverse inspection outcomes. Similarly, the Invalidated Out-of-Specification Rate is a lab-specific indicator linked to compliance failures in Quality Control labs.

4. Industry project case study with predictive analyses

Data access and collection in the context of manufacturing site internal information, corporate quality functions, and external regulators is challenging. Therefore, the research team is grateful for the opportunity to access the proprietary data of one of the world's largest pharma companies. Predictive models typically require longitudinal data to be trained. This access to longitudinal data around operational KPIs, internal quality audits, and regulatory agency inspection outcomes allows deepening and extending the findings of the previous chapters. More specifically, the findings from literature research and analyses of the cross-company data can be challenged using this independent dataset. Building on that, the predictive powers of operational performance metrics for inspection outcomes are tested in traditional regression and machine learning models of varying complexity.

The chapter begins with an introduction of the partnering company and its approach to Quality Management. Secondly, an outline of the collaboration project that forms the foundation of the research shown in this chapter is provided. Section 4.3 describes the available dataset with special regards to the inspection outcomes, their comparability, and normalizations. The following sub-chapter provides descriptive analyses of the dataset and connects to the previous chapter, as well as sets a baseline for the predictive analyses. These are the key content of Section 4.5, where the methodologies are introduced and results are presented. The last section in this chapter concludes the findings and reflects upon them through a discussion with the project partner and a panel of industry experts.

4.1 Introduction to case study industry partner

The company we worked with is a global leader in the pharmaceutical industry and ranks within the top five companies in terms of sales volume. Based in Europe, it employs more than 100,000 employees and is present in around 100 countries. The company places a strong emphasis on maintaining a healthy balance between in-house production, and the management of external manufacturing operators. It employs more than 30.000 employees at more than 50 manufacturing sites distributed all around the globe. In 2019 alone, the company distributed close to 5 billion doses of pharmaceuticals, consumer healthcare products, and vaccines. The pharmaceutical products can be

allocated to several major treatment areas such as oncology, cardiovascular, and central nervous system diseases.

The company established a holistic Quality System that includes all stages of the product life cycle more than ten years ago. Since then, the model has not remained static but is subject to evolution under the maxim of continuous improvement. The goal is to maintain a system that shows the right focus and simplifies procedures. The Quality System and the dedication to high-quality standards, as well as the development of a quality mindset in the whole organization, are very important to the company and publicly backed by the CEO.

The Quality organization is headed by a C-level resort for all Quality-related matters at all stages of the life cycle. Different operational groups and staff functions directly report to it. Operational groups are defined by product type or stage of the life cycle. Our main contact for the project was the Head of Internal Quality Auditing. As described in Figure 1, this function serves as second in the line of defense ensuring the quality of the products. The group performs surveillance audits of development, manufacturing, and distribution sites, as well as related administrative offices. In addition to the auditing process, the team also supports establishments in improving their state of quality compliance and inspection readiness. The staff mostly consists of senior employees with extensive professional experience and successful careers in the company or related industry. Previous roles of team members include leading local operations at manufacturing sites or managing Quality Control laboratories.

4.2 Project setting

The initial starting point was an internal re-organization of the quality audit process. The company intended to use more data to improve audit prioritization in scheduling and preparation of individual audits since the current process only considered static factors. The status quo of the auditing schedule are regular intervals in which all sites are audited regardless of their current operational or compliance status. The standard interval of auditing was every two years, except for sites producing sterile drug products that are evaluated annually. Additional ad-hoc inspections could be initiated when there is a specific need such as a negative regulatory inspection or to prepare for a new product

application. The team could also waive audits in case consistently good results were achieved in previous audits and inspections.

The audit function uses several data sources to inform and support the audit planning and scheduling. This applies to the overall focus areas and priorities in the production network, as well as the individual audit of a given establishment. The primary orientation for the team is the compliance history consisting of regulatory inspection records and their own past audits. First, this provides information on the open follow-ups on corrective and preventive actions taken based upon past deficiencies. Secondly, the state of quality is characterized by the overall picture of past audit and inspection outcomes. Additional information is gathered by a preparatory survey for the audit. It captures specific quality problems, the complexity of production processes, and also any management changes in quality and operations[12]. While several operational and quality indicators or performance metrics from the sites are available to the audit function, they are not used in the planning process.

The project was initiated after the company expressed interest in intensifying its analyses of existing quality data to gain valuable insight beyond the limits of monitoring. The research team in St. Gallen already worked competently in this field and also in the context of FDA inspection outcomes, providing the ideal platform for potential collaboration. The company agreed to provide access to their data and expertise. The research team was provided the task to identify links between operational and compliance data that could potentially help in increasing the transparency around the compliance of the manufacturing sites.

The project followed an iterative approach and spanned about 15 months with several breaks due to meeting schedule constraints. Overall, there were four joint workshops and working phases in between them. The last project workshop was conducted virtually, whereas the earlier ones took place at different sites of the partnering company. The final step in the collaboration involved an expert interview to review the overall project and its contribution to current managerial practice that was held in late 2020. The questionnaire to guide the semi-structured interview is attached in Appendix A.

[12] Exemplary surveys were available during the project, but not included in this thesis due to confidentiality reasons.

4.3 Dataset

The analyses in this chapter fully rely on data provided by the company. More precisely, three pillars made the research in this chapter possible. First, these are GMP inspection outcomes of each site documented for all regulatory agencies that oversee a given site. Secondly, internal audits conducted by the corporate quality audit group were also included for each manufacturing site. The normalization of both data types is described in the following sub-section. A brief description of the available operational performance metrics, the third pillar, is introduced in Section 4.3.2. The available data contains information from 37 production sites, typically larger ones with an international presence in various markets. Internal Quality compliance audit information and inspection results from 2013 to mid-2019 were available. Operational data was available for the time frame since 2016.

4.3.1 Regulatory inspections and internal GMP audits

Inspections of a particular manufacturing site by any regulating authority are rather infrequent, i.e., typically fewer than once in a year. Nevertheless, manufacturing sites with a global delivery footprint experience numerous inspections per year due to the involvement of multiple agencies. This creates a remarkably complex landscape of interactions with the regulators that the manufacturing sites have to handle. Figure 22 shows the number of inspections by different agencies[13] conducted in the selected plants in a time of five years.

[13] Since the specific combination of regulatory agencies could indicate the partnering company and its manufacturing network, they remain anonymous in this figure.

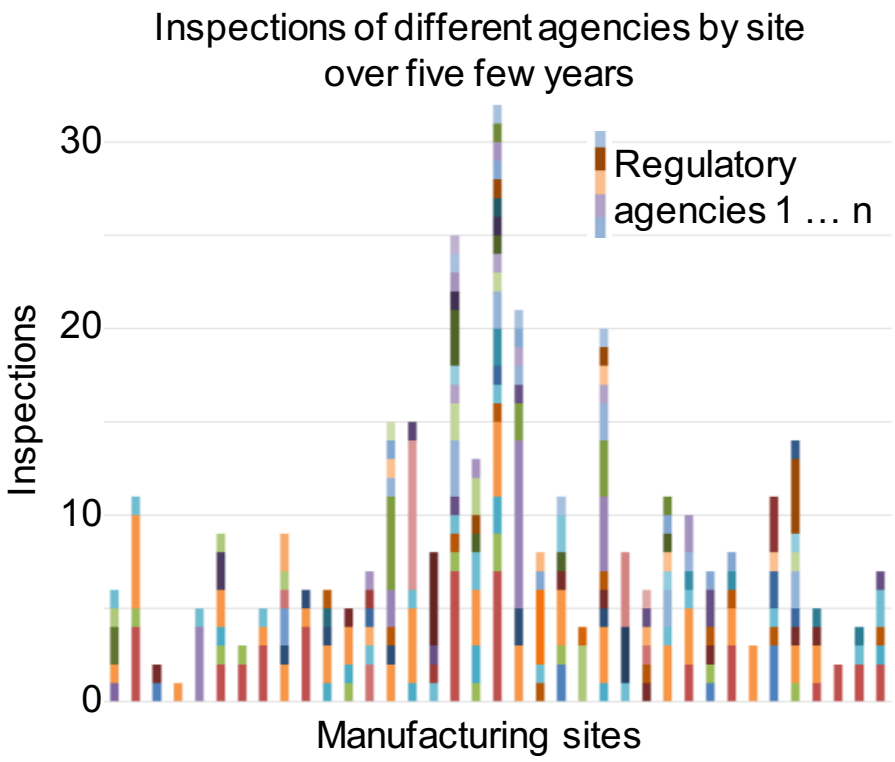


Figure 22: Cumulative number of inspections of different agencies by site

The company deals with various regulators around the world, whereas the original focus was the evaluation by the US FDA. There are more than 50 drug FDA inspection records for the selected sites in the respective time frame. To increase the dataset, it was decided to expand the dataset by also integrating FDA recognized national agencies that are part of the European Medicines Agency's (EMA) network. This resulted in an overall dataset of 149 inspections open to evaluation. Different inspections of the same site were treated as independent data records. Allocation of sites and site types to the individual groups of data records was subsequently controlled for.

FDA and EMA agencies both mutually recognize their GMP procedure statures, but they do not share the same inspection protocols. To ensure comparability, their typical patterns in evaluating inspections were examined. Unlike the FDA, the inspections of European agencies, as well as the internal quality audits distinguish findings into three categories (critical/ major/ minor). Furthermore, they tend to have very different outcomes in terms of the number of findings in their inspections. Figure 23 presents an overview of the inspection results of the regulatory authorities in question.

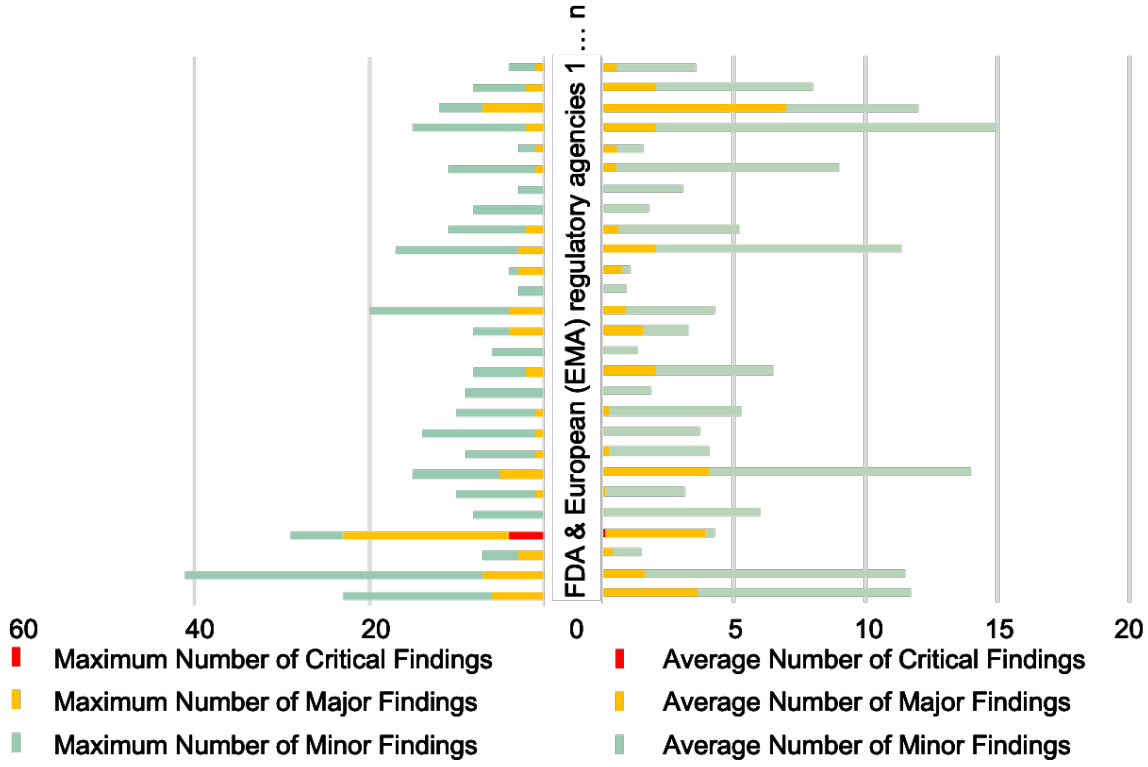


Figure 23: Number of findings per inspection by agency

Since the rating mechanisms across the agencies are different, a normalization process was necessary. Across the different agencies, there is no easily comparable measure such as the final decision following an FDA inspection. However, for every inspection, the datasets include the number of citations in three different criticality groups. These offer a greater amount of information around the individual citations. Therefore, the focus of the analyses was set on the number of findings or citations in the inspections. Because agencies do not average around the same number of citations, these are normalized by comparison to the maximum number of citations in that category per agency.

The normalized number of citations then was aggregated for a preliminary inspection/ audit score in a weighted sum using the following weights:

- Critical findings/citations ≙ 9
- Major findings/citations ≙ 3
- Minor findings/citations ≙ 1

The preliminary inspection scores were then normalized again to account for the characteristics of manufacturing sites and regulatory agencies. Thereby, two distinct differences were incorporated:

- In internal quality audits and regulatory inspections, API sites seem to receive significantly fewer citations than drug product sites. Therefore, the respective normalization streams were separated.
- The European agencies' and the FDA preliminary inspection scores are quite different due to the weighted different criticalities of citations. Hence, separate normalization streams were followed.

Based on the above-mentioned conclusions, the following normalization equations were formalized:

$$Preliminary\ inspection\ score = 9 * \frac{\#\ critical\ findings}{\max\left(critical\ findings_{i,j}\right)} + 3 * \frac{\#\ major\ findings}{\max\left(major\ findings_{i,j}\right)} + \frac{\#\ minor\ findings}{\max\left(minor\ findings_{i,j}\right)}$$

$$i = inspecting\ agency, j = value\ chain\ step\ (API\ or\ drug\ product)$$

The preliminary score was then processed as follows:

$$Inspection\ score = 1 - percentrank(Preliminary\ inspection\ score)_{j,k}$$

$$j = value\ chain\ step\ (API\ or\ drug\ product), k = inspecting\ agency\ (European\ or\ FDA)$$

Since the FDA does not provide classify citations based on their severity, the calculation was simple and is illustrated with a fictional example:

The given FDA inspection of an API site was concluded with three citations. The example assumes that the maximum number of citations in FDA inspections for any API site in the sample is seven.

$$Preliminary\ inspection\ score = \frac{\#\ findings}{\max\left(findings_{FDA,\ \ API}\right)} = \frac{3}{7}$$

The overall population of Preliminary FDA inspection scores for API sites is (0, $\frac{2}{7}$, $\frac{3}{7}$, $\frac{5}{7}$, $\frac{5}{7}$, $\frac{7}{7}$). The percentrank-function[14], computes the following inspection score:

$$Inspection\ score = 1 - percentrank\left(\frac{3}{7}\right)_{FDA,\ \ API} = 1 - 0.4 = 0.6$$

At 0.6, the given inspection ranks in the better half of inspections.

In conclusion, all inspections are equipped with a comparable score between 0 and 1 with a higher score being more favorable, representing fewer findings. The audits are processed similarly excluding any differentiations by agencies since there is only one internal corporate audit function for quality compliance.

In practice, the generated inspection scores show an average of 0.55 with a median of 0.53, and a standard deviation of 0.35. Figure 24 depicts the resulting distribution in a histogram.

[14] percentrank.inc in MS excel turns any values into the percentile it represents in a given sample, it therefore ranks values by how many values in the sample are bigger/smaller on scale from 0 to 1. A return value of 1 thereby represents the highest value in the sample.

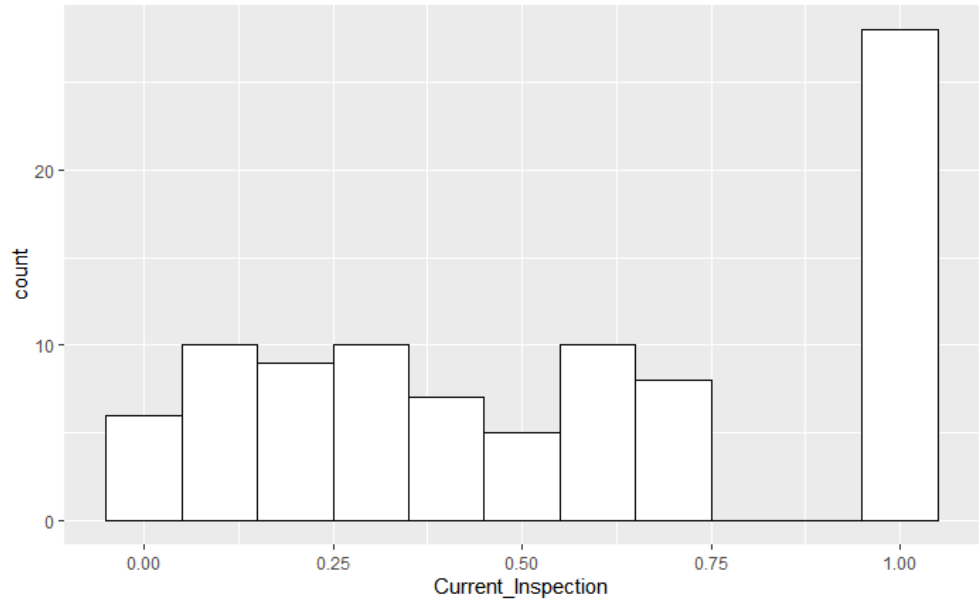


Figure 24: Distribution of inspection scores in the sample

Generally, there is no direct link between the number of findings and the actual severity of the overall inspection outcome. An inspection with more citations is not necessarily worse than an inspection with only one. This is especially true for FDA inspections where minor or critical citations cannot be differentiated. However, the citations of FDA inspections were partly assigned to the criticality levels by quality compliance experts of the company. Still, the normalization bears a certain limitation with regards to its representation of the severity of inspections which was accepted and hereby highlighted.

As a secondary option for operationalizing inspections, a measure similar to the one used in Chapter 3 was chosen. To do so, a binary classification was created based on whether or not any citations were issued in the given inspection. The result is a trivial two-group differentiation, and as such lacks some detail of the operationalization. However, it is ideally suited for various prediction algorithms which will be introduced in Chapter 4.5. Figure 25 visualizes the distribution of inspection outcomes as a binary measure in a pie chart. It shows that about one-third of the inspections in question are concluded without any findings. Therefore, it is slightly more imbalanced than the analyses of manufacturing sites based on the survey reported in Chapter 3.2.

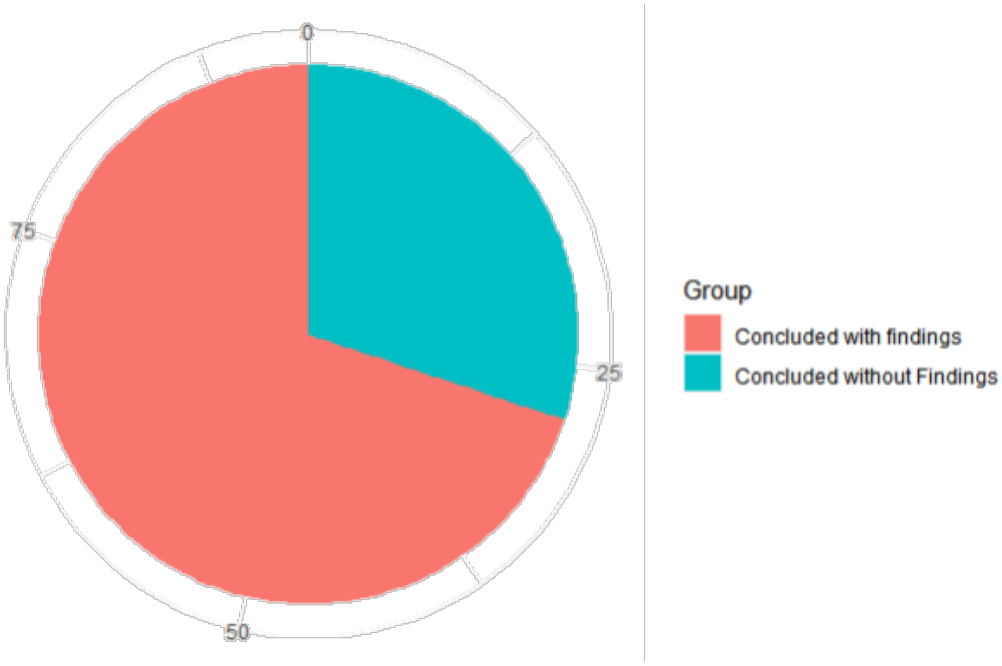


Figure 25: Distribution of inspection outcomes as a binary measure

Throughout the collaboration, the two ways of operationalizing inspection outcomes were discussed in workshops and thereby validated against the subject matter expert's (SME) opinion and experience. Particularly the normalization into an inspection score was strongly valued as an insightful way to make inspection outcomes broadly comparable. The operationalization of inspection outcomes has also been discussed with several representatives of the FDA's Center for Drug Evaluation and Research in an on-site workshop and follow-up meetings. Given the limitations introduced above, they were deemed valid also from the agency's perspective.

4.3.2 Operational performance data

The performance KPIs shown in this chapter were selected based on their availability within the project. Of the overall dataset of 149 inspections, only 93 of these records also contain operational performance data, since it was not captured in the earlier years. Three performance metrics were available across most sites and therefore, a sufficient dataset for the analyses. These include:

- Lot Acceptance Rate
  The company uses a definition quite close to the one the FDA uses and that was introduced in Section 3.2.1.

- Right First-Time Rate
  The "Right First Time" KPI is another production KPI that is important for managing quality control. It refers to all product batches that pass internal quality control without any defects in the first run. It is therefore a good quality indicator to further increase the percentage of defect-free products in production. High values of this indicator represent an efficient and high-quality production process.
- Customer Complaint Rate
  Again, the company uses a definition quite close to the one the FDA uses and that was introduced in Section 3.2.1.

The data is available on a monthly basis per site. In addition to the current value of any given month, the moving average of the previous six months was computed. Since there can be notable volatilities on a month-to-month basis, this aims to improve the robustness of the performance values.

$$\bar{y}_t = \frac{1}{6}\sum_{t-5}^{t} y_t$$

4.4 Descriptive analyses

The first analyses presented in this chapter serves a dual purpose. First, they offer an exploratory view into the given dataset of the partnering company and thereby shall help inform the later predictive analyses that follow in Chapter 4.5. Secondly, with regard to the available predictors of inspection results, there is some overlap with the coefficients used in Chapter 3.2 which allows challenging the previous findings.

4.4.1 Methodology

The data records are assigned to groups based on their current regulatory evaluations. The groups are then compared by their operational KPIs, and preceding quality compliance performances. Differences between the groups were tested for significance using t-tests. The compilation of sites' data records in terms of technologies and drug products (DP) are shown in Figure 28. Higher

inspection scores are the ones with no or few citations. Such inspections will also be referred to as “good” or favorable in the following sections.

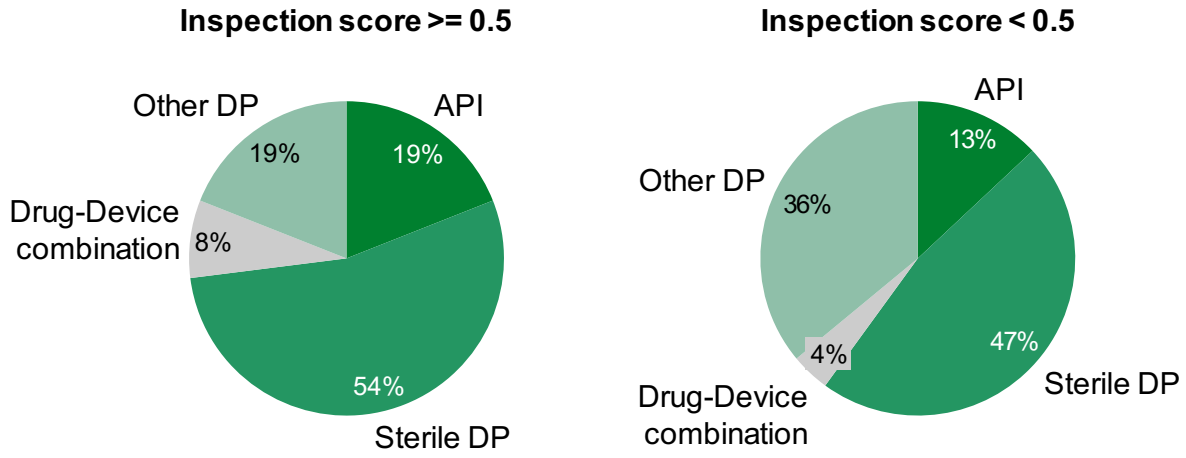


Figure 26: Product types of site data records by inspection score above/below the median

Overall, the distribution of technologies is similar. However, the group with comparatively more findings/ citations in their current inspection shows a higher share of non-sterile drug product manufacturing sites (typically solids & liquids). This does somewhat contradict the general assumption of sterile drug product manufacturing sites being more prone to manufacturing instabilities (Friedli, Köhler, *et al.*, 2018), and risk of compliance failures (FDA - CDER, 2020b). However, this trend applies most to ‘no application’ (i.e., generic) drugs and off-shore location manufacturers. The partnering company on the other hand is specialized in sterile drug manufacturing and has longstanding experience with the associated technologies and quality requirements.

4.4.2 Results

This section shows the results of the t-tests performed in this study. For the Rejected Batches Rate, there is only a small difference between the current values of sites that had inspection scores above or below 0.5, respectively (cf. Table 18). However, regarding the six months moving average, the t-test showed a significant result ($p < 1\%$). This could be due to reduced stability in monthly rates, as they are usually influenced by a relatively small number of overall batches as the denominator. This assumption is supported by standard deviations being much lower for the moving averages compared to the individual

months. Sites with good inspections on average have a Rejected Batches Rate that is almost one percent level lower compared to the other sites.

Table 18: T-Tests for Lot Acceptance Rate and current inspection score

T-Test results: Rejected Batches Rate

Metric	Inspection score	n	Average	Median	Standard Deviation	Significance p-value
Current month	>= 0.5	49	1.12%	0.00%	2.94%	19.05%
	< 0.5	44	1.69%	0.55%	3.24%	
6 months moving average	>= 0.5	53	0.86%	0.31%	1.63%	0.94%
	< 0.5	44	1.78%	0.69%	2.08%	

As introduced above, the Right First-Time Rate, similarly to Rejected Batches Rate, is an indicator for the stability, and robustness of production processes. As shown in Table 19, sites with good scores for an inspection at the time show higher rates of lot acceptance without any deviations. This trend is significant ($p < 5\%$) at the time of the inspection, and slightly below the confidence level for the six months moving average.

Table 19: T-Tests for Right First-Time Rate and current inspection score

T-Test results: Right First-Time Rate

Metric	Inspection score	n	Average	Median	Standard Deviation	Significance p-value
Current month	>= 0.5	38	92.2%	96.4%	12.16%	4.62%
	< 0.5	39	86.2%	93.3%	17.77%	
6 months moving average	>= 0.5	42	92.5%	96.2%	11.69%	5.22%
	< 0.5	39	87.6%	94.4%	14.70%	

Thirdly, Customer Complaint Rates were analyzed again in an overlap with the analyses in Chapter 3.2. Analogously to the metrics presented above, the current value for Product Complaint Rate is the primary variable to be analyzed. In addition to that, a lagging moving average was chosen. In consultation with the partnering company's experts, it was determined that production failures take about three to six months to result in complaints. This was translated into a corresponding lagging average metric shown below. Customers complaining

about the product quality could be a very helpful measure of process stability but certainly is less helpful in prediction due to its indirect nature, and the corresponding time gap.

$$\bar{y}_t = \frac{1}{4}\sum_{t+3}^{t+6} y_t$$

Other than the KPIs shown above, the results achieved with the St. Gallen benchmarking dataset cannot be confirmed here. There is no statistically significant connection between the assignment of groups of inspection based on their score, and the complaints received by the site. The full results of the analyses are tabulated in Table 20.

Table 20: T-Tests for Customer Complaint Rate and current inspection score

T-Test results: Customer Complaint Rate

Metric	Inspection score	n	Average	Median	Standard Deviation	Significance p-value
Current month	>= 0.5	32	0.25	0.17	0.36	18.80%
	< 0.5	37	0.18	0.10	0.33	
3-6 months lagging average	>= 0.5	32	0.23	0.12	0.36	32.88%
	< 0.5	43	0.20	0.13	0.26	

Digging deeper into that analysis, sites that produce drug-device combinations are over proportionally present in the group with good inspections. In discussions with the subject matter expert, it became imminent that such products often result in much higher complaint rates due to their increased difficulty in handling. Therefore, the uneven product mix across groups could explain the inconsistency with the results from Chapter 3.2.

As shown in Table 21, there are no significant differences between the two groups in terms of their previous internal quality compliance audit, which might be attributed to several factors. Audits and inspections can have different focus areas. Also, a site evaluated with many findings in an internal audit can correct these deficits and may obtain a better evaluation in the subsequent inspection,

so that the results may turn out to be very different. After all, there is a slight tendency for good inspections to follow sooner after better audits.

Table 21: T-Tests for previous quality audit and current inspection score

T-Test results: Previous internal quality audit

Metric	Inspection score	n	Average	Median	Standard Deviation	Significance p-value
Audit score	>= 0.5	75	0.51	0.59	0.32	13.02%
	< 0.5	65	0.57	0.62	0.30	
Months since audit	>= 0.5	75	7.25	6	5.64	34.53%
	< 0.5	65	7.63	7	5.43	

Contrasting the analysis of previous quality audits, a much clearer picture shows for preceding inspections. As Table 22 summarizes, sites with good inspections also had significantly ($p < 5\%$) better evaluations in their previous inspections. Sites with current inspection scores below 0.5 also had notable time lag since their previous inspections ($p < 10\%$).

Table 22: T-Test for previous and current inspection score

T-Test results: Previous regulatory inspection outcome

Metric	Inspection score	n	Average	Median	Standard Deviation	Significance p-value
Inspection score	>= 0.5	64	0.59	0.56	0.33	1.51%
	< 0.5	48	0.45	0.45	0.30	
Months since inspection	>= 0.5	64	8.91	6	8.85	8.66%
	< 0.5	48	11.23	10.5	8.72	

These results fit a pattern the SME observed in his professional experience. While quality is always a priority at the manufacturing site, there can be a deteriorating effect to the attention placed on being compliant with Good Manufacturing Practices. This can for example occur when other large projects, such as new product introductions or technical changes, require more managerial attention.

4.4.3 Conclusion and Discussion

This section serves as an addition to the work on the St. Gallen Operational Excellence benchmarking database. The relationships of operational performance metrics and quality compliance risk were analyzed in a single case study based on proprietary longitudinal compliance and operational data from one of the global top five pharma companies. Data records that are momentary snapshots of manufacturing sites were grouped according to current inspection scores, and several significant differences between the two groups were observed.

Sites that receive few citations in the current inspection are likely to also have received few citations in previous inspections. Similarly, before an inspection with more findings, typically more time has passed since the previous inspection. This confirms some of the factors used in the FDA's risk-based site selection methodology (FDA, 2018). Furthermore, these results are well aligned with existing literature on the issue (Anand, Gray, and Siemsen, 2012; Gray and Massimino, 2014; Seiss, 2018). These mentioned sources all base their research on larger datasets. Being able to reproduce results on the given data reflects the robustness of the performed analyses. It provides increased confidence to the newly uncovered, previously unpublished findings.

Specifically, these are that better quality compliance evaluations, meaning inspections with fewer (weighted) citations, can be predicted with better performance in operational metrics. This confirms the findings from Chapter 3.2 except for Customer Complaint Rate, which might be attributed to an imbalanced product mix across the sample groups.

From the author's perspective, there could also exist other relationships between the observed phenomena. In a fictional example, manufacturing sites with weaker operational performance KPIs could be under stronger pressure to improve. Such pressure could shift attention from manufacturing according to GMP requirements towards attempting production stabilities using unsustainable short-term fixes. One could imagine this effect to increase if sites remain without regulatory interaction for a longer period of time. Thereby, lacking operational performance values could be an indirect reason for adverse inspection outcomes while the true reason lies in deviation or shortfall from target values. With the currently available data and means of analyses, these must remain hypotheses to be tested in the future and food for thought for practitioners in the industry.

4.5 Predictive Modelling

Based on the data analyzed before and their corresponding results, this research aims to show the benefits of using operational KPIs for the risk evaluation at production sites. This could serve as a proof of concept for a possible extension of the current risk-based site selection model utilized by the regulator (FDA - CDER, 2018). Alternatively, there are also use cases in industrial practice to enable more proactive Quality Risk Management.

4.5.1 Introduction

Analogously to the two ways of operationalizing inspection outcomes, this chapter has two main sections that focus on the prediction of inspection scores and binary inspection outcomes, respectively. The general methodology of training and testing the models and evaluating the performance is explained in Section 4.5.2. A brief introduction to the prediction algorithms employed, as well as the consolidated results of the numeric prediction of inspection scores, follow in Section 4.5.3. Section 4.5.4 analogously introduces the reader to the mathematical models used for binary prediction and the achieved results[15].

4.5.2 Methodology

As introduced in Chapter 4.3, the dataset contains 149 inspection records. Out of these, 93 datasets include operational KPIs and therefore form the basis of the following analyses. The prediction model performance was evaluated using ten-fold cross-validation. This high number of segmentations was chosen to maximize available training datasets in each iteration. The computational stress of the operation is acceptable in this case given the relatively small dataset. In practice, this means that every model was trained on 90% of the data, and tested on 10% of the data ten times. Thereby, every data record was part of the testing dataset once and thus has a stored prediction value. Afterwards, the performance was evaluated across the stored predicted outcomes of all the iteration steps. In all cases, the model fitting function of the R ‘caret’ package was employed.

[15] The full R code utilized for this dissertation thesis is provided as a R notebook in the appendix.

Incidents of missing data occurred in a relatively low number of data records and some of the algorithms used are incapable of handling missing data. Therefore, the dataset was treated with an imputing method established in machine learning, whereby missing variables were replaced with the average, and dummy variables were coded (c.f. Lantz, 2019, Chapter 9). To ensure comparability between the different algorithms, the dataset with imputed values was used for testing and training of all algorithms, including the ones that do not require it.

The prediction accuracy for numerical prediction was done using the mean average error (MAE), and Root mean square error (RMSE). Both measures are among the most popular performance metrics for model prediction performance measurement. The error values were in the unit of the target dimension, in this case ranging from 0 to 1. The RMSE specifically discriminates against higher variance in prediction error and thereby, shows a larger value compared to MAE when prediction error is distributed unevenly (Chai and Draxler, 2014).

With the error measure being contingent on the scale of the prediction target, it is important to visualize its distribution. To recap from Section 4.3.1: The generated inspection scores average at 0.55 with the median of 0.53 and a standard deviation of 0.3508. Figure 27 shows the distribution in a box plot diagram.

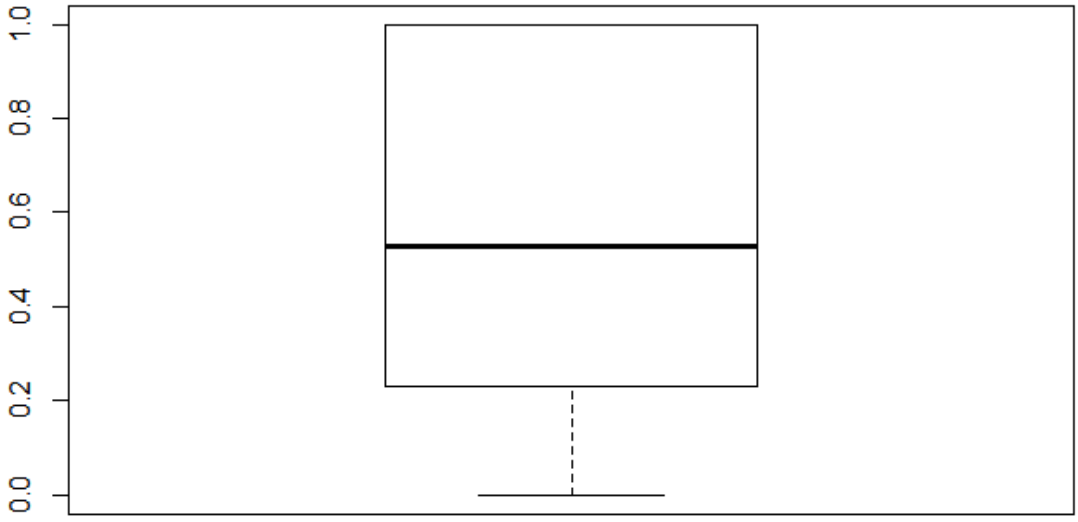


Figure 27: Box plot distribution of prediction target current inspection score

As the standard deviation is the natural deviation in the sample from the average, it is the minimum benchmark any prediction must undercut with its RMSE. Otherwise, a naïve guess would be the better option compared to the prediction model.

For the classification of a binary target, the area-under-the-curve (AUC) of the receiver-operating-characteristic curve (ROC) curve is the standard performance measure. The concept is illustrated in Figure 28. The ROC calculation requires the model to provide a probability for the predicted outcome event to be between 0 and 1. The algorithm then tests several probabilities and cut-offs to evaluate the number of true and false predictions. Generally, the more the curve deviates from the straight diagonal line to the top left, the better the model performs. An ideal prediction model would result in a curve that stretches to the top left corner of the chart.

The area-under-the-curve metric then provides a simplified measure for the prediction performance. Theoretically, the AUC can range from 0 to 1. However, a naïve guess achieves an AUC of 0.5. Beyond that, there are general conventions on interpreting the performance of a classification model based on its AUC as noted by Lantz (2019):

- Outstanding = 0.9 to 1.0
- Excellent/Good = 0.8 to 0.9
- Acceptable/Fair = 0.7 to 0.8
- D: Poor = 0.6 to 0.7
- No Discrimination = 0.5 to 0.6

The machine learning expert also notes that this is a subjective classification, and therefore the valuation of such results must take the specific context, and task into consideration (Lantz, 2019).

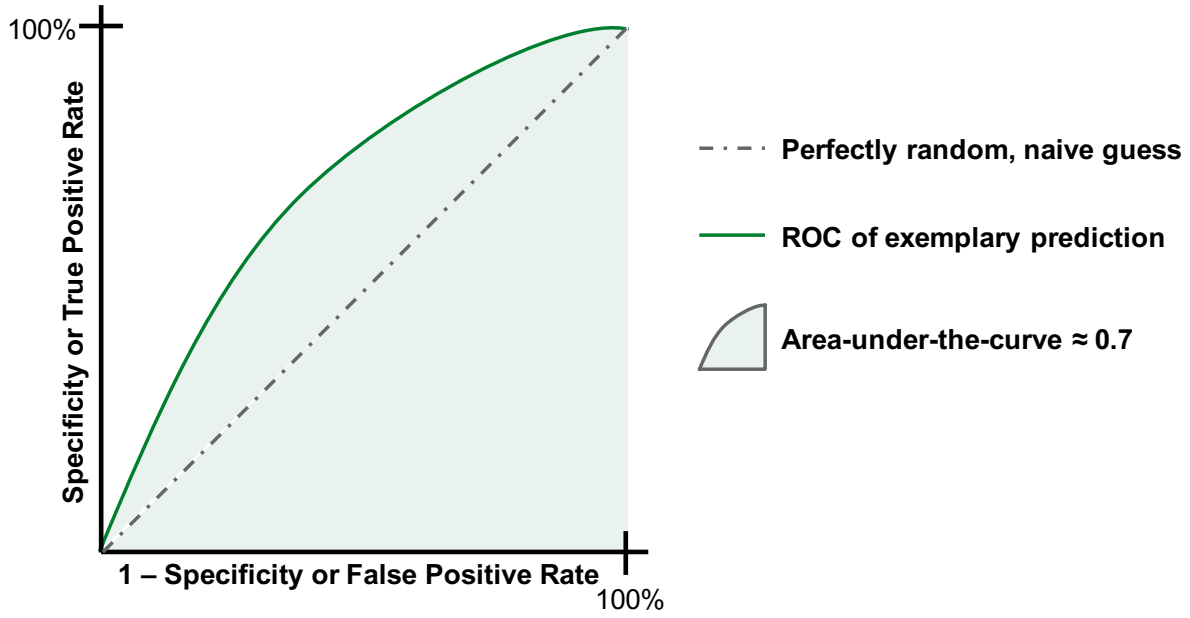


Figure 28: Illustration of ROC and AUC

While there are several other measures based on the ratios between predicted and actual values, such as accuracy, specificity, and sensitivity, this thesis uses AUC for its universal recognition. Besides, it also enables benchmarking the achieved results against the one similar publication by Seiss (2018).

Many machine learning algorithms have one or several parameters that determine their size and complexity. If applicable, several options from small to medium size variants were tested which shall be sufficient given the small size of the dataset. More complex models are often likely to over-fit when trained on relatively small datasets. This means it could near a perfect representation of the known training data, but then be unfit to be generalized to any new data. The tested breadth of specifications has been introduced along with the model types in the following section. The results section includes the best fitting model and specifies the associated model properties with the results.

Across all methodologies, models are trained in several data configurations to evaluate the potential of different types of data for informing the predictions. These include:

1. Inspection legacy
    a. Previous inspection score

   b. Time since last inspection
2. Inspection and audit legacy
   a. Previous inspection score
   b. Time since last inspection
   c. Previous quality audit score
   d. Time since last quality audit
3. Inspection and audit legacy, as well as operational data
   a. Previous inspection score
   b. Time since last inspection
   c. Previous quality audit score
   d. Time since last quality audit
   e. Right First Time Rate
   f. Lot Acceptance Rate
   g. Product Type

The machine learning models also were provided with the missing value dummy variables as a separate, secondary test. Since missing data is rather scarce, the dummy variables only mark very few sites. Therefore, it may induce overfitting of the models in the learning algorithms. As a consequence, the datasets with dummy variables for missing values were tested as a redundant, additional option (data configuration 4).

4.5.3 Prediction of numeric inspection scores

Some numerous algorithms and functions are suitable to produce numeric estimates of a target. In this study, multiple linear regression models, decision trees, and artificial neural networks were selected. The combination offers a span from intuitively explainable algebraic functions to sophisticated black-box machine learning models. All of these algorithms have been well explored in by academicians and research-based literature (e.g. Tabachnick and Fidell, 2014; Lantz, 2019). Therefore, they are only introduced here briefly.

4.5.3.1 Linear regression models

Linear regression models assume a linear relationship between a dependent variable, and one or several independent variables. Coefficients are calculated for every input variable, and the model is fitted using the ordinary least squares

method. The method has several strengths and weaknesses as also compiled by Lantz (2019), which are tabulated in Table 23:

Table 23: Strengths and weaknesses of Linear Regression models for prediction. Table adapted from Lantz (2019).

Strengths	**Weaknesses**
Very common and widely used method in prediction and traditional analyses	Linear relationships are a strong and not always justifiable assumption about the variables
Versatile to almost any modeling task	The model form is specified in advance and does not build itself according to the sample data
Does deliver an indication of the size and the strengths of potential links between independent and dependent variables	Missing data must be excluded or imputed first
	Any categorical data has to be turned into numerical dummy variables first

For this study, while using statistical software R, the standard ‘lm’ model was selected.

4.5.3.2 Decision and regression trees

Decision trees consist of a sequence of binary decisions that can be represented as a tree. Decision trees that are modeled to fit continuous, numerical targets are called regression trees. In each node- a "yes-no" decision is taken (parent node). If the decision is "Yes" in the parent node, the decision in the next node follows on the left side (child node), while if the decision is "No", it is represented on the right side. The decisions are made in such a way that the nodes represent a "pure" class if possible. In each terminal node, an assignment to a value or classification is made, which represents the associated data points the best. The properties are enumerated in Table 24. For the given prediction task, the models were tested with two to ten nodes.

Table 24: Strengths and weaknesses of decision tree models for prediction. Table adapted from Lantz (2019).

Strengths	Weaknesses
Computationally efficient algorithm	Relatively high risk of over- and under-fitting
Capable of handling numerical, nominal, and missing data	Often very little robustness of model on small datasets
Suitable for small datasets	Larger trees can be difficult to interpret
By design only includes relevant predictors in the final model	Not suitable for all problems due to axis-parallel splits

For the prediction of inspection scores, the ‘rpart’ function with a maximum number of nodes of ten (tuneLength) was selected.

4.5.3.3 Neural networks

Artificial neural networks describe the relationship between several inputs and one or multiple return values. The inputs are processed through interconnected nodes that each consists of a weighting function for all the inputs. Other than the number of inputs, neural network models are best characterized by their size. The structure and size of a neural network are determined by the number of nodes in it. More precisely, the number of hidden layers and the number of nodes per layer determine the total number of nodes. Often described as modeled after the human brain’s neurons, artificial neural networks today often only have several hundred nodes. In comparison, the brain of highly developed species has billions of neurons. Today, neural network models are the benchmark for a wide variety of machine learning tasks, but they have their limitations as described in

Table 25 (Lantz, 2019).

Table 25: Strengths and weaknesses of neural networks for prediction. Table adapted from Lantz (2019).

Strengths	**Weaknesses**
Versatile towards classification and numeric prediction tasks	Large networks can require extensive computation in training
Capable of modeling highly complex problems	Likely to over-fit data
Variable towards different types of relationships between the variables	Black box model that is hardly explainable

As explained above, the size of a neural network is described in two dimensions. With its great adaptability, this type of algorithm is at risk of overfitting the data. Since the chance of such behavior increases with the size of the neural network, only small configurations were tested as shown in Table 26. In R, the 'neuralnet' function was used.

Table 26: Tested configurations of neural networks

Nodes per Layer	Configuration 1	Configuration 2	Configuration 3	Configuration 4
Hidden layer 1	10	5	3	3
Hidden layer 2	10	5	3	-
Hidden layer 3	10	5	-	-

4.5.3.1 Numeric prediction results

The prediction performance results of the different linear regression models are displayed in

Table 27. Overall, the performance of all models is relatively bad, with the RMSE only slightly below the standard deviation for the better model configurations.

Table 27: Numeric prediction errors for linear regression

Data configuration	RMSE	MAE	Multiple R-squared	Adjusted R-squared	Best model
1	0.334152	0.161944	0.05397	0.04092	*
2	0.336601	0.080991	0.05878	0.03245	
3	0.353676	0.109945	0.08717	0.02349	
4	0.367778	0.060834	0.1107	0.02602	

The regression tree models show a similar picture in terms of the ideally achieved performance. The full results of the performance evaluation and model size are shown in Table 29. The best model performance is achieved using a relatively small tree with operational performance data as the input.

Table 28: Numeric prediction errors for regression trees

Data configuration	RMSE	MAE	Model size	Best model
1	0.336042	0.284764	4 terminal nodes	
2	0.343326	0.282478	5 terminal nodes	
3	0.328668	0.278185	3 terminal nodes	*
4	0.368335	0.318781	3 terminal nodes	

The impact of the availability of operational performance data is also directly visible in the model itself. The main split of the model is based on the differentiation of manufacturing sites with Right First Time Rates below 96% as an indication of lower inspection scores. Similar to previous analyses, the time elapsed since the last inspection is used as secondary differentiation of the better and worse inspection scores.

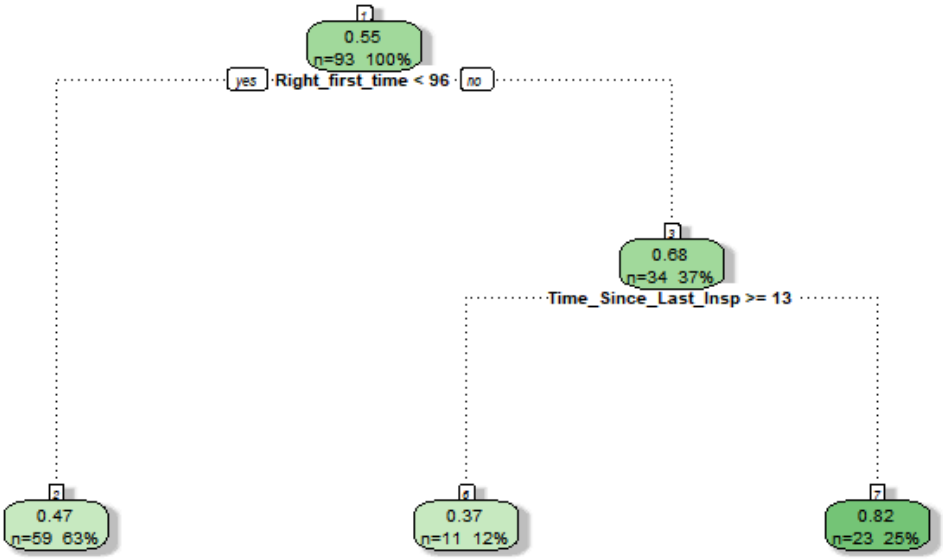


Figure 29: Regression tree model visualization for prediction of inspection scores

The best performance across neural networks was achieved with a model with two layers of five nodes each. It is trained on tested on data configuration three, meaning that in addition to compliance history, it also uses the operational performance data to predict the inspection score.

Table 29: Numeric prediction errors for neural network models

Data configuration	RMSE	MAE	Model size	Best model
1	0.352529	0.298077	1 Layer with 3 nodes	
2	0.357002	0.308127	1 Layer with 3 nodes	
3	0.337481	0.299192	2 Layers with 5 nodes each	*
4	0.354072	0.307131	2 Layers with 3 nodes each	

The resulting model is shown in Figure 30, whereby the constants for every node were hidden to reduce complexity. Even with a relatively small model, the result is a visually messy set of interconnected polynomials and factor weights which are hard to understand for humans and almost impossible to interpret contextually.

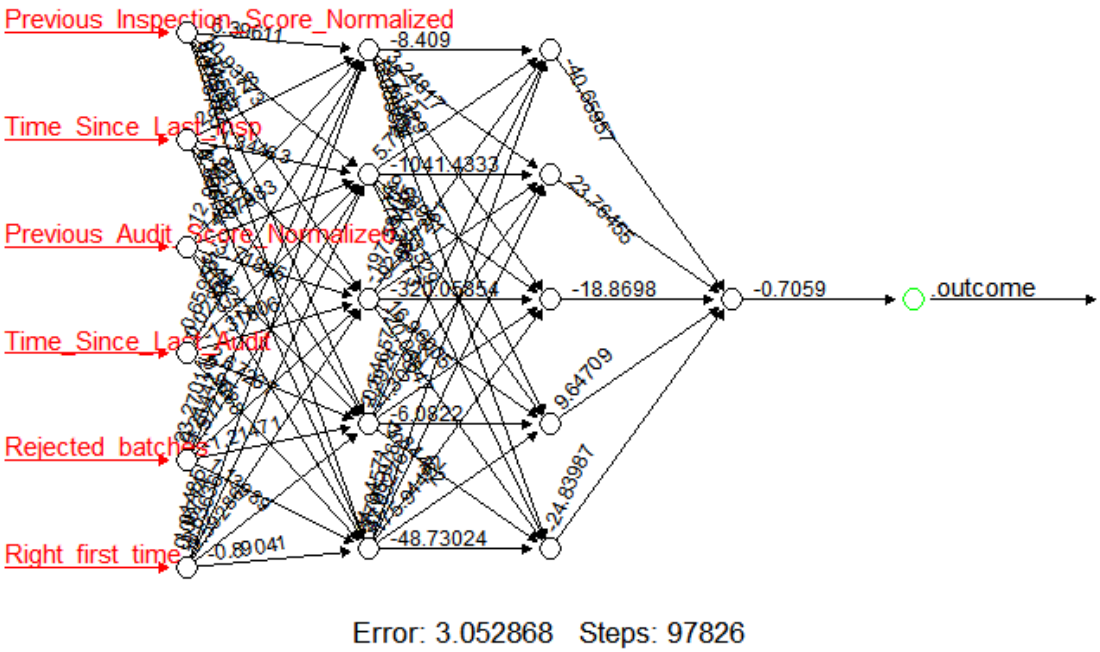


Figure 30: Neural network model visualization for prediction of inspection scores.

Figure 31 provides a comparison of the performance of all three algorithms across the different data configurations. The models trained and tested with operational performance data (configuration 3) were most successful for regression trees and neural networks. Thereby, the regression tree shows the smallest RMSE at approximately 0.329. On the other hand, the availability of dummy variables for missing data consistently delivered results with errors higher than the standard deviation.

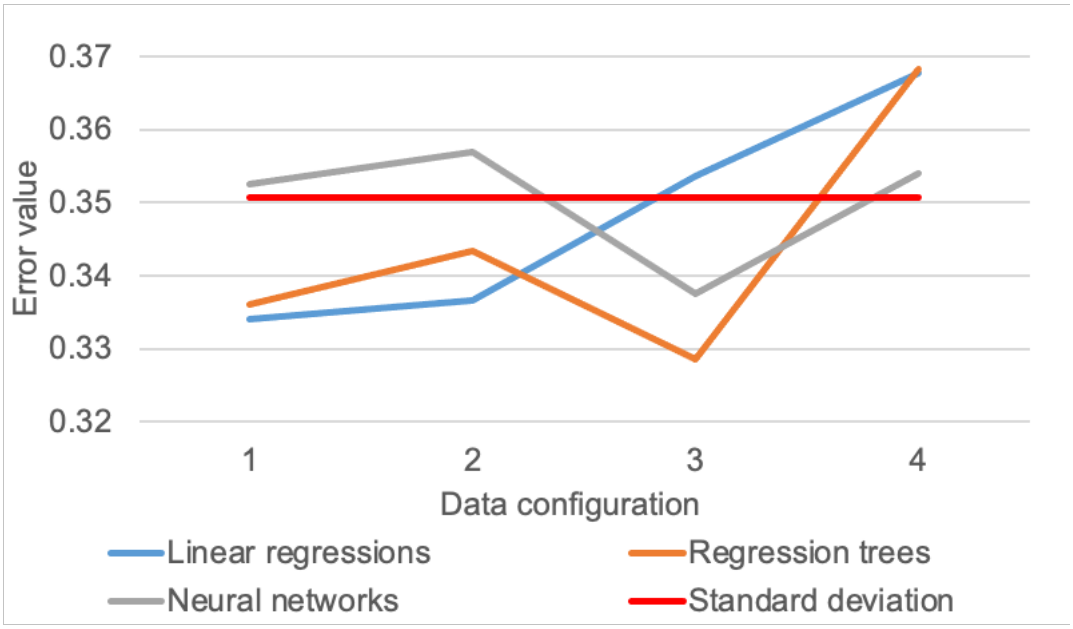


Figure 31: Root-mean-square-error summary chart for all predictions of inspection scores by data configuration

In conclusion, the tested algorithms were not very successful in predicting the inspection scores, regardless of the nature of the input data. This means that none of them showed a strong improvement of error compared to the standard deviation in the sample. The possible reasons for that are manifold. The insufficient size of the training dataset or the skewed nature of the inspection score with an accumulation of inspections that have a score of 1 (without any citations) may be primary contributors.

4.5.4 Prediction of binary inspection outcomes

For the practitioner, a simple and reasonably balanced, binary prediction status of the manufacturing site could serve as an ideal mechanism for scheduling an evaluation of a manufacturing site. In addition to that, it allows using more and other established algorithms for prediction. Using the binary measure for inspection outcomes also has the advantage of not depending on the normalization process of generating inspection scores and its limitations.

4.5.4.1 Logistic regression models

Logistic regression models are a simple, yet effective way of connecting various numerical or categorical input variables to a categorical output variable. Logistic

regression models were employed for most of the hypothesis testing in this thesis. Like linear regression models, they can also be used for prediction.

4.5.4.2 Decision trees

For the introduction of decision trees, please refer to Section 4.5.3.2. Again, 'rpart' models with identical settings were used for the binary prediction.

4.5.4.3 k-nearest neighbor algorithm

The k-nearest neighbor (k-NN) algorithm classifies instances based on their proximity or similarity of the input values. It computes the similarity of every instance to all known training instances across all variables. Then, it selects a specified number (k) of its closest neighbors within the training data records. Finally, the classifications on the k neighbor instances are evaluated and the majority classification among them is assigned to the instance currently to predict.

Table 30: Strengths and weaknesses of k-NN models for prediction. Table adapted from Lantz (2019).

Strengths	**Weaknesses**
Simple and effective	Does not produce a model, therefore it is difficult to reconstruct the link between input variables and assigned class. Also, every classification takes relatively long.
Agnostic to all data distributions	Selection of suitable k is critical
No training phase, therefore very fast	Unable to handle nominal or missing data

4.5.4.4 Random forest models

Random forest models are combinations of several decision trees. From the training data, several non-exclusive, smaller datasets are created. The datasets are smaller in terms of the number of features or variables that are present. The random forest models then are trained on the new datasets. Thereby, less prominent features are more likely to be included as decision nodes than in one singular decision tree. This can be particularly useful when using datasets with

a high number of features. Lantz (2019) lists several further strengths and weaknesses of the technique:

Table 31: Strengths and weaknesses of Random Forest models for prediction. Table adapted from Lantz (2019).

Strengths	**Weaknesses**
All-purpose model agnostic to the data type or missing data	Black box model that is not easily interpretable
Suitable to datasets with a high number of variables	

The models are tuned across two parameters that are briefly explained here. The minimum number of observations in any terminal node (‘min.node.size’) specifies the depth of the model. Thereby it indirectly also limits the number of terminal nodes when the minimum number is compared to the overall number of observations in the dataset. Requiring the model to include more observations in every terminal node stops the model from overfitting the dataset. If done otherwise, the model could be trained to represent outlier and out of pattern values, which would potentially harm the generalization of the model and its performance on new, unseen data.

The second parameter ‘mtry’ specifies the number of variables that are considered for each tree split or node. The impact of the variable on prediction performance is disputed, but a general suggestion is to use the square root of the number of variables included (Duke University, 2020). In accordance with the mentioned study, the models used in this research were tuned across the parameters as {mtry = 1, 2, 3, 4, 5, 6; min.node.size = 5, 8, 10, 15, 20}.

4.5.4.5 Neural networks

Artificial neural networks have already been introduced in Section 4.5.3.3. While they generally are suitable for all prediction tasks, the ‘neuralnet’ function in R does not work for classification problems. Instead, the ‘nnet’ function is used. It comes with a different structure and parameters, which are ‘size’ and ‘decay’. ‘Size’ specifies the number of nodes in the single hidden layer, the algorithm is limited to. Secondly, ‘decay’ is a regularization parameter to avoid overfitting.

The number of nodes was varied between 1 and 10 by steps of 1. The decay parameter was tested in 0.1 increments between 0.1 and 0.5. Between the two parameters, all possible combinations were tested.

4.5.4.6 Binary prediction results

As introduced in Section 4.5.2, AUC was used as the primary performance measure. Where available in the model summary in R, also the sensitivity and specificity of the models are displayed. The sensitivity is the ratio of positives that are predicted correctly (true positives), whereas the specificity measures the share of negative results that are identified, also referred to as true negatives (Fawcett, 2006). Within the context of this application, and the coding of the data, positives are sites that are diagnosed with quality compliance deficiencies in their inspection.

The first classification models tested were logistic regressions. The performance evaluation of the resulting models for all data configurations is displayed in Table 32.

Table 32: Binary prediction performance for logistic regression models targeting inspection outcome

Data configuration	AUC	Specificity	Sensitivity	Best model
1	0.49	0.7230769	0.3571429	
2	0.549	0.7538462	0.4285714	
3	0.72	0.7230769	0.3571429	*
4	0.659	0.6308	0.1429	

As expected, the model with access to operational performance data does perform the best and delivers a notable improvement of performance displayed in the ROC curves shown in Figure 32. Additionally supplying information on missing variables in data configuration 4 results in slightly lower performance.

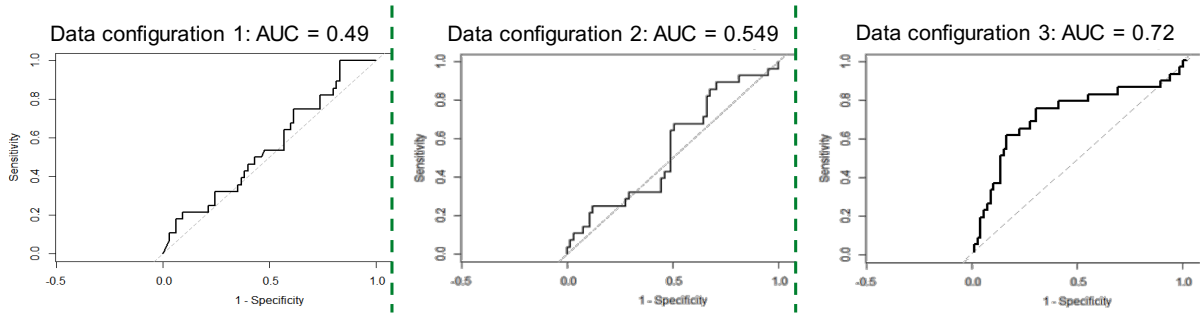


Figure 32: ROC charts for logistic regression models across data configurations 1-3

The decision tree models appear to not work well on the given prediction task with the maximum AUC still below 0.6. Also, the performance does not improve with more variables being added to the available data. The full comparison of the models based on different configurations is shown in Table 33.

Table 33: Binary prediction performance for decision tree models targeting inspection outcome

Data configuration	AUC	Specificity	Sensitivity	Model size	Best model
1	0.541	0.79127	0.288889	5 terminal nodes	
2	0.595	0.713492	0.227778	5 terminal nodes	*
3	0.535	0.790476	0.266667	5 terminal nodes	
4	0.543	0.869048	0.094444	7 terminal nodes	

As expected given the relatively low area-under-the-curve, the receiver-operating-characteristic did not deviate significantly from the bisector line as shown in Figure 33. This in turn results in the shown values for the AUC. The small differences in AUC do not allow much interpretation of the results. Nevertheless, the model based on data configuration 2, being the best performer, also heavily relied on quality audit data for its decision nodes. Models 3 and 4 also featured the previous audit score and/or the time passed since the previous audit in the decision tree. This is pointed out here as there were no significant findings regarding the links between these two metrics and inspection outcomes when testing hypotheses in Chapter 4.4.2.

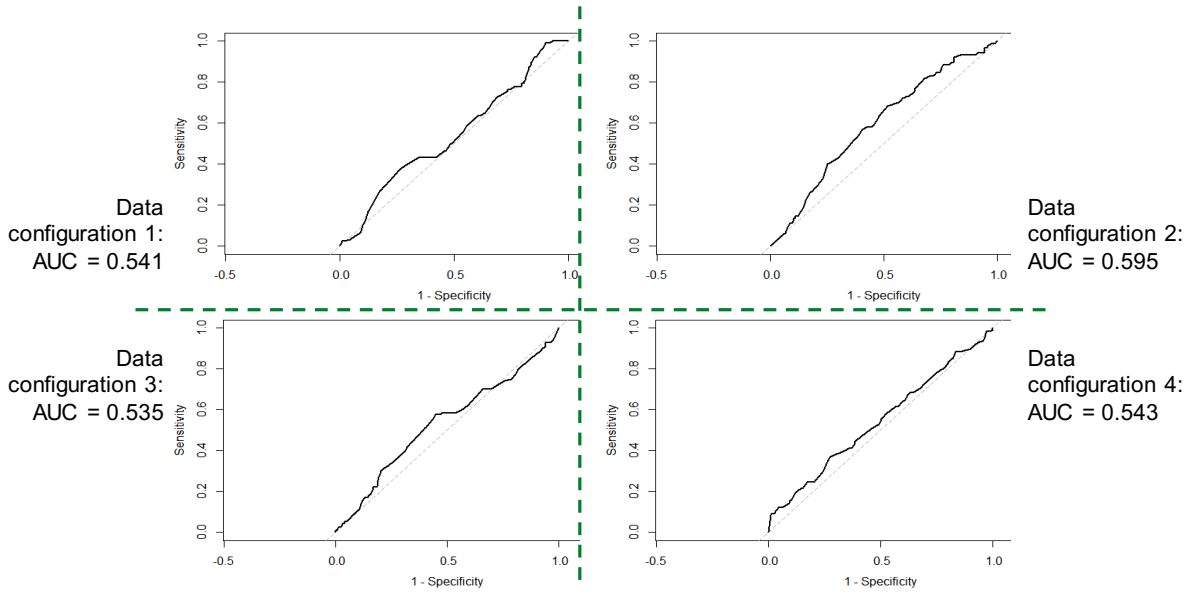


Figure 33: ROC charts for decision tree models across all data configurations

The random forest prediction model similarly achieved no notable performance improvement against the naïve guess benchmark. Table 34 reveals that the model tuning always gravitated to a minimum number of observations per terminal node that is the specified upper limit. This is interpreted as forcing the model to stop overfitting. The downside of this is that the technique thereby loses the opportunity to profit from the existence of many decision nodes and incorporating various predictors. This behavior can probably be attributed to the model being too complex for the given task with a relatively small dataset in the context of machine learning.

Table 34: Binary prediction performance for random forest models targeting inspection outcome

Data configuration	AUC	Specificity	Sensitivity	min.node.size	mtry	Best model
1	0.568	0.859	0.2167	20	4	*
2	0.528	0.986	0.1667	20	3	
3	0.55	0.845	0.4	5	6	
4	0.568	0.938	0.2	20	3	*

While the interpretation of a random forest model is arguably difficult, it is possible to calculate the relative importance of the variables for the prediction results in R. Doing so for model 4, produces the chart shown in Figure 34. It shows that the three most important variables are two operational performance metrics[16] and the previous inspection scores. At the same time, the dummy variables highlighting any missing data are mostly irrelevant. This could be expected in the overall picture and is also logical since they represent isolated cases while the minimum size of terminal nodes required relatively large groups of observations for this particular model.

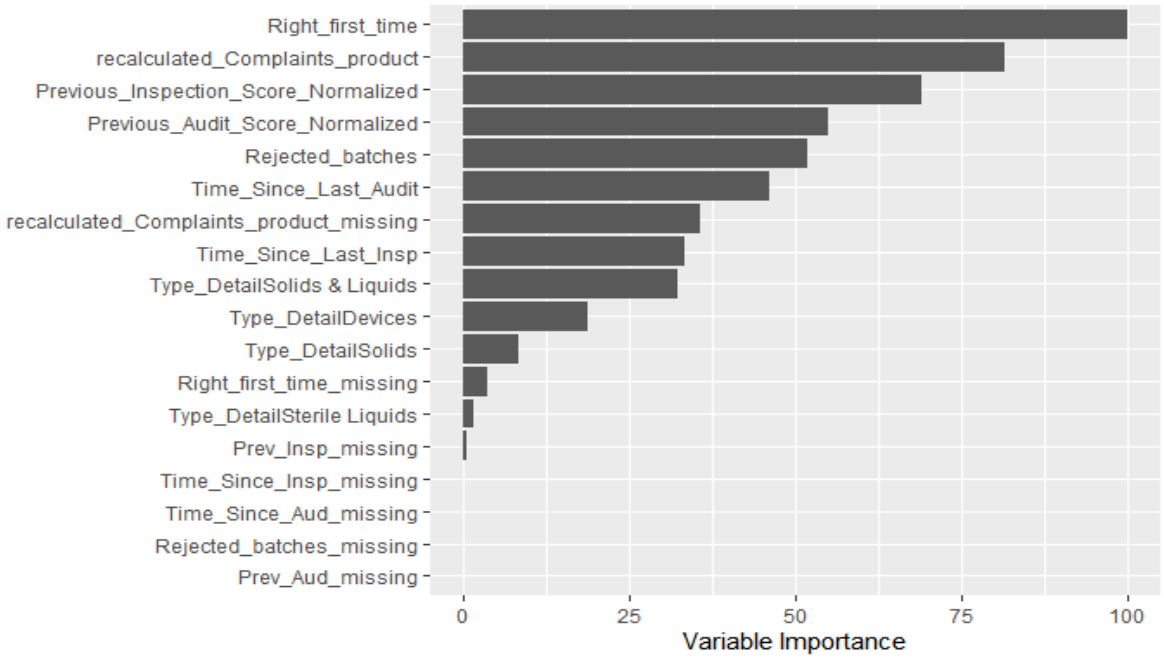


Figure 34: Variable importance in random forest prediction of inspection outcomes

Unlike the previously introduced models, the kNN technique does not compute a probability of data records belonging to a certain group but just provides a classification. Therefore, it is not possible to create a ROC curve or compute an AUC value when using kNN[17]. Instead of AUC, Table 35 provides the accuracy of the models to judge their performance against each other. While the metric is suitable to do so, it is not ideal to judge the performance of the models in

[16] "recalculated_Complaints_product" is the abbreviated variable name for the lagging average of Customer Complaint Rate shown in Table 20.

[17] While there are ways to have R compute AUC values for kNN, it is not possible to do so using it within the 'caret' package applied throughout this research for training and cross-validation of prediction models.

general, because it does not take into account that the target variable is not balanced (cf. Figure 25).

Table 35 Binary prediction performance for k-nearest neighbor models targeting inspection outcome

Data configuration	Accuracy	Specificity	Sensitivity	k	Best model
1	0.5645	0.730476	0.18	2	
2	0.5763	0.710476	0.263333	2	
3	0.7333	0.897143	0.355	7	*
4	0.7194	0.884524	0.348333	8	

The artificial neural network algorithms, as with the numerical inspection scores, did not perform the task of predicting the binary inspection outcomes well. The ‘nnet’ function reached a maximum AUC of 0.593 when inspection, audit, and operational data were included. The full overview of the model is shown in Table 36. The selected models are relatively small as the optimization mechanism most probably attempted to limit overfitting.

Table 36: Binary prediction performance for neural network models targeting inspection outcome

Data configuration	Accuracy	Specificity	Sensitivity	Size	Decay	Best model
1	0.515	0.698889	0	1	0.5	
2	0.554	0.699444	0	1	0.5	
3	0.593	0.745278	0.253024	4	0.2	*
4	0.565	0.741111	0.23674	6	0.4	

In conclusion of this section, only the logistic regression model delivered a prediction result notably above 0.5, which is also shown in Figure 35. As shown in Chapter 4.4, operational KPIs are significantly different for sites inspected with fewer or more citations. Consequently, comparing the different configurations of logistic regression models, these KPIs are suitable to strongly improve the models in predicting inspection outcomes.

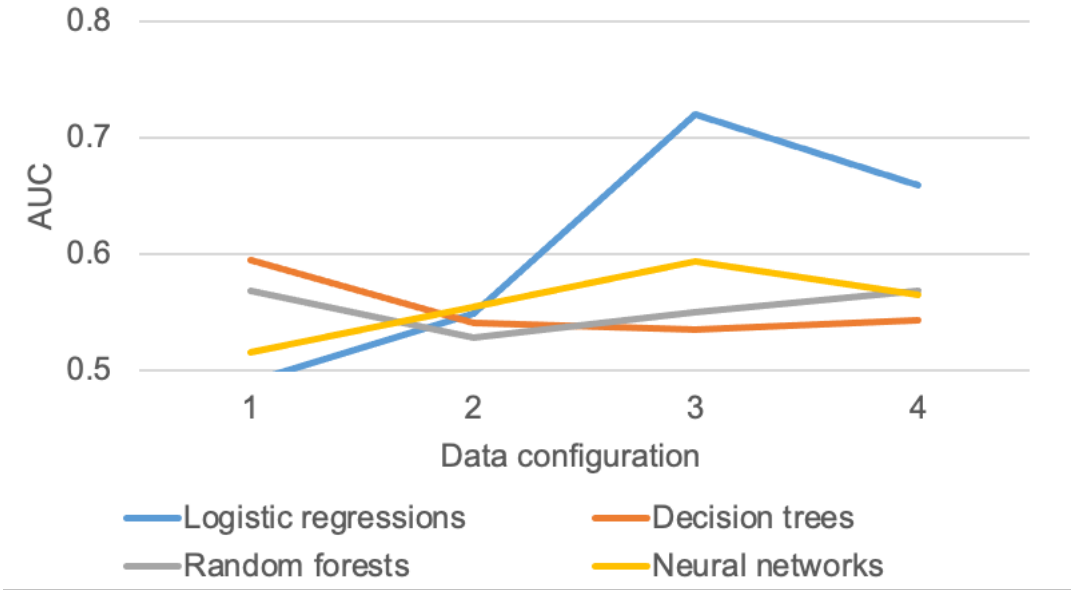


Figure 35: Area-under-the-curve summary chart of all predictions of inspection outcomes by data configuration

Besides that, there are two other observations. Where available, the operational performance data was used as an integral part of the prediction models. Even though the performance of random forest or decision tree models remained rather poor, this indicates the predictive power of these variables. Secondly, there are strong signs of the complex models overfitting the data and therefore performing below expectation

4.5.5 Conclusion and Discussion

Even though the available dataset is relatively small for a machine learning application, the logistic regression model shows significant performance in predicting inspection outcomes at an AUC of 0.72. While generally rated as “acceptable/fair” by Lantz (2019), these results must be taken in the context of the given application. Across the screened literature, there is one comparable study that also aims at predicting regulatory inspection outcomes of pharmaceutical manufacturing sites. Utilizing 3,429 records of FDA inspection outcomes, Seiss (2018) used logistic regression, scorecard, and random forest models to predict inspections that resulted in OAI and OAI or VAI, respectively. The predictors included product type, past inspection outcomes, time since last inspection, and FDA information on hazard signals. In addition to that, the study

made use of proprietary measures for the financial and compliance risk of the establishments.

Even though using a dataset almost 40 times larger, Seiss (2018) too experienced an inferior performance by random forest algorithms compared to other models. He concluded that using random forest or other sophisticated machine learning is not viable on the small dataset and generally requires "*at least ~100K records*" (Seiss, 2018, p.12). The study then tested logistic regression and scorecard models in various configurations and operationalizations. The logistic regression models generally performed better than the scorecard models. While several model configurations neared an AUC value of 0.7, it ultimately was not reached. The best variant thereby showed an AUC of 0.6945. Figure 36 depicts an overlay of the corresponding receiver-operating-characteristic curve and the one obtained in this thesis (Figure 32).

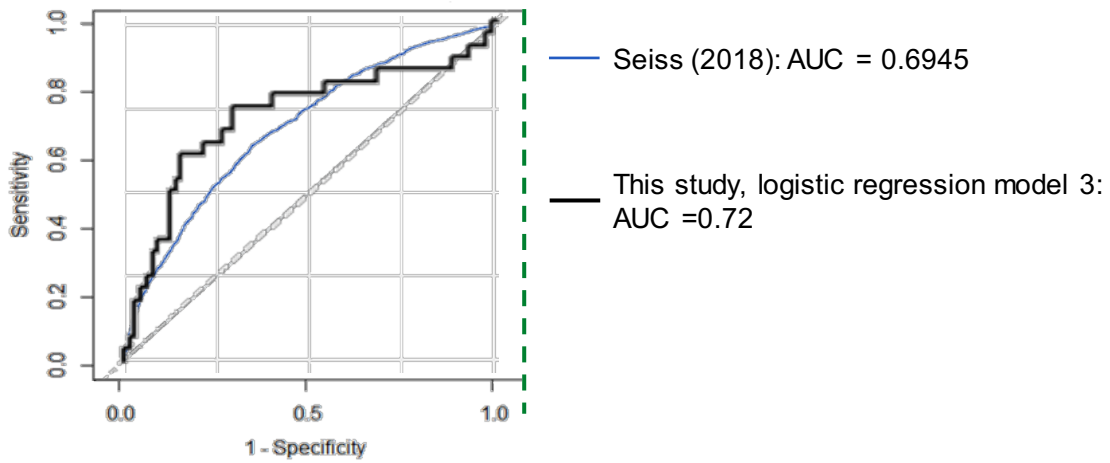


Figure 36: Benchmark of achieved prediction performance AUC. Blue lined chart taken from Seiss (2018).

The benchmark shows that there is a notable gain in prediction performance achieved across the two studies attempting a very similar prediction task. The underlying datasets are very different in size, but most probably overlap for a small part. Although operationalized differently, the types of data generally are quite alike. This of course excludes the financial and compliance data in Seiss' study (2018) compared to the operational performance data used in this research. Given the similar conditions, it can be deduced that operational performance data is much more predictive of negative inspection outcomes than financial data is.

It shall also be noted that Seiss (2018) produced better prediction results of AUC exceeding 0.76 for only targeting OAI inspections. While this is fascinating and highly relevant to the regulator, it is not feasible to produce a comparison with the dataset at hand. Nevertheless, it points to the potential that future work on extended datasets still holds, particularly with the focus on operational performance data.

The research in this chapter certainly is at the absolute lower limit of data size in terms of available data records when using it for prediction. Nevertheless, working with such small datasets is of great relevance to the industry. While regulatory inspections can be numerous from the perspective of an industry practitioner, they are very scarce events when compared to typical machine learning applications. All in all, the dataset used in this chapter represents the available data of one of the biggest pharmaceutical companies. The industry ramped up its ability to centrally manage and store operational performance data only a few years ago. Available datasets within companies will therefore certainly grow over the next years, but only in an evolutionary manner. Availability of datasets that are larger by one or more orders of magnitude on the other hand can only happen with the regulating agencies.

In conclusion, the performed results for predicting inspection outcomes using logistic regression models can be deemed good. In the light of the fourth sub-research question, prediction of inspection results is possible to a certain extend. Operational performance data does play a decisive role in achieving a good quality of predictions and therefore should be included in any future considerations of solving this problem. Along with the previously known indicators, it can help the industry to adopt a data-based Quality Risk Management strategy at the plant level and enable better oversight for the regulating agencies.

The results of the company-specific case study are well in line with the analyses in Chapter 3. The analyses underline the existence of a relationship between production process stability measured in operational performance metrics and the state of quality compliance. In addition to that, they make the aspect of the predictive power of such operational performance metrics for regulatory inspection outcomes more tangible.

4.6 Project impact and future quality audit planning at the industry partner

Within the industry collaboration project, the final step was to predict the likeliness of negative inspection outcomes for all manufacturing sites based on the most current available data. The algorithm used was the best performing logistic regression model. The result was a ranking of all sites by their predicted risk of receiving citations in their next GMP inspection as illustrated in Figure 37. In discussion with the company expert, the ranking resonated with his perception of risk to some extent.

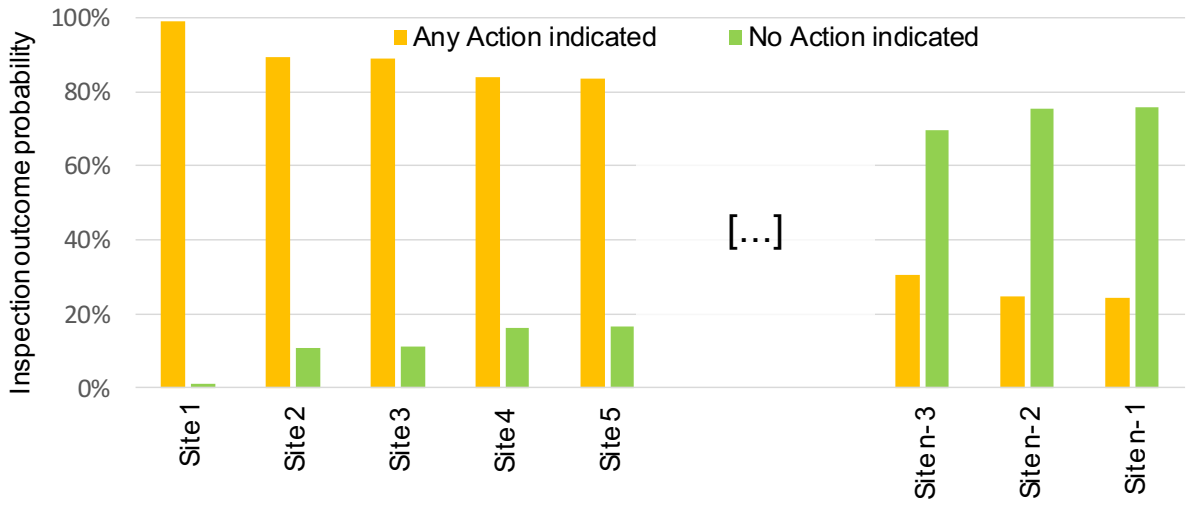


Figure 37: Adverse inspection risk ranking prediction based on a logistic regression model.

The company had already started adapting how inspection readiness and quality audits were planned. One of the changes was to incorporate existing reports on Operational Excellence and particularly quality-related practices implementation of the manufacturing sites in their planning. While the data was not available in the given project, this can only be encouraged drawing from the results of Chapter 3. When questioned about how the state of quality compliance of a manufacturing site can be monitored after the project, the expert's first answer was that they now implemented a survey the sites have to take on the implementation of Quality Systems and data integrity.

After the conclusion of the joint project, the expert pledged to also take into consideration the current operational performance levels of the sites. The existence of a link between operational performance in quality is found highly

intuitive but never realized before. The two dimensions had not been reviewed in combination in the past, but the team will incorporate that now. The details of data incorporation in the process are yet to be defined, but could potentially include variability and deviation from target.

Directly applying prediction algorithms for predicting current risk levels and scheduling quality audits was discussed, but not pursued at the moment. While the technique could benefit from more research and validation of accuracy, the main reason for the decision was not based on technical aspects alone. The context of GMP compliance is a very conservative space, which is not very open to change. Furthermore, corporate audit planning can have strong consequences for the manufacturing sites in question, in terms of resource consumption. There are doubts about the acceptance of decision making that is based on an algorithm, even if it is an explainable model and not black-box machine learning. In the project, the model in Figure 38 was developed to illustrate the journey towards a mature implementation of data analytics in the decision-making process around Quality Risk Management on a site level.

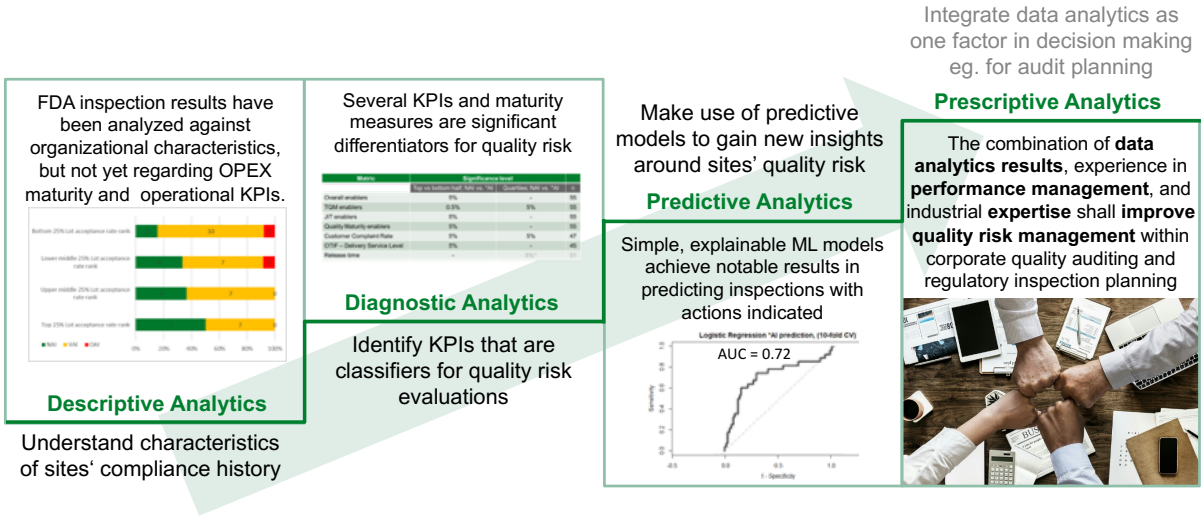


Figure 38: Maturity model for the integration of Data Analytics in QRM

About a year after the conclusion of the project, the model serves as a roadmap for the company. The team there is working on the implementation of several aspects of the project contents. Descriptive analytics will be strengthened by the implementation of a scoring system for past inspections which relates quite closely to Section 4.3.1. It shall also include details of the agencies' inspectors involved, which links to the study compiled by Ball, Siemsen, and Shah (2017).

Diagnostic techniques will be employed to also better understand the auditing process and identify which parts of the work are most beneficial. Incorporating all the different forms of data for a more predictive approach to Quality Risk Management is the next goal to achieve. To do so, our partner company aims to transform the function towards a more quantitative work approach and plans to recruit data analytics resources for the first time to enhance the central audit function.

4.7 Validation of findings with industry experts

Throughout the research and industry project work, contacts with several subject matter experts in various companies were established. These contacts were utilized to discuss the results of quantitative analyses regarding their applicability in practice. The validation aims to engage in discussions with pharmaceutical companies (Section 4.7.1) and the regulating agencies (Section 4.7.2).

4.7.1 Discussion with industry representatives

Representatives of eleven companies were interviewed in virtual meetings on Quality Management systems, management reviews, and operational performance measurement. Their professional job profiles include managing roles within corporate quality assurance and global quality systems management. The discussions took place between July and September 2020. Analyses approaches and preliminary results of this thesis were shared and discussed, which serves as a validation for the developed approaches with the customer in industrial practice. None of the companies uses predictive modeling or any machine learning applications to assess or manage quality compliance risks. However, all of them showed general interest in the principles behind such a solution. Two companies have specific projects in place to explore the option of creating their own predictive system for quality compliance. It should be noted that both companies already were briefed on the research and earlier results prior to the validation interviews. It is unclear if and to what extent the initiated projects were related to the input from this research.

In many cases, there is also an organizational barrier against such ways of working with data. Operational data is traditionally generated and worked with at the respective manufacturing sites. Central corporate organizations typically

have no function for working with or analyzing operational performance or quality data. They therefore solely review the data and take decisions. More recently, this distribution of tasks is changing with companies building or planning to build central data analytics functions. While their main tasks at the moment are not specifically in predictive quality compliance, but for example focus on text-mining product complaints, creating use-cases like the one proposed in this thesis could be the next step.

Out of all the companies, about half of them do integrate operational performance metrics into the quality dashboards that also inform planning from a quality compliance perspective. This leaves about the other half of the companies that still do not see the two aspects as being closely interrelated. The hypotheses of low Lot Acceptance Rates or higher IOOS rates potentially indicating adverse conditions affecting quality compliance did resonate with the interviewees. This was particularly the case with the IOOS rate as it can very well be interpreted as a symptom of a behavioral instead of a technical root-cause. Accordingly, a focus on measuring quality culture and changes in attitudes is one company's main approach to become more proactive in decision making around Quality Management.

The adaption of predictive analytics for corporate quality audit planning and/or inspection readiness assessments is in its early stages for some industry leaders. For most companies, it is an interesting perspective on how data could be utilized but refrained from for the moment. The reasons for that could be plentiful, including a general hesitation against moving too early in such a heavily regulated environment. Instead, many companies wait for best practices to emerge and then adapt. The general importance of operational performance data for quality compliance is widely noticed but has not yet translated into widespread adaption.

4.7.2 The perspective of the regulators

Other than application within companies, the research is designed to inform decision making within the regulating agencies as well. Throughout the time of this dissertation project, several workshops were held with policymakers of the FDA's Center for Drug Evaluation and Research (CDER) in Washington DC, St. Gallen, as well as virtually.

The FDA was the starting point of a now long ongoing dispute of a possible link between operational performance data and quality compliance in pharmaceutical manufacturing (cf. Chapter 1.1). Since then, the agency had to concede that a mandatory reporting of such data by the industry has not yet been implemented. Nevertheless, there are several programs to foster exchange between the FDA and the industry directly, as well as a large-scale data gathering program by Dun & Bradstreet and the University of St. Gallen (Dun & Bradstreet, 2020). Beyond the practitioners' perspective, the FDA has also funded research on the issue to be conducted in various universities across the US with research grants of several millions of US dollars (Pharmaceutical Technology, 2017; Ohio State University, 2019).

Since the introduction of the FDA's draft guidance, the FDA has also adapted their understanding of operational performance metrics and the importance of the selection. In a conference presentation on September 17th, 2020, FDA CDER's Division Director of the Office of Quality Surveillance, Alex Viehmann, included OTIF as part of the new, voluntary quality metrics program. OTIF was highlighted as a predictive metric in Chapter 3.2 of this thesis and mentioned earlier to the FDA during the collaboration. Similarly, the FDA now includes the current states of quality culture and quality systems programs into their considerations. (Viehmann, 2020)

5. Contributions to Theory and Practice

A successful management dissertation project should not only contribute to its specific academic discourse but also provide insights for future practice. The contributions of this dissertation underline the significance of the research project and the practical demands it meets. Therefore, this final chapter leads with a crisp summary of managerial implications that are derived from the results of the performed research. The subsequent paragraphs first highlight the theoretical contribution before explaining the potential practical impact. This thesis closes by reviewing the limitations of the results and an outlook on future research opportunities.

5.1 Implications for practitioners in Quality Management and regulators

The main research question of this thesis is targeted at improving decision making for practitioners with the knowledge of these relationships:

MRQ: How can data analytics be used for predictive quality risk management on a manufacturing plant level?

In short, the previous chapters have shown that data analysis can be used to identify risk factors and directly inform predictive algorithms for quality compliance risk assessments of the production sites. With regards to the research question and specific management advice, the main implications for practitioners are as follows:

5.1.1 Establishing a quality compliance score for tracking inspection outcomes

A quality compliance score that boils down heterogeneous inspection results into a single figure is a valuable tool for quality management across a group of sites. A manufacturing site’s history of quality compliance inspection results is important and must be tracked. It can to some degree predict future outcomes and influence the likeliness of future inspections by the regulator.

Manufacturing sites producing drugs for multiple international markets are subject to inspections by numerous regulators for the respective markets. Such inspections by different agencies or of different types of production sites are difficult to compare since they use different scales and documenting deficiencies differently. Therefore, they always require the judgement of a subject matter expert to be put into perspective. An overview across several inspections of different regulators is difficult to maintain, even for an experienced expert. In many cases, this results in the management mostly focusing on inspections by the domestic agencies and/or the ones of a few key markets.

Normalization of the number of findings by agency and product type allows simplification of the inspection outcomes to an intuitive dimension from zero citations to more and most serious citations observed. This creates comparability between different agencies and the opportunity to aggregate different inspection outcomes into an average. In conclusion, such a compliance score allows comparability among and aggregation across different agencies, as well as simplified reporting of the overall state of quality compliance. However, it in no way replaces the detailed discussion of any individual inspection by experts.

5.1.2 A good state of quality compliance can deteriorate over time

Building on earlier research, it becomes clear that adherence to GMP regulations is not a one-time effort. Several examples show that sites with good inspection outcomes tend to worsen if not inspected again over a longer period of time. While this insight is already incorporated in some of the regulators' processes and intuitive to experts in the industry, it should motivate managers to continuously focus on working on the matter with their respective teams.

5.1.3 The importance of Quality Management practices for GMP compliance

The implementation of Operational Excellence and, more specifically, Quality Management practices is beneficial for the quality compliance risk level of manufacturing sites. Implementing Operational Excellence and deploying Continuous Improvement in manufacturing sites has been a strong priority in the industry since the beginning of this millennial. The idea of including the quality departments in this endeavor is a more recent development. Generally, these measures aim to improve quality results, service levels, and save cost.

However, this research shows that an advanced level of Quality Management practices at the sites and labs is linked to better evaluations in terms of GMP inspections. Thus, it is recommended to not only measure the state of implementation but also consider it in the context of GMP compliance.

5.1.4 Stable, in-control manufacturing processes benefit quality compliance

This study proves the existence of a critical relationship between operational performance indicators and regulatory inspection outcomes. The discussion of KPIs in pharmaceutical production or quality metrics has been ongoing ever since the FDA released their guidelines on the issue. Performance metrics are measured very actively in the industry and used to track the impact of improvement initiatives, target attainment, and benchmark sites against each other. This thesis outlines potential metrics suitable to predict the potential risk of GMP violations. This aspect holds true for both, manufacturing sites as a whole, as well as quality control labs. Therefore, these should be included in corporate quality auditing and regulatory inspection planning.

5.1.5 Creating a central hub for quality data analytics is the key

Enabling the identification of quality risk related patterns in operational data is only possible if it is captured, compared, and analyzed. While regulators like the FDA have invested a lot of resources into becoming more data-driven in their quality surveillance, the industry is yet to catch up. Although it is a very scientific and empirical business, working with quality data on a management level has not evolved beyond reporting selected metrics. Most companies have no function that works with quality data across individual manufacturing sites. Nevertheless, some early movers now show initial success building one. Broadening and utilizing the results shown in this thesis requires a data analytics function that does not have to be extensive, but must be well connected. To be successful, it should combine access to operations data, as well as functional and quality compliance experts for drawing the right conclusions.

5.2 Theoretical Contribution

From an academic perspective, this research mainly contributes to the literature stream of several publications in high ranking journals that discuss the outcomes of FDA inspections in various contexts. By advancing this existing quality

compliance research with a perspective into the manufacturing sites and their operations, this research opens a new level of granularity. The first sub-research question is an initial consolidation of the existing research, which has not yet been compiled. Answering the second and third SRQ then enables the research field to move from an econometric to a more operations management perspective. It explores the possibility of quantitatively measuring the quality risk state of a specific manufacturing site. Existing literature on regulatory inspection outcomes links them to process and system maturity constructs as evaluated in an ISO9001 certification. In contrast to that, this research adds a new level of detail to that discussion by identifying bundles of Operational Excellence practices with a strongly significant impact on quality compliance while also distinguishing manufacturing and QC laboratory operations. To the best of the author's knowledge, this research is the first to link operational performance metrics to quality compliance. It allows future researchers to utilize this relationship in studies of manufacturing sites as interconnected socio-technical systems. It can also be utilized in a research perspective on constraints in such organizations.

Generally, the emerging literature on predictive analytics and their potential in improving industrial operations mostly focuses on direct connections. Of undisputable importance, applications such as predictive maintenance make use of physical causal relationships between temperature, noise levels, vibrations, or power intake of a motor. Relationships in an overall manufacturing site as a socio-technical system are of much higher complexity and often not trivial. Prototyping the usability of predictive analytics in such an operations management context is a valuable addition to the tool-set being reported in current publications. This study could allow the transition from smart silo solutions to a more holistic, evidence-based, and data-informed approach to production management.

5.3 Practical Contribution

This study contributes to two major discussions among practitioners. First, a comprehensive understanding of the relationship between operational KPIs and inspection outcomes adds tremendous value to the ongoing argument on the FDA's Quality Metrics proposal. It can help resolve the politically driven deadlock and dispute with evidence for the first time. These results will enable

the FDA or other agencies to re-evaluate the proposed metrics and potentially select the best possible set of indicators for risk classification. The gained knowledge could on the other hand also help to overcome the industry's resistance around the missing proof of the concept's validity. If introduced successfully, such a new policy will mean important progress and ultimately result in the biggest change of pharmaceutical quality compliance in many years.

Secondly, the application of data analytics for implementing appropriate management decisions increases in popularity across industry and regulators. However, in many cases the specific use-cases and quality of the results are unclear. Proving the concept of operational data informing management decisions viable could increase adoption of these practices. The specific use-case can help corporate quality auditing and regularity inspection planning to better assign resources to the right targets. Establishing the use of such data analytics could enhance current QRM practices while being in line with the ICH Q9 standard. This can bear efficiency improvements, but more desirably and particularly a reduction of quality failures and shortages of critical medicines. Five specific pieces of advice for Quality Management practitioners are compiled in Chapter 5.1.

5.4 Limitations and outlook

Quantitative research efforts are always accompanied by certain risks, which include, but are not limited to the amount and quality of the data gathered. To compile large datasets for plausible and significant answers to the research questions raised in this thesis, the risk was spread across several data sources and data gathering endeavors. For successfully training predictive models, the size of individual datasets and their longitudinal structure are important. The datasets compiled for this thesis allow a unique combination of pharmaceutical production site internals and regulatory inspection outcomes. However, the datasets remained quite small considering the scales applied in the machine learning domain. To the author's best knowledge, this risk was mitigated through the application of suitable model types and training methodologies, such as regression models and 10-fold cross-validation. Nevertheless, the predictive modeling techniques shown in this thesis will benefit from further research with larger datasets and yield a more robust performance.

The shown research results are specifically intended to be used by practitioners in the industry and regulatory authorities. In practical application, far-reaching decisions could be implemented based on the results of such analyses and the recommendations of the algorithms. However, it is advisable to replicate the results on larger datasets before implementation for increased robustness.

The quality of the data can be directly influenced in cases where such data is newly surveyed. This applies to data sources as introduced in Chapter 3. In those cases, special attention was paid to the clear definition of metrics and explanations within the survey. Secondly, data feedback was validated against the overall database and clarified or confirmed in case any outliers existed. The data on Operational Excellence implementation in production sites and quality control labs additionally has a high degree of subjectivity. These so-called enablers are based on a self-assessment on a Likert scale in the questionnaire. During data gathering, participants were equipped with examples or definitions of the different levels of implementation and encouraged to fill out the survey as a group of qualified colleagues. Nevertheless, a certain level of blurriness due to personal perception cannot be ruled out completely.

Despite the best efforts of the author and the research team at St. Gallen, the datasets underlying the research show biases. The analyses in Chapter 3 are skewed to a majority of the pharmaceutical establishments in Europe and the Western world, while the production sites evaluated in Chapter 4 all belong to the same parent company. By comparing the achieved results from both sources with pre-existing literature, several findings were confirmed thanks to similar outcomes. Nevertheless, if compiled successfully, a larger and more geographically diverse dataset could prove beneficial for future research and yield more robust insights.

This thesis highlights conditions in Operational Excellence implementation and operational performance that predict certain outcomes in regulatory inspections. It does not claim to determine the causalities of the relationships between these factors. Further exploration of these aspects, such as testing the posed hypotheses from Section 4.4.3 could generate highly valuable, additional insights. Such research will be necessary for guiding practitioners on how to steer their business and potentially hold interventions in case risk to quality compliance is detected.

6. Declaration of funding and conflict of interest

The project presented in this thesis was partially funded by the Food and Drug Administration [Grant no. 1U01FD005675-01]. The views expressed in this publication do not necessarily reflect the official policies of the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organization imply endorsement by the Government of the United States of America.

7. References

van Aken, J., Chandrasekaran, A. and Halman, J. (2016) 'Conducting and publishing design science research: Inaugural essay of the design science department of the Journal of Operations Management', *Journal of Operations Management*. Elsevier B.V., 47–48, pp. 1–8. doi: 10.1016/j.jom.2016.06.004.

Altamuro, J. L. M., Gray, J. V. and Zhang, H. (2017) 'Organizational Non-Compliance: A Study of FDA-Regulated Industries', *SSRN Electronic Journal*, pp. 1–64. doi: 10.2139/ssrn.2658225.

Anand, G., Gray, J. V. and Siemsen, E. (2012) 'Decay, shock, and renewal: Operational routines and process entropy in the pharmaceutical industry', *Organization Science*, 23(6), pp. 1700–1716. doi: 10.1287/orsc.1110.0709.

van Arnum, P. (2017) *Industry Comments on FDA's Quality Metrics Guidance*, *Value Chain Insights*. Available at: https://www.dcatvci.org/4025-industry-comments-on-fda-s-quality-metrics-guidance (Accessed: 13 March 2019).

Ball, G., Siemsen, E. and Shah, R. (2017) 'Do Plant Inspections Predict Future Quality? The Role of Investigator Experience', *Manufacturing & Service Operations Management*, 19(4), pp. 534–550. doi: 10.1016/j.brainresbull.2018.03.008.

Banker, G. S. and Rhodes, C. T. (2002) *Modern Pharmaceutics*. 4th Editio. Edited by J. Swarbrick. New York, Basel: Marcel Dekker,Inc.

Baryannis, G. *et al.* (2019) 'Supply chain risk management and artificial intelligence: state of the art and future research directions', *International Journal of Production Research*. Taylor & Francis, 57(7), pp. 2179–2202. doi: 10.1080/00207543.2018.1530476.

Bates, H. *et al.* (2006) 'Motor vehicle recalls: Trends, patterns and emerging issues', *Omega*, 35(2), pp. 202–210. doi: 10.1016/j.omega.2005.05.006.

Belhadi, A. *et al.* (2019) 'Understanding Big Data Analytics for Manufacturing Processes: Insights from Literature Review and Multiple Case Studies', *Computers and Industrial Engineering*. Elsevier Ltd, 137, p. 106099. doi: 10.1016/j.cie.2019.106099.

Von Bertalanffy, L. (1968) *General System Theory Foundations, Development, Applications*. New York: Braziller, George.

Birkie, S. E. *et al.* (2018) 'Implementation challenges affecting the environmental improvement performance in pharmaceutical production: Results of a green kaizen pilot', in *IFIP Advances in Information and Communication Technology*. Springer New York LLC, pp. 58–66. doi: 10.1007/978-3-319-99704-9_8.

Borodin, V. *et al.* (2014) 'A quality risk management problem: Case of annual crop harvest scheduling', *International Journal of Production Research*. Taylor and Francis Ltd., 52(9), pp. 2682–2695. doi: 10.1080/00207543.2013.877172.

Bortolotti, T. *et al.* (2015) 'Leveraging fitness and lean bundles to build the cumulative performance sand cone model', *International Journal of Production Economics*. Elsevier B.V., 162, pp. 227–241. doi: 10.1016/j.ijpe.2014.09.014.

Bunge, M. (1980) *Epistemologia*. São Paulo: TA Queiroz Editora Ltda.

Cao, L. (2018) *Data Science Thinking*. Cham, Switzerland: Springer International Publishing (Data Analytics). doi: 10.1007/978-3-319-95092-1.

Chai, T. and Draxler, R. R. (2014) 'Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature', *Geoscientific Model Development*, 7(3), pp. 1247–1250. doi: 10.5194/gmd-7-1247-2014.

Charoo, N. A. and Ali, A. A. (2013) 'Quality risk management in pharmaceutical development', *Drug Development and Industrial Pharmacy*, 39(7), pp. 947–960. doi: 10.3109/03639045.2012.699065.

Claycamp, H. G. (2015) 'Expert Judgements in Quality Risk Management Where Quality Risk Management Can "Go Wrong"', *Journal of Validation Technology*, 21(4), pp. 51–55. Available at: http://www.ivtnetwork.com/article/expert-judgements-quality-risk-management-where-quality-risk-management-can-go-wrong.

Coffey, S. (2008) *Achieving Business and Operational Excellence in the Pharmaceutical Industry*, *Master's Thesis*. Massachusetts Institute of Technology (MIT).

Colledani, M. *et al.* (2014) 'Design and management of manufacturing systems for production quality', *CIRP Annals*, 63(2), pp. 773–796. doi: 10.1016/j.cirp.2014.05.002.

Cooper, H. M. (1988) 'Organizing knowledge syntheses: A taxonomy of literature reviews', *Knowledge in Society*, 1(1), pp. 104–126. doi: 10.1007/BF03177550.

Crathorne, A. R. and Shewhart, W. A. (1933) 'Economic Control of Quality of Manufactured Product.', *The American Mathematical Monthly*, 40(6), p. 353. doi: 10.2307/2301530.

Crosby, P. B. (1979) *Quality is Free: The Art of Making Quality Certain*. 1st Editio. New York: McGraw-Hill.

Curtis, D. J. (2020) *Trends in FDA Data Integrity 483s and Warning Letters for Pharmaceutical*, *astrix technology group*. Available at: https://astrixinc.com/trends-in-fda-data-integrity-483s-and-warning-letters-for-pharmaceutical-companies/ (Accessed: 15 May 2020).

Dahlgaard, J. J., Kristensen, K. and Khanji, G. K. (2005) *Fundamentals of Total Quality Management*. 1st Editio. London & New York: Taylor & Francis.

Dai, Y. *et al.* (2020) 'Twenty years in the making: The evolution of the journal of manufacturing & service operations management', *Manufacturing and Service Operations Management*, 22(1), pp. 1–10. doi: 10.1287/msom.2019.0791.

Dale, B. G., Dehe, B. and Bamford, D. (2016) 'Quality Management Systems and the ISO 9000 series', in Dale, Barrie G., van der Wiele, T., and Bamford, David (eds) *Managing quality: an essential guide and resource gateway*. 6th Editio. Chichester, UK: John Wiley & Sons, Ltd, pp. 161–180. doi:

10.1002/9781119302735.ch8.

Das, P., Mishra, S. and Nayak, B. (2017) 'The Colossal Role of QbD and PAT Tools in the Pharmaceutical Process Automation', *International Journal of Pharma Research and Health Sciences*, 5(6), pp. 1909–14. doi: 10.21276/ijprhs.20.

Deidda, R. *et al.* (2018) 'Risk-based approach for method development in pharmaceutical quality control context: A critical review', *Journal of Pharmaceutical and Biomedical Analysis*. Elsevier B.V., 161, pp. 110–121. doi: 10.1016/j.jpba.2018.07.050.

Deming, W. E. (1982) *Quality, Productivity and Competitive Position*. 1st Editio, *Quality and Reliability Engineering International*. 1st Editio. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Dresch, A., Lacerda, D. P. and Antunes Jr, J. A. V. (2015) *Design Science Research*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-07374-3.

Dresch, A., Pacheco Lacerda, D. and Cauchick-Miguel, P. A. (2019) 'Design science in operations management: conceptual foundations and literature analysis', *Brazilian Journal of Operations & Production Management*, 16(2), pp. 333–346. doi: 10.14488/BJOPM.2019.v16.n2.a13.

Duke University (2020) *Fit Random Forest Model*. Available at: http://code.env.duke.edu/projects/mget/export/HEAD/MGET/Trunk/PythonPackage/dist/TracOnlineDocumentation/Documentation/ArcGISReference/RandomForestModel.FitToArcGISTable.html (Accessed: 23 October 2020).

Dun & Bradstreet (2020) *Global Pharmaceutical Manufacturing Quality Study*. Available at: https://www.dnb.com/marketing/pharmaceutical-manufacturing-quality-study.html (Accessed: 27 October 2020).

Eger, F. *et al.* (2018) 'Zero Defect Manufacturing Strategies for Reduction of Scrap and Inspection Effort in Multi-stage Production Systems', *Procedia CIRP*. The Author(s), 67, pp. 368–373. doi: 10.1016/j.procir.2017.12.228.

Eich, S. *et al.* (2020) 'Progress of Operational Excellence implementation and performance in pharmaceutical manufacturing sites - descriptive study from 2003 to 2018', in *Procedings of the 27th EurOMA*. Warwick, UK: European Operations Management Association.

Eich, S. and Friedli, T. (2021) 'Analysis of the effects of Operational Excellence implementation on Inspection Outcomes in the Pharmaceutical Industry: An Empirical Study', *Brazilian Journal of Operations & Production Management*, 18(Special Issue-IJCIEOM).

Fahmy, R. *et al.* (2012) 'Quality by Design I: Application of Failure Mode Effect Analysis (FMEA) and Plackett–Burman Design of Experiments in the Identification of "Main Factors" in the Formulation and Process Design Space for Roller-Compacted Ciprofloxacin Hydrochloride Immediat', *AAPS PharmSciTech*, 13(4), pp. 1243–1254. doi: 10.1208/s12249-012-9844-x.

Fawcett, T. (2006) ‘An introduction to ROC analysis’, *Pattern Recognition Letters*, 27(8), pp. 861–874. doi: 10.1016/j.patrec.2005.10.010.

FDA - Center for Drug Evaluation and Research (2002) *Pharmaceutical CGMPs for the 21st Century: A Risk Based Approach*.

FDA - Center for Drug Evaluation and Research (2004) *Risk-Based Method for Prioritizing CGMP Inspections of Pharmaceutical Manufacturing Sites - A Pilot Risk Ranking Model*. Available at: http://www.fda.gov/oc/guidance/gmp.html (Accessed: 23 January 2019).

FDA - Center for Drug Evaluation and Research (2015) *Request for Quality Metrics - Draft Guidance*. Washington DC, USA. Available at: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm (Accessed: 4 October 2019).

FDA - Center for Drug Evaluation and Research (2016) *Submission of Quality Metrics Data Guidance for Industry - Draft Guidance*. Silver Spring, MD. Available at: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm (Accessed: 1 March 2019).

FDA - Center for Drug Evaluation and Research (2018) *Current Good Manufacturing Practice (CGMP) Regulations*. Available at: https://www.fda.gov/drugs/pharmaceutical-quality-resources/current-good-manufacturing-practice-cgmp-regulations (Accessed: 21 January 2020).

FDA - Center for Drug Evaluation and Research (2019a) *Drug Shortages: Root Causes and Potential Solutions*. Washington DC, USA.

FDA - Center for Drug Evaluation and Research (2019b) *Drug Shortages*. Washington DC: Center for Drug Evaluation and Research. Available at: https://www.fda.gov/Drugs/DrugSafety/DrugShortages/ucm441585.htm (Accessed: 13 March 2019).

FDA - Center for Drug Evaluation and Research (2020a) *Inspectional Observations and Citations*. Available at: https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-references/inspectional-observations-and-citations#A_Short_Description_of_Citations (Accessed: 21 May 2020).

FDA - Center for Drug Evaluation and Research (2020b) *Report on the State of Pharmaceutical Quality*. Available at: www.fda.gov (Accessed: 28 July 2020).

FDA - Center for Drug Evaluation and Research (2020c) *Results: Inspection Classification Database Search*, *Access Data*.

Ferdows, K. and De Meyer, A. (1990) ‘Lasting Improvements in Manufacturing Performance: In Search of a New Theory’, *Journal of Operations Management*, 9(2), pp. 168–184.

Ferdows, K. and Thurnheer, F. (2011) ‘Building factory fitness’, *International Journal of Operations & Production Management*, 31(9), pp. 916–934. doi: 10.1108/01443571111165820.

Fisher, M., Olivares, M. and Staats, B. R. (2020) 'Why Empirical Research Is Good for Operations Management, and What Is Good Empirical Operations Management?', *Manufacturing & Service Operations Management*, 22(1), pp. 170–178. doi: 10.1287/msom.2019.0812.

Friedli, T. *et al.* (2006) *Operational Excellence in the Pharmaceutical Industry*. 1st edn. ECV. Available at: https://www.ecv.de/Buecher/Operational_Excellence_in_the_Pharmaceutical_Industry (Accessed: 4 August 2020).

Friedli, T. *et al.* (2010) *The Pathway to Operational Excellence in the Pharmaceutical Industry: Overcoming the internal inertia*. Der pharma. Aulendorf: Editio Cantor Verlag. Available at: https://www.alexandria.unisg.ch/publications/70035 (Accessed: 19 December 2019).

Friedli, T. *et al.* (2013) *Leading Pharmaceutical Operational Excellence*. Edited by T. Friedli et al. Heidelberg, New York, Dordrecht, London: Springer. doi: 10.1007/978-3-642-35161-7.

Friedli, T. *et al.* (2017) *FDA Quality Metrics Research - 1st Year Report*. St.Gallen.

Friedli, T., Basu, P. K., *et al.* (2018) *21c Quality Management in the Pharmaceutical Industry - The Journey from Compliance to Excellence*. 1st Editio. Edited by T. Friedli et al. Aulendorf: Editio Cantor Verlag. Available at: http://dnb.ddb.de. (Accessed: 13 December 2018).

Friedli, T., Köhler, S., *et al.* (2018) *FDA Quality Metrics Research - 2nd Year Report*. St.Gallen.

Friedli, T. *et al.* (2019) *FDA Quality Metrics Research - 3rd Year Report*. St.Gallen.

Friedli, T., Goetzfried, M. and Basu, P. K. (2010) 'Analysis of the Implementation of Total Productive Maintenance, Total Quality Management, and Just-In-Time in Pharmaceutical Manufacturing', *Journal of Pharmaceutical Innovation*, 5(4), pp. 181–192. doi: 10.1007/s12247-010-9095-x.

Frost, J. (2019) *Regression Analysis*. 1st Editio. doi: 10.1016/S0140-6736(86)92832-1.

Garza-Reyes, J. A. *et al.* (2018) 'Lean readiness – the case of the European pharmaceutical manufacturing industry', *International Journal of Productivity and Performance Management*, 67(1), pp. 20–44. doi: 10.1108/IJPPM-04-2016-0083.

Gassmann, O. (1999) 'Praxisnähe mit Fallstudienforschung - Nutzen für das Technologiemanagement ist gegeben', *Wissenschaftsmanagement*, pp. 11–16.

Gassmann, O., Frankenberger, K. and Sauer, R. (2016) *Exploring the Field of Business Model Innovation*. doi: 10.1007/978-3-319-41144-6.

Gottlieb, S. (2018) *Statement from FDA Commissioner Scott Gottlieb, M.D., on the agency's global efforts to help assure product quality and transparency at*

foreign drug manufacturing facilities, *FDA - Center for Drug Evaluation and Research*. Available at: https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-agencys-global-efforts-help-assure-product-quality-and (Accessed: 4 October 2019).

Gray, J. V., Anand, G. and Roth, A. V. (2015) 'The influence of ISO 9000 certification on process compliance', *Production and Operations Management*, 24(3), pp. 369–382. doi: 10.1111/poms.12252.

Gray, J. V. and Massimino, B. (2014) 'The effect of language differences and national culture on operational process compliance', *Production and Operations Management*, 23(6), pp. 1042–1056. doi: 10.1111/poms.12212.

Gray, J. V., Roth, A. V. and Leiblein, M. J. (2011) 'Quality risk in offshore manufacturing: Evidence from the pharmaceutical industry', *Journal of Operations Management*. Elsevier B.V., 29(7–8), pp. 737–752. doi: 10.1016/j.jom.2011.06.004.

Gray, J. V., Roth, A. V. and Tomlin, B. (2016) *Contract Manufacturing and Quality Risk: Theory and Empirical Evidence*. Tuck School of Business Working Paper No. 2815520.

Gray, J. V., Siemsen, E. and Vasudeva, G. (2015) 'Colocation Still Matters: Conformance Quality and the Interdependence of R&D and Manufacturing in the Pharmaceutical Industry', *Management Science*, 61(11), pp. 2760–2781. doi: 10.1287/mnsc.2014.2104.

Greene, A., O'Donnell, K. and Calnan, N. (2018) *An Audience with International Regulators in the Manufacture of Medicines: Quality Risk Management (QRM) and Knowledge Management (KM)*. Dublin. doi: 10.21427/D7SZ5W.

Greener, S. (2008) *Business Research Methods*. Ventus Publishing.

Haddad, G. (2019) *Quality Risk Management: The Development of Role-Based Competency Model for the Biopharmaceutical Sector*. Technological University Dublin.

Haleem, R. M. *et al.* (2015) 'Quality in the pharmaceutical industry - A literature review', *Saudi Pharmaceutical Journal*. King Saud University, 23(5), pp. 463–469. doi: 10.1016/j.jsps.2013.11.004.

Hall, K. *et al.* (2016) 'Characteristics of FDA drug recalls: A 30-month analysis', *American Journal of Health-System Pharmacy*. Narnia, 73(4), pp. 235–240. doi: 10.2146/ajhp150277.

Handley, S. M. and Gray, J. V. (2015) 'Managing Quality in a Heterogeneous Contract Manufacturing Environment', *Decision Sciences*, 46(6), pp. 1011–1048. doi: 10.1111/deci.12147.

Harrell, F. E. J. *et al.* (1984) 'Regression modelling strategies for improved prognostic prediction', *Statistics in Medicine*, 3(2), pp. 143–152. doi: 10.1002/sim.4780030207.

Harrell, F. E. J., Lee, K. L. and Mark, D. B. (1996) 'Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy,

and measuring and reducing errors', *Statistics in Medicine*, 15(4), pp. 361–387. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

Hata, N. (2017) 'Risk Assessment and Lifecycle Management in Stability Testing', *Journal of GXP Compliance*, 21(5). Available at: http://www.ivtnetwork.com (Accessed: 11 March 2019).

Hayes, R. H. and Wheelwright, S. C. (1984) *Restoring Our Competitive Edge: Competing through Manufacturing.*, *Administrative Science Quarterly*. New York, NY: John Wiley & Sons. Available at: https://www.jstor.org/stable/2393118?origin=crossref (Accessed: 18 December 2019).

Hevner, A. R. *et al.* (2004) 'Design Science in Information Systems Research', *MIS Quarterly*, 28(1), p. 75. doi: 10.2307/25148625.

Ho, T. H. *et al.* (2017) 'OM forum - Causal inference models in operations management', *Manufacturing and Service Operations Management*, 19(4), pp. 509–525. doi: 10.1287/msom.2017.0659.

Ho, W. *et al.* (2015) 'Supply chain risk management: A literature review', *International Journal of Production Research*, 53(16), pp. 5031–5069. doi: 10.1080/00207543.2015.1030467.

ICH - International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (2005) *Quality Risk Management Q9*. Available at: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guideline.pdf (Accessed: 20 December 2018).

ICH - International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (2019) *Technical and regulatory considerations for pharmaceutical product lifecycle Q12*.

ISO, I. O. for S. (2016) *ISO/TS 9002:2016(en), Quality management systems — Guidelines for the application of ISO 9001:2015*. Available at: https://www.iso.org/obp/ui/#iso:std:iso:ts:9002:ed-1:v1:en (Accessed: 9 January 2020).

Jackson, T. L. (1996) *Implementing a Lean Management System*. 1st Editio. New York: Productivity Press. doi: 10.4324/9780367804893.

Johnson, R. A., Kast, F. E. and Rosenzweig, J. E. (1964) *Systems Theory and Management*, *Source: Management Science*. Available at: https://www.jstor.org/stable/pdf/2627306.pdf (Accessed: 8 April 2019).

Johnston, A. and Holt, D. W. (2014) 'Substandard drugs: a potential crisis for public health', *British Journal of Clinical Pharmacology*. Wiley-Blackwell, 78(2), pp. 218–243. doi: 10.1111/bcp.12298.

Jüttner, U. and Christopher, M. (2003) 'Supply Chain Risk Management: Outlining an Agenda for Future Research', *International Journal of Logistics : Research & Applications*, 6(4), pp. 197–210. doi: 10.1080/1367556031000162701 6.

Kalska-Zimnicka, M., Calnan, N. and O'Sullivan, F. (2017) 'De-risking Your Supply Chain: Considerations from a Supply Chain Maps Management Program', *Journal of GXP Compliance*, 21(5). Available at: http://www.ivtnetwork.com (Accessed: 11 March 2019).

Kaynak, H. (2003) 'The relationship between total quality management practices and their effects on firm performance', *Journal of Operations Management*, 21(4), pp. 405–435. doi: 10.1016/S0272-6963(03)00004-4.

Köhler, S. (2019) *Measuring Operational Excellence Performance – A Mixed-methods Conceptualization and Application in Pharmaceutical Quality Control Laboratories*. University of St.Gallen.

Köhler, S., Friedli, T. and Basu, P. K. (2019) 'Operational Excellence in Pharmaceutical Quality Control Labs: Driver of an Effective Quality System', *Journal of Pharmaceutical Innovation*.

Krafcik, J. F. (1998) 'Triumph of the Lean Production System', *Sloan Management Review*, 31(1).

Kubicek, H. (1976) 'Heuristische Bezugsrahmen und heuristisch angelegte Forschungsdesigns als Elemente einer Konstruktionsstrategie empirischer Forschung', *Empirische und handlungstheoretische Forschungskonzeptionen in der Betriebswirtschaftslehre*, pp. 3–36. Available at: http://www.tim.ethz.ch/education/courses/courses_fs_2007/course_docsem_fs_2007/literature/12_Kubicek_Heuristische_Bezugsrahmen.pdf.

Lantz, B. (2019) *Machine Learning with R*. Third Edit, *Machine Learning with R*. Third Edit. Birmingham - Mumbai: Packt Publishing Ltd. doi: 10.1007/978-981-10-6808-9.

Lee, N. and Lings, I. (2008) *Doing Business Research: A Guide to Theory and Practice*. Los Angeles, CA: SAGE Publications Ltd. Available at: http://dx.doi.org/10.1016/j.lisr.2010.03.001 (Accessed: 8 January 2020).

Mänder, C. and Friedli, T. (2018) 'Applicability of the St. Gallen Quality System Model – Case Study Reflection', in Friedli, T. et al. (eds) *21c Quality Management in the Pharmaceutical Industry*. 1st Editio. Aulendorf: Editio Cantor Verlag.

Martins, Y. S., Eduardo, C. and da Silva, C. E. S. (2019) 'Risk and ISO 9001: A systematic literature review', in Reis, J., Pinelas, S., and Melão, N. (eds) *Industrial Engineering and Operations Management, Springer Proceedings in Mathematics and Statistics*. IJCIEOM Lisbon: Springer Nature Switzerland, pp. 257–269. doi: 10.1007/978-3-030-14969-7_22.

Mcfarland, A. B. and Waldron, K. (2015) 'Considerations Regarding the Use of Detectability In Risk Assessments', *Institute of Validation Technology*, 21(4). Available at: http://www.ivtnetwork.com (Accessed: 11 March 2019).

Medina, C. and Richmond, F. J. (2015) 'Incentivizing Quality in the Manufacture of Pharmaceuticals', *Therapeutic Innovation & Regulatory Science*, 49(5), pp. 730–738. doi: 10.1177/2168479015589821.

Mezher, M. (2015) ‘Quality Metrics: Industry Reaction to FDA’s Draft Guidance’, *Regulatory Focus*, 25 August. Available at: https://www.raps.org/regulatory-focus™/news-articles/2015/8/quality-metrics-industry-reaction-to-fda-s-draft-guidance (Accessed: 4 October 2019).

Milmo, S. (2014) ‘Quality by design-bridging the gap between concept and implementation’, *BioPharm International*, 27(4), pp. 12–13. doi: 10.1208/s12248-014-9598-3.

Mišić, V. V. and Perakis, G. (2020) ‘Data analytics in operations management: A review’, *Manufacturing and Service Operations Management*, 22(1), pp. 158–169. doi: 10.1287/msom.2019.0805.

Moktadir, A. *et al.* (2018) ‘Decision modeling of risks in pharmaceutical supply chains’, *Industrial Management and Data Systems*, 118(7), pp. 1388–1412. doi: 10.1108/IMDS-10-2017-0465.

Mollah, A. H., Long, M. and Baseman, H. S. (2013) *Risk Management Applications in Pharmaceutical and Biopharmaceutical Manufacturing*. Edited by H. S. Mollah, A. Hamid Long, Mike Baseman. Hoboken, NJ, USA: John Wiley & Sons, Ltd.

Muse, A., Njeru, A. W. and Waiganjo, E. (2016) ‘Influence of Operational Excellence on the management of efficiency levels in the pharmaceutical industry in Kenya’, *International Journal of Academic Research in Business and Social Sciences*. Human Resource Management Academic Research Society, International Journal of Academic Research in Business and Social Sciences, 6(9), pp. 266–296.

Nagle, D. J. (2019) ‘Periodic Review of Validated Systems’, *Journal of Validation Technology*, 25(1).

Nally, J. D. (2007) *Good manufacturing practices for pharmaceuticals*. 6th Editio. Edited by J. D. Nally. Boca Raton: Informa Healthcare.

Nenni, M. E., Giustiniano, L. and Pirolo, L. (2014) ‘Improvement of manufacturing operations through a lean management approach: A case study in the pharmaceutical industry’, *International Journal of Engineering Business Management*, 6(1), pp. 1–6. doi: 10.5772/59027.

O’Connor, T. F. *et al.* (2017) ‘Quality risk management for pharmaceutical manufacturing: The role of process modeling and simulations’, in Pandey, P. and Bharadwaj, R. (eds) *Predictive Modeling of Pharmaceutical Unit Operations*. Amsterdam: Elsevier Ltd, pp. 15–37. doi: 10.1016/B978-0-08-100154-7.00002-8.

O’Connor, T. F., Yu, L. X. and Lee, S. L. (2016) ‘Emerging technology: A key enabler for modernizing pharmaceutical manufacturing and advancing product quality’, *International Journal of Pharmaceutics*. Elsevier B.V., 509(1–2), pp. 492–498. doi: 10.1016/j.ijpharm.2016.05.058.

O’Donnell, K. (2010) ‘Strategies for addressing the problems of subjectivity and uncertainty in quality risk management exercises: Part I: the role of human heuristics’, *Journal of Validation Technology*. UBM LLC, 16(3), pp. 76–85.

Available at: http://go.galegroup.com/ps/anonymous?id=GALE%7CA238946523&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=10796630&p=AONE&sw=w (Accessed: 25 January 2019).

O'Donnell, K. and Greene, A. (2007) 'Failure Modes: Simple Strategies for Improving Qualitative Quality Risk Management Exercises during Qualification, Validation, and Change Control Activities', *Journal of Validation Technology*, 13(2), pp. 100–112. Available at: https://search.proquest.com/docview/205482897/fulltextPDF/199807CDD08740EBPQ/1?accountid=28962 (Accessed: 25 January 2019).

Ohio State University (2019) *Fisher among five institutions selected as part of $1.7M FDA research contract | Fisher College of Business*. Available at: https://fisher.osu.edu/news/fisher-among-five-institutions-selected-part-1.7m-fda-research-contract#.Xc3HKwbmEU4.linkedin (Accessed: 26 October 2020).

Ohno, T. (1988) *Toyota Production System - Beyond Large-Scale Production*. New York: Productivity Press.

Olsen, T. L. and Tomlin, B. (2020) 'Industry 4.0: Opportunities and challenges for operations management', *Manufacturing and Service Operations Management*, 22(1), pp. 113–122. doi: 10.1287/msom.2019.0796.

Patil, A. S. and Pethe, A. M. (2013) 'Quality by design (QbD): A new concept for development of quality pharmaceuticals', *International Journal of Pharmaceutical Quality Assurance*, 4(2), pp. 13–19. doi: 10.1007/s11095-007-9511-1.

Pavlović, K. and Božanić, V. (2010) 'Lean and Six Sigma Concepts – Application in Pharmaceutical Industry', *International Journal for Quality Research*, 5(2), pp. 259–268.

Peduzzi, P. *et al.* (1996) 'A simulation study of the number of events per variable in logistic regression analysis', *Journal of Clinical Epidemiology*, 49(12), pp. 1373–1379. doi: 10.1016/S0895-4356(96)00236-3.

Pettersen, J. (2009) 'Defining lean production: Some conceptual and practical issues', *TQM Journal*, 21(2), pp. 127–142. doi: 10.1108/17542730910938137.

Pfeifer, P. D.-I. h. c. P. h. c. T. (2002) 'Produkt- und Prozessqualität in Netzwerken', in Milberg, J. and Schuh, G. (eds) *Erfolg in Netzwerken*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-56175-7.

Pharmaceutical Technology (2017) *FDA to Fund Manufacturing Research Programs at NIPTE*. Available at: https://www.pharmtech.com/view/fda-fund-manufacturing-research-programs-nipte (Accessed: 26 October 2020).

PIC/S - Pharmaceutical Inspection Convention | Pharmaceutical Inspection Co-Operation Scheme (2012) *A Recommended Model for Risk-Based Inspection Planning in the GMP Environment*.

Pollack, A. (2010) 'Genzyme Drug Shortage Leaves Users Feeling Betrayed', *The New York Times*, 15 April.

Porsche Consulting (2013) *Operational Excellence in the Pharmaceutical Industry Survey 2013*.

Qing, L. *et al.* (2014) 'Quality Risk Management Model for Railway Construction Projects', *Procedia Engineering*, 84, pp. 195–203. doi: 10.1016/j.proeng.2014.10.426.

Raban, C. and Turner, L. (2006) 'Quality risk management. Modernising the architecture of quality assurance', *perspective*, 10(2). doi: 10.1080/13603100600644522.

Radnor, Z. J. and Barnes, D. (2007) 'Historical analysis of performance measurement and management in operations management', *International Journal of Productivity and Performance Management*, 56(5–6), pp. 384–396. doi: 10.1108/17410400710757105.

Ramnarine, E. (2015) 'Understanding Problems of Subjectivity and Uncertainty in Quality Risk Management', *Journal of Validation Technology*, 21(4), p. 2015.

Ramnarine, E. *et al.* (2017) 'PDA Points to Consider: Technical Product Lifecycle Management. Pharmaceutical Quality System Effectiveness for Managing Post-approval Changes', *PDA Journal of Pharmaceutical Science and Technology*. Parenteral Drug Association (PDA), 71(3), pp. 252–258. doi: 10.5731/pdajpst.2017.007575.

Ramnarine, E. and O'Donnell, K. (2018) 'Demonstrating Pharmaceutical Quality System Effectiveness and Driving Continual Improvement: Evidence-Based Risk Reduction', *PDA Journal of Pharmaceutical Science and Technology*, 72(3), pp. 338–345. doi: 10.5731/pdajpst.2017.008524.

Rampini, G. H. S., Berssaneti, F. T. and Saut, A. M. (2019) 'Insertion of Risk Management in Quality Management Systems with the Advent of ISO 9001:2015: Descriptive and Content Analyzes', in Reis, J., Pinelas, S., and Melão, N. (eds) *Industrial Engineering and Operations Management, Springer Proceedings in Mathematics and Statistics*. IJCIEOM Lisbon: Springer Nature Switzerland, pp. 209–221. doi: 10.1016/0377-2217(85)90018-9.

Richter, L. and Haddad, G. (2015) 'Role of Senior Leadership in Quality Risk Management', *Journal of Validation Technology*, 21(4), p. 18.

Roth, A. and Rosenzweig, E. (2020) 'Advancing empirical science in operations management research: A clarion call to action', *Manufacturing and Service Operations Management*, 22(1), pp. 179–190. doi: 10.1287/msom.2019.0829.

Saraph, J. V., Benson, P. G. and Schroeder, R. G. (1989) 'An Instrument for Measuring the Critical Factors of Quality Management', *Decision Sciences*, 20(4), pp. 810–829. doi: 10.1111/j.1540-5915.1989.tb01421.x.

Schuh, G. (2013) *Lean Innovation*. Berlin, Heidelberg: Imprint: Springer Vieweg.

Seiss, M. (2018) *FDA CDER ORA Site Selection Model Improvement Pilot Study*.

Shah, R., Ball, G. and Netessine, S. (2017) 'Plant Operations and Product Recalls in the Automotive Industry: An Empirical Investigation', *Management*

Science, 63(8). doi: 10.2139/ssrn.2356315.

Shah, R. and Ward, P. T. (2003) 'Lean manufacturing: context, practice bundles, and performance', *Journal of Operations Management*, 21(2), pp. 129–149. doi: 10.1016/S0272-6963(02)00108-0.

Shah, R. and Ward, P. T. (2007) 'Defining and developing measures of lean production', *Journal of Operations Management*, 25(4), pp. 785–805. doi: 10.1016/j.jom.2007.01.019.

Shingo, S. and Dillon, A. P. (1989) *A Study of the Toyota Production System: From an Industrial Engineering Viewpoint*. Re-Transla. New York: Productivity Press. Available at: https://books.google.ch/books?id=RKWU7WElJ7oC&pg=PA22&redir_esc=y#v=onepage&q&f=false (Accessed: 8 January 2020).

Sirkia, R. and Laanti, M. (2015) 'Adaptive Finance and Control: Combining Lean, Agile, and Beyond Budgeting for Financial and Organizational Flexibility', in *2015 48th Hawaii International Conference on System Sciences*. IEEE, pp. 5030–5037. doi: 10.1109/HICSS.2015.596.

Smiddy, H. and Naum, L. (1954) 'Evolution of a "science of managing" in America', *Management Science*, 1(1), pp. 1–31.

Soto, I. (2018) 'How to Implement a Risk-Based Requalification Program', *Journal of Validation Technology*, 24(6).

Stephen Barlas (2017) 'Drug Industry Groups Oppose FDA "Quality Metrics" Plan - The Agency Wants to Use Manufacturing Data to Attack Drug Shortages', *Pharmacy & Therapeutics*, 42(7), pp. 446–448. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5481295/pdf/ptj4207446.pdf (Accessed: 13 March 2019).

Stricker, N. and Lanza, G. (2014a) 'The concept of robustness in production systems and its correlation to disturbances', *Procedia CIRP*. Elsevier B.V., 19(C), pp. 87–92. doi: 10.1016/j.procir.2014.04.078.

Stricker, N. and Lanza, G. (2014b) 'The concept of robustness in production systems and its correlation to disturbances', *Procedia CIRP*. Elsevier B.V., 19(C), pp. 87–92. doi: 10.1016/j.procir.2014.04.078.

Tabachnick, B. G. and Fidell, L. S. (2014) *Using multivariate statistics*.

Tennant, G. (2001) *Six Sigma: SPC and TQM in Manufacturing and Services*, *Work Study*. Burlington, VT: Gower. doi: 10.1108/ws.2002.07951cae.002.

The FDA Group (2019) *2019 FDA Warning Letter & Inspection Observation Trends*. Available at: https://www.thefdagroup.com/blog/2019-fda-warning-letter-inspection-observation-trends (Accessed: 15 May 2020).

The Institute of Internal Auditors (2013) *IIA Position Paper: The Three Lines of Defense in Effective Risk Management and Control*. Available at: https://na.theiia.org/standards-guidance/Public Documents/PP The Three Lines of Defense in Effective Risk Management and Control.pdf (Accessed: 7 June 2019).

Tomczack, T. (1992) ‘Abhandlungen Forschungsmethoden in der Marketingwissenschaft’, *Marketing Zeitschrift für Forschung und Praxis*, 14(2), pp. 77–87.

Torkko, M. *et al.* (2013) ‘Quality KPIs in Pharmaceutical and Food Industry’, *Journal of Pharmaceutical Innovation*. Springer US, 8(4), pp. 205–211. doi: 10.1007/s12247-013-9159-9.

Tse, Y. K. *et al.* (2011) ‘Quality risk in global supply network’, *Journal of Manufacturing Technology Management*, 22(8), pp. 1002–1013. doi: 10.1108/17410381111177458.

Ulrich, H. (1982) ‘Anwendungsorientierte Wissenschaft’, *Die Unternehmung: Swiss journal of business research and practice; Organ der Schweizerischen Gesellschaft für Betriebswirtschaft (SGB)*, 36(1), pp. 1–10.

Ulrich, H. (1984) ‘Die Betriebswirtschaftslehre als anwendungsorientierte Sozialwissenschaft’, in Ulrich, H., Dyllick, T., and Probst, G. (eds) *Management*. Bern/Stuttgart: Paul Haupt, pp. 168– 199.

Vesper, J. (2015) ‘Q9+ Ten Years Examining Risk Communication’, *Journal of Validation Technology*, 21(4). Available at: http://www.ivtnetwork.com/article/q9-ten-years-examining-risk-communication.

Viehmann, A. (2020) ‘Quality Metrics and Quality Management Maturity’. Washington DC, USA: FDA - Center for Drug Evaluation and Research.

Voss, C. A. *et al.* (1995) ‘The Competitiveness of European Manufacturing - A Four Country Study’, *Business Strategy Review*, 6(1), pp. 1–25. doi: 10.1111/j.1467-8616.1995.tb00169.x.

Waldron, K. (2017) *Managing Risk to the Patient: Recoding Quality Risk Management for the Pharmaceutical and Biopharmaceutical Industries*, *Doctoral*. doi: 10.21427/D7G165.

Waller, M. A. and Fawcett, S. E. (2013) ‘Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management’, *Journal of Business Logistics*, 34(2), pp. 77–84. doi: 10.1111/jbl.12010.

Walters, L. and Barneva, R. P. (2017) ‘Prioritization of Process Improvement Using Risk Evaluation in the Manufacturing of Biologics’, *Quality Management Journal*, 24(2), pp. 31–40. doi: 10.1080/10686967.2017.11918507.

Webster, J. and Watson, R. T. (2002) *Analyzing the Past to Prepare for the Future: Writing a Literature Review*, *MIS Quarterly*. Available at: http://www.misq.org/misreview/announce.html (Accessed: 1 December 2019).

Weckenmann, A., Akkasoglu, G. and Werner, T. (2015) ‘Quality Management – History and Trends’, *The TQM Journal*. Edited by P. Vidosav Majstorovic and Dr Albert W, 27(3), pp. 281–293. doi: 10.1108/TQM-11-2013-0125.

Wheelwright, S. C., Hayes, R. H. and Wheelwright, S. C. (1985) ‘Competing Through Manufacturing’, *Harvard Business Review*, 63(1), pp. 99–109. Available at: https://hbr.org/1985/01/competing-through-manufacturing (Accessed: 18 December 2019).

Womack, J. P. and Jones, D. T. (2003) *Lean Thinking*. 2nd Editio. New York, NY: Free Press.

Womack, J. P., Jones, D. T. and Roos., D. (1990) *The Machine that changed the world. Simon and Schuster*. New York: Rawson Associates.

Woodcock, J. *et al.* (2012) 'Reliable Drug Quality: An Unresolved Problem', *PDA Journal of Pharmaceutical Science and Technology*. Parenteral Drug Association (PDA), 66(3), pp. 270–272. doi: 10.5731/pdajpst.2012.00868.

Yu, L. X. and Kopcha, M. (2017) 'The future of pharmaceutical quality and the path to get there', *International Journal of Pharmaceutics*, 528(1–2), pp. 354–359. doi: 10.1016/j.ijpharm.2017.06.039.

Zhang, L.; Narkhede, B. E.; Chaple, A. P. (2017) 'Evaluating lean manufacturing barriers: an interpretive process', *Journal of Manufacturing Technology Management*. Edited by N. B. Eknath. Emerald Publishing Limited, 28(8), pp.

A Appendix

A.1 Literature review table on Inspection outcomes

The following pages show the literature review on research based on FDA inspection outcomes and quality risk in an enlarged version for better readability. Please see Table 3 for context.

Article	Sample	Industry	Independent variables	IV/DV	Dependent variables	Control variables
Gray et al., 2011	672 inspections of 60 plants	Pharma	Regional high school education share; regional industry cluster, site distance from company HQ, US mainland / offshore	DV	Delphi-process based score of inspection final decision and form 483 issuance	Plant age (years), plant size (number of employees), and company size (sales)
Anand, Gray, & Siemsen, 2012	8080 inspections of 1493 plants	Pharma	Time since last inspection; merger/acquisition prior to inspection	DV	Factor operationalization: „OAI“ / „VAI“ / „NAI“	Inspection type, previous inspection outcome, year, foreign location
Gray & Massimino, 2014	7696 inspections	Pharma	Language and cultural differences between site and head quarter	DV	Factor operationalization: „OAI“ / „VAI“ / „NAI“	Previous inspection outcome, time since previous inspection, Ownership change, head quarter-plant distance
Gray, Siemsen, et al., 2015	8832 inspections	Pharma	Firm size, site (co-)location with R&D, information and communication technologies (ICT) availability, Tacit process knowledge	DV	Binary operationalization: „OAI“ vs. rest	Plant type, age, population density, R&D intensity

Terminology	Methods	Results	Year
Quality risk	Paired sample; Ordinary least squares regression	Used individually in regression models, regional high school completion quota, industry agglomeration, proximity to headquarter, and mainland location reduce quality risk	2011
Process entropy	Random Coefficients Ordered-Probit Analysis	Good states of process entropy deteriorate over time; mergers negatively affect entropy whereas acquisitions do positively	2012
Process compliance	Ordered probit regression, polynomial regression	Time since last inspection and language difference clearly harm process compliance; no clear picture for cultural differences, but similar power distance at HQ and site has positive impact	2014
Conformance quality	T-test, (two-stage) probit regression	Colocation improves quality conformance regardless of ICT access or time; colocation especially important for large companies and complex products	2015

Article	Sample	Industry	Independent variables	IV/DV	Dependent variables	Control variables
Gray, Anand, et al., 2015	Overall, 552 plants	Medical devices	(Early) ISO 9000 certification and time since certification	DV	See Gray et al., 2011	Public companies, standalone sites, plant size
Handley & Gray, 2015	106 supplier – customer pairs	Food, pharma, and medical device	CM's Manufacturing heterogeneity, coordination mechanisms eg. Cooperative relationship (moderator), attention mechanisms eg. Formal performance assessments (moderator)	-	Four level scale survey	Regulatory intensity, CM size, plant age, lenght of relationship, process type
Gray et al., 2016	1253 inspections	Pharma	Contract manufacturing (CM) vs. Internal plant, manufacturing experience (moderator), inspection intensity (moderator)	DV	See Gray et al., 2011	Company sales
Altamuro et al., 2017	216 companies	Pharma	Internal control weakness	DV	See Gray et al., 2011; Financial non-compliance (accounting restatements)	Management ability, ROA, sales growth & volatility, loss, leverage ratio (total debt / assets), firm size, complexity (number of geographic & business segments), Foreign income / Foreign plant, 'Big Four' auditor, free cash, new debt / equity

Terminology	Methods	Results	Year
Process compliance	Pearson Bivariate Correlation, T-Test	ISO 9000 certification has a positive effect on process compliance if obtained early; the effect decays over time	2015
Conformance Quality Performance	Regression	Negative link between heterogeneity and CM‘s conformance quality performance; contractual coordination provisions and formalized performance assessments can minimize the adverse effect of heterogeneity on the CM’s conformance quality performance	2015
Quality risk	Ordinary least squares regression	CMs operate at higher risk, but more experience reduces risk; higher inspection intensity reduces risk particularly for CMs	2016
Operational non-compliance	Probit regression	Weak compliance culture harms operational and financial non-compliance; contemporary financial and operational non-compliance correlate	2017

Article	Sample	Industry	Independent variables	IV/DV	Dependent variables	Control variables
Ball et al., 2017	2231 recalls	Medical device	Inspection outcome, inspector (site specific) experience	IV	Recall classification (class I vs. II)	Sales, Firm, Public companies, local college completion ratio, Time since last recall, previous recalls, product scope, number of products, year, products‘ technological level, product novelty, inspection type
Shah, Ball, & Netessine, 2017	67 recalls	Auto-motive	Product and plant variety, utilization	-	Recall	Plant size (volume, employees), year, manufacturer, plant age, years since launch, segment
Seiss, 2018 (non-public	3429 inspections	Pharma	Previous inspection outcomes, time since last inspection, financial data	DV	Binary operationalization: „NAI“ vs. rest	-

Terminology	Methods	Results	Year
-	Cox regression analysis	Inspection outcomes does predict future recall severity; site-specific experience of the investigator does harm that relationship	2017
-	Correlation matrix, generalized estimating equation GEE regression	High variety and utilization increase the likeliness of recalls, particularly when combined	2017
-	Logistic regression & scorecard prediction models	Quality compliance history does somewhat predict future inspection outcomes; financial data slightly improves the predictions	2018

A.2 Case interview guidelines

Dissertation case interview with partner company

Institute of Technology Management
University of St.Gallen

– Steffen Eich

1. Introduction

- Dissertation-Topic: Using data science to find predictors of adverse regulatory inspections –

Enhancing quality risk management in the pharmaceutical industry.

- Interview-Objective: Validating analyses results and understanding practical impact.

2. Corporate Approach to Quality Compliance

- Do you have a measure for the state of quality compliance or inspection readiness?

- Do you track past inspections with a measure beyond individual inspection results? How?

- Do you have a data analysis approach to ensuring GMP compliance at the sites? What are the specific use cases?

Institute of Technology Management

University of St.Gallen

Dissertation case interview with regulatory agency

– Steffen Eich

1. Introduction

- Dissertation-Topic: Using data science to find predictors of adverse regulatory inspections –

Enhancing quality risk management in the pharmaceutical industry.

- Interview-Objective: Validating analyses results and understanding practical impact.

2. Organizational Approach to Quality Compliance

- Do you have a measure for the state of quality compliance or inspection priority?

- Do you track past inspections with a measure beyond individual inspection results? How?

- Do you have a data analytics approach to ensuring GMP compliance at the sites? What are the specific use cases?

A.3 R Code: Analyses for Chapter 3

The following code is provided as two separate R notebooks with commentary and an interactive table of contents for the overall manufacturing site and QC lab benchmarking database, respectively.

A.3.1 Operational Excellence benchmarking database

````
---
title: "OPEX-BM_FDA"
author: "Steffen Eich"
date: "23 Januar 2020"
output: word_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
````

 # 1. Setup

Install necessary packages:

````
```{r}
#Only run this once unless there are problems with the file
rm(list=ls())
remove.packages(c("ggplot2", "data.table","tidyverse", "Rcpp", "dplyr"))
install.packages('Rcpp', dependencies = TRUE)
install.packages('broom', dependencies = TRUE, type="binary")
install.packages('ggplot2', dependencies = TRUE)
install.packages('data.table', dependencies = TRUE)
````
````

```
install.packages("tidyverse", dependencies = TRUE)
install.packages("dplyr", dependencies = TRUE)
```

Load packages

```{r}
library(Rcpp)
# load packages
#install.packages(readxl)
library(readxl)
#install.packages("xlsx")
library(xlsx)
#install.packages("tidyverse")
library(tidyverse)
#install.packages("reshape2")
library(reshape2)
#install.packages("zoo")
library(zoo)
library(ggplot2)
#install.packages("ggcorrplot")
library(ggcorrplot)
library(tibble)
library(dplyr)
library(rpart)
library(rpart.plot)
library(caret)
library(rattle)
library(magrittr)
library(pROC)
```

```
#install.packages("gee")
library(gee)
#install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)
```

 # 2. Load Data

Next, working directory and input filename are defined.

```{r}
rm(list=ls())
setwd("C:/Users/SEich/switchdrive/01 Forschung/04 Daten")
INPUT_FILE <- "C:/Users/SEich/switchdrive/01 Forschung/04 Daten/OPEX BM data + FDA.xlsx"
```

Data import

```{r}
Enabler <- read_xlsx(INPUT_FILE)
```

 # 3. Data preparation

3.1. Normalization & Missing data

```{r}
#Enabler_Clean <- Enabler[!is.na(Enabler$Enabler_Overall),]
Enabler_Insp <- Enabler[!is.na(Enabler$FDA_Insp_Outc),]
Enabler_Insp <- Enabler_Insp %>%
mutate(FDA_Insp_Bin = ifelse(FDA_Insp_Outc == "NAI", 1, 0))
Enabler_Insp[, c(1,3:50)] <- sapply(Enabler_Insp[, c(1,3:50)], as.numeric)
Enabler_Insp[, c(2,51)] <- sapply(Enabler_Insp[, c(2,51)], as.factor)
```

```
str(Enabler_Insp)
#Enabler_Insp <- Enabler_Insp %>%
#  mutate(
#    FDA_Insp_Outc = FDA_Insp_Outc - 1)
Enabler_Insp_Scale <- Enabler_Insp %>% mutate_at(c(3:49), funs(c(scale(.))))
#Enabler_Insp_Scale <- Enabler_Insp %>% mutate_at(c("Enabler_Overall", "Enabler_TPM", "Enabler_TQM", "Enabler_JIT", "Enabler_EMS", "Enabler_BE", "PQS_Effectiveness", "OS_SCORE", "PrevMaint", "Tech", "FiveS", "ProcMgmt", "XfunctPD", "CustInvol", "SupplQual", "Setup", "Pull", "Layout", "PlanAdh", "Hoshin", "LeadCult", "CI", "FuncIntQual", "Standard", "VisMgmt", "UnplannedMaintenance", "OEE", "LotAcceptanceRate", "DeviationsPerBatch", "Yield", "ScrapRate", "ProductionScheduleAccuracy", "ReleaseTime", "DeviationClosureTime", "CustomerComplaintRate", "RightFirstTime"), ~(scale(.) %>% as.vector))
Enabler_Insp[, c(1,3:50)] <- sapply(Enabler_Insp[, c(1,3:50)], as.numeric)
Enabler_Insp[, c(2,52)] <- sapply(Enabler_Insp[, c(2,52)], as.factor)
summary(Enabler_Insp_Scale)
str(Enabler_Insp_Scale)
```

3.2. Colinearity

```{r}
corrCat <- round(cor(Enabler_Insp_Scale[,c(4:8)]), 1)
head(corrCat)
ggcorrplot(corrCat, hc.order = TRUE, type = "lower", lab = TRUE)
chart.Correlation(Enabler_Insp_Scale[,c(4:8)], histogram=TRUE, pch=19)
chart.Correlation(Enabler_Insp_Scale[,c(11:27)], histogram=TRUE, pch=19)
chart.Correlation(Enabler_Insp_Scale[,c(28:38)], histogram=TRUE, pch=19)
```

3.3. Missing data - Performance

```{r}
```

```
All_Insp <- Enabler_Insp_Scale[!is.na(Enabler_Insp_Scale$OS_SCORE),]
All_Insp[, c(1:11)] <- sapply(All_Insp[, c(1:11)], as.numeric)
str(All_Insp)
summary(All_Insp)
#str(Enabler_Insp_Scale)
Enabler_Insp_Scale <- Enabler_Insp %>% mutate_at(c(#"FDA_Insp_Outc",
  "Year",  "Enabler_Overall",  "Enabler_TPM",  "Enabler_TQM",  "Enabler_JIT",
"Enabler_EMS", "Enabler_BE", "OS_SCORE"), ~(scale(.) %>% as.vector))
OS_Clean <- Enabler[!is.na(Enabler$OS_SCORE),]
OS_Insp <- OS_Clean[!is.na(OS_Clean$FDA_Insp_Bin),]
OS_Insp[, c(1:11)] <- sapply(OS_Insp[, c(1:11)], as.numeric)
#OS_Insp[, c(2:3)] <- sapply(OS_Insp[, c(2:3)], as.factor)
#Enabler_Insp <- Enabler_Insp %>%
#  mutate(
#    FDA_Insp_Outc = FDA_Insp_Outc - 1)
```

3.4. Colinearity - Performance

```{r}
corr <- round(cor(Enabler[,c(31, 38, 45)]), 1)
head(corr)
ggcorrplot(corr, hc.order = TRUE, type = "lower", lab = TRUE)
chart.Correlation(Enabler[,c(31, 38, 45)], histogram=TRUE, pch=19, na.action =
na.omit)
```

 # 4. GEE regression

```{r}
```

```
gee_mod <- gee(formula = FDA_Insp_Bin ~ #Enabler_Overall +
    Enabler_TPM + Enabler_TQM + Enabler_JIT + Enabler_EMS + Enabler_BE
   , id = Year, data = Enabler_Insp_Scale, na.action = na.omit, family = gaussian)
summary(gee_mod)
```

```
# 5. MLR
```{r}
mlr <- lm(formula = FDA_Insp_Outc ~ #Enabler_Overall +
    Enabler_TPM + Enabler_TQM + Enabler_JIT + Enabler_EMS + Enabler_BE
   , data = Enabler_Insp_Scale, na.action = na.omit)
summary(mlr)
```
```

 # 6. LGR

6.1. Enabler

```
```{r}
# Overall Enablers
LGROverall <- glm(formula = FDA_Insp_Bin ~ Enabler_Overall
   , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGROverall)
LGROverallContr <- glm(formula = FDA_Insp_Bin ~ Enabler_Overall + LC + Size
   , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGROverallContr)
1-pchisq(74.095-64.189,55-52)
```
```

```
# Category Enablers
LGRCat <- glm(formula = FDA_Insp_Bin ~ Enabler_TPM + Enabler_TQM + Enabler_JIT + Enabler_EMS + Enabler_BE
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRCat)
1-pchisq(76.992-63.145,57-52)
LGRCatContr <- glm(formula = FDA_Insp_Bin ~ Enabler_TPM + Enabler_TQM + Enabler_JIT + Enabler_EMS + Enabler_BE + LC + Size
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRCatContr)
1-pchisq(74.095-54.611,55-48)

# TQM Enablers
LGRTQM <- glm(formula = FDA_Insp_Bin ~ Enabler_TQM - 1
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRTQM)
1-pchisq(80.405-71.617,58-57)
LGRTQM <- glm(formula = FDA_Insp_Bin ~ Enabler_TQM - 1
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRTQM)
1-pchisq(80.405-71.617,58-57)
LGRTQMContr <- glm(formula = FDA_Insp_Bin ~ Enabler_TQM + LC + Size
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRTQMContr)
1-pchisq(74.095-58.925,55-52)

# Sub-category Enablers
LGRSub <- glm(formula = FDA_Insp_Bin ~ #Enabler_Overall +
```

```
      PrevMaint + Tech + FiveS + ProcMgmt + XfunctPD + CustInvol + SupplQual
+ Setup + Pull + Layout + PlanAdh + Hoshin + LeadCult + CI + FuncIntQual +
Standard + VisMgmt
   , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRSub)
1-pchisq(75.025-51.640,56-39)
LGROS <- glm(formula = FDA_Insp_Bin ~ #Enabler_Overall +
      OS_SCORE
   , family = "binomial", data = OS_Insp, na.action = na.omit)
summary(LGROS)
LGRCatOS <- glm(formula = FDA_Insp_Bin ~ #Enabler_Overall +
      TPM + TQM + JIT + EMS + BE + OS_SCORE
   , family = "binomial", data = All_Insp, na.action = na.omit)
summary(LGRCatOS)

#LGR <- train(FDA_Insp_Bin ~ #Enabler_Overall +
#           Enabler_TPM + Enabler_TQM + Enabler_JIT + Enabler_EMS +
Enabler_BE
#           , data = Enabler_Insp_Scale, method="glm", trControl = ctrl,
family="binomial")
```

6.2. Performance

```
# OS_SCORE, PQS_Effectiveness, SR_SCORE, UnplannedMaintenance,
      OEE, LotAcceptanceRate, DeviationsPerBatch, Yield,
      ScrapRate, ProductionScheduleAccuracy, ReleaseTime,
      DeviationClosureTime, CustomerComplaintRate, RightFirstTime,
MaintConvCost, QualConvCost, SetupClean,
      SupplierComplaintRate, SupplierServiceLevel, OTIF, TPM_Score,
      TQM_Score, JIT_Score, OPEX_Score
```

```
LGRPerf <- glm(formula = FDA_Insp_Bin ~ LotAcceptanceRate + CustomerComplaintRate # + OTIF
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRPerf)
1-pchisq(49.461-25.166,35-32)
LGRPerfContr <- glm(formula = FDA_Insp_Bin ~ OTIF # + CustomerComplaintRate + LotAcceptanceRate
                + LC + Size
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRPerfContr)
1-pchisq(61.513-47.296,46-43)
LGRPerfContr <- glm(formula = FDA_Insp_Bin ~ OTIF + CustomerComplaintRate + LotAcceptanceRate
                + LC + Size
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRPerfContr)
1-pchisq(49.461-24.101,35-30)
```

6.3. Combined

```{r}
# OS_SCORE, PQS_Effectiveness, SR_SCORE, UnplannedMaintenance,
        OEE, LotAcceptanceRate, DeviationsPerBatch, Yield,
        ScrapRate, ProductionScheduleAccuracy, ReleaseTime,
        DeviationClosureTime, CustomerComplaintRate, RightFirstTime,
MaintConvCost, QualConvCost, SetupClean,
        SupplierComplaintRate, SupplierServiceLevel, OTIF, TPM_Score,
        TQM_Score, JIT_Score, OPEX_Score
```

```
LGRComb <- glm(formula = FDA_Insp_Bin ~ CustomerComplaintRate + LotAcceptanceRate + OTIF + Enabler_TQM
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRComb)
1-pchisq(49.461- 18.725,35-31)

LGRCombContr <- glm(formula = FDA_Insp_Bin ~ CustomerComplaintRate + LotAcceptanceRate + OTIF + Enabler_TQM
                    + LC + Size
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRCombContr)
1-pchisq(49.461- 17.366,35-29)
```

A.3.2 QC Lab benchmarking database

```
---
title: "OPEX-BM_FDA"
author: "Steffen Eich"
date: "23 Januar 2020"
output: word_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

 # 1. Setup

Necessary packages:

```{r}
# load packages
library(readxl)
#install.packages("xlsx")
library(xlsx)
#install.packages("tidyverse")
library(tidyverse)
#install.packages("reshape2")
library(reshape2)
#install.packages("zoo")
library(zoo)
library(ggplot2)
#install.packages("ggcorrplot")
library(ggcorrplot)
```
```

```
library(tibble)
library(dplyr)
library(rpart)
library(rpart.plot)
library(caret)
library(rattle)
library(magrittr)
library(pROC)
#install.packages("gee")
library(gee)
#install.packages("PerformanceAnalytics")
library("PerformanceAnalytics")
#install.packages("aod")
library("aod")
install.packages("ggpubr")
library(ggpubr)
```

 # 2. Load Data

Next, working directory and input filename are defined.

```{r}
rm(list=ls())
setwd("C:/Users/SEich/switchdrive/01 Forschung/04 Daten")
INPUT_FILE <- "C:/Users/SEich/switchdrive/01 Forschung/04 Daten/QC_Database + FDA Data.xlsx"
```

Data import

```{r}
```

```
Enabler <- read_xlsx(INPUT_FILE)
```

```

# 3. Data preparation

## 3.1. Missing data

```{r}

Enabler_Clean <- Enabler[!is.na(Enabler$Overall),]

Enabler_Clean[, c(2:17,22:51)] <- sapply(Enabler_Clean[, c(2:17,22:51)], as.numeric)

Enabler_Clean[, c(18:21,52)] <- sapply(Enabler_Clean[, c(18:21,52)], as.factor)

Enabler_Insp <- Enabler_Clean[!is.na(Enabler_Clean$Last_Insp_outcome),]

Enabler_Insp[, c(2:17,22:51)] <- sapply(Enabler_Insp[, c(2:17,22:51)], as.numeric)

Enabler_Insp[, c(18:21,52)] <- sapply(Enabler_Insp[, c(18:21,52)], as.factor)

#Enabler_Insp <- Enabler_Insp %>%

#  mutate(

#    FDA_Insp_Outc = FDA_Insp_Outc - 1)

#Enabler_Insp_Scale <- Enabler_Insp %>% mutate_at(c(#"FDA_Insp_Outc",

#"MaintQualSys", "PrevMaint", "Tech", "FiveS", "ProcMgmt", "PlanSteer", "Setup", "Pull", "Layout", "PlanAdh", "VisMgmt", "MgmtSys", "LeadCult", "CI", "FuncIntQual", "Overall"), ~(scale(.) %>% as.vector))


Enabler_Insp_Scale <- Enabler_Insp %>% mutate_at(c(2:17,22:51), funs(c(scale(.))))

Enabler_Insp_Scale[, c(2:51)] <- sapply(Enabler_Insp_Scale[, c(2:51)], as.numeric)

#Enabler_Insp_Scale[, c(18:21)] <- sapply(Enabler_Insp_Scale[, c(18:21)], as.factor)


#missing data
```
```

```
Enabler_Insp_Scale <- Enabler_Insp_Scale %>%
mutate(ProcMgmt = ifelse(is.na(ProcMgmt), 1, ProcMgmt)) %>%
mutate(Last_Insp_outcome = ifelse(Last_Insp_outcome == 1 , 0, Last_Insp_outcome)) %>%
mutate(Last_Insp_outcome = ifelse(Last_Insp_outcome == 2 , 1, Last_Insp_outcome)) %>%
mutate(Next_Insp_outcome = ifelse(Next_Insp_outcome == 1 , 0, Next_Insp_outcome)) %>%
mutate(Next_Insp_outcome = ifelse(Next_Insp_outcome == 2 , 1, Next_Insp_outcome))
#summary(Enabler_Insp_Scale)
Enabler_Insp_Lab <- Enabler_Insp_Scale[!is.na(Enabler_Insp_Scale$Last_Insp_Lab),]
#str(Enabler_Insp_Scale)
```

3.1b. Missing data visualization

```{r}
# Missing value visualization
MD_IOOS <- Enabler_Clean[ , c("iOOS_Proportion","iOOS_RATE", "iOOS_Proportion_DP", "iOOS_RATE_DP")]
missing.values <- MD_IOOS %>%
  gather(key = "key", value = "val") %>%
  mutate(isna = is.na(val)) %>%
  group_by(key) %>%
  mutate(total = n()) %>%
  group_by(key, total, isna) %>%
  summarise(num.isna = n()) %>%
  mutate(pct = num.isna / total * 100)
levels <- (missing.values  %>% filter(isna == T) %>%
```

```
        arrange(desc(pct)))$key
percentage.plot <- missing.values %>%
    ggplot() +
      geom_bar(aes(x = reorder(key, desc(pct)),
                   y = pct, fill=isna),
               stat = 'identity', alpha=0.8) +
      scale_x_discrete(limits = levels) +
      scale_fill_manual(name = "",
                        values = c('steelblue', 'tomato3'),
                        labels = c("Present", "Missing")) +
      coord_flip() +
      labs(title = "Percentage of missing values",
           x = 'Variable', y = "% of missing values")
percentage.plot


MD_Performance <- Enabler_Clean[ , c("InvestPerTest","InvestPerTest_DP",
"ATS", "ATS_DP",
                                     "ATL","ATL_DP", "CycleTime", "CycleTime_DP",
                                     "RFT","RFT_DP", "Tests/FTE", "DP_Tests/FTE")]
missing.values <- MD_Performance %>%
  gather(key = "key", value = "val") %>%
  mutate(isna = is.na(val)) %>%
  group_by(key) %>%
  mutate(total = n()) %>%
  group_by(key, total, isna) %>%
  summarise(num.isna = n()) %>%
  mutate(pct = num.isna / total * 100)
levels <- (missing.values  %>% filter(isna == T) %>%
```

```
        arrange(desc(pct)))$key
percentage.plot <- missing.values %>%
    ggplot() +
      geom_bar(aes(x = reorder(key, desc(pct)),
                   y = pct, fill=isna),
               stat = 'identity', alpha=0.8) +
      scale_x_discrete(limits = levels) +
      scale_fill_manual(name = "",
                        values = c('steelblue', 'tomato3'),
                        labels = c("Present", "Missing")) +
      coord_flip() +
      labs(title = "Percentage of missing values",
           x = 'Variable', y = "% of missing values")
percentage.plot
summary(MD_Performance)
```

3.2. Multi-colinearity

```{r}
#Enabler Categories
corrCat <- round(cor(Enabler_Insp_Scale[,c(2,7,13)]), 1)
head(corrCat)
ggcorrplot(corrCat, hc.order = TRUE, type = "lower", lab = TRUE)
chart.Correlation(Enabler_Insp_Scale[,c(2,7,13)], histogram=TRUE, pch=19)
#Enabler Item
corr <- round(cor(Enabler_Insp_Scale[,c(3:6,8:12,14:16)]), 1)
head(corr)
ggcorrplot(corr, hc.order = TRUE, type = "lower", lab = TRUE)
```

```
chart.Correlation(Enabler_Insp_Scale[,c(3:6,8:12,14:16)], histogram=TRUE, pch=19)
```

3.3. Missing data - Performance

```{r}
All_Insp <- Enabler_Insp_Scale[!is.na(Enabler_Insp_Scale$OS_SCORE),]
All_Insp[, c(2:3)] <- sapply(All_Insp[, c(1:11)], as.numeric)
str(All_Insp)
summary(All_Insp)
#str(Enabler_Insp_Scale)
Enabler_Insp_Scale <- Enabler_Insp %>% mutate_at(c(#"FDA_Insp_Outc",
  "Year", "Overall", "TPM", "TQM", "JIT", "EMS", "BE", "OS_SCORE"), ~(scale(.) %>% as.vector))
OS_Clean <- Enabler[!is.na(Enabler$OS_SCORE),]
OS_Insp <- OS_Clean[!is.na(OS_Clean$FDA_Insp_Bin),]
OS_Insp[, c(1:11)] <- sapply(OS_Insp[, c(1:11)], as.numeric)
#OS_Insp[, c(2:3)] <- sapply(OS_Insp[, c(2:3)], as.factor)
#Enabler_Insp <- Enabler_Insp %>%
#  mutate(
#    FDA_Insp_Outc = FDA_Insp_Outc - 1)
```

3.4. Multi-colinearity - Performance

```{r}
corr <- round(cor(All_Insp[,c(5:9, 11)]), 1)
head(corr)
ggcorrplot(corr, hc.order = TRUE, type = "lower", lab = TRUE)
chart.Correlation(All_Insp[,c(5:9, 11)], histogram=TRUE, pch=19)
#iOOS variants
```

```
corrIOOS <- round(cor(Enabler_Insp_Scale[,c(32:35)]), 1)
head(corrIOOS)
ggcorrplot(corrIOOS, hc.order = TRUE, type = "lower", lab = TRUE)
chart.Correlation(Enabler_Insp_Scale[,c(32:35)], histogram=TRUE, pch=19)
#Performance Metrics
corrPerf <- round(cor(Enabler_Insp_Scale[,c(3:6,8:12,14:16)]), 1)
head(corrPerf)
ggcorrplot(corrPerf, hc.order = TRUE, type = "lower", lab = TRUE)
chart.Correlation(Enabler_Insp_Scale[,c(3:6,8:12,14:16)], histogram=TRUE, pch=19)
#Significant metrics from LinRegs
corrSel <- round(cor(Enabler_Insp_Scale[,c(19,22,24,37)]), 1)
head(corrSel)
ggcorrplot(corrPerf, hc.order = TRUE, type = "lower", lab = TRUE)
chart.Correlation(Enabler_Insp_Scale[,c(19,22,24,37)], histogram=TRUE, pch=19)
```

3.5. Lab inspection citations

```{r}
# Histogram with density plot
ggplot(Enabler_Insp_Lab, aes(x=Last_Insp_Lab)) +
 geom_histogram(binwidth=.5, position="identity", color="black", fill="darkgreen")+
 labs(y= "Count", x = "Weighted citations for QC Lab inspections")+
 geom_vline(aes(xintercept=mean(Last_Insp_Lab)), color="blue", linetype="dashed", size=1)
```

3.6. Distribution Lab citations

```{r}
shapiro.test(Enabler_Insp_Lab$Last_Insp_Lab)
ggqqplot(Enabler_Insp_Lab$Last_Insp_Lab)
```

 # 4. GEE regression

```{r}
gee_mod <- gee(formula = FDA_Insp_Bin ~ #Enabler_Overall +
    Enabler_TPM + Enabler_TQM + Enabler_JIT + Enabler_EMS + Enabler_BE
   , id = Year, data = Enabler_Insp_Scale, na.action = na.omit, family = gaussian)
summary(gee_mod)
```

 # 5. MLR

5.1. Enabler

```{r}
# Linear Regression Lab citations
MLROverall <- lm(Last_Insp_Lab ~ Overall
   , data = Enabler_Insp_Lab)
summary(MLROverall)

MLROverallContr <- lm(Last_Insp_Lab ~ Overall + LC + Size + SizeQC
   , data = Enabler_Insp_Lab)
summary(MLROverallContr)

MLRCat <- lm(Last_Insp_Lab ~ MaintQualSys + PlanSteer + MgmtSys
```

```
    , data = Enabler_Insp_Lab)
summary(MLRCat)

MLRMaintQual <- lm(Last_Insp_Lab ~ MaintQualSys
    , data = Enabler_Insp_Lab)
summary(MLRMaintQual)

MLRMaintQualContr <- lm(Last_Insp_Lab ~ SizeQC#MaintQualSys + LC + Size
+ SizeQC
    , data = Enabler_Insp_Lab)
summary(MLRMaintQualContr)
```

5.2. Performance

```{r}
# Various iOOS rates Linear regression with Inspection outcome
#"iOOS_Proportion","iOOS_RATE", "iOOS_Proportion_DP",
"iOOS_RATE_DP", "iOOS_Proportion_Stab", "iOOS_RATE_Stab"
LRiOOS <- lm(Last_Insp_Lab ~ iOOS_Proportion
    , data = Enabler_Insp_Lab)
summary(LRiOOS)

# Various Performance KPIs with Inspection outcome
#"InvestPerTest","InvestPerTest_DP", "InvestPerTest_Stab", "ATS",
"ATS_DP", "ATS_Stab","ATL","ATL_DP", "ATL_Stab", "CycleTime",
"CycleTime_DP", "CycleTime_Stab", "RFT","RFT_DP", "RFT_Stab",
"TestsFTE", "DP_TestsFTE", "Stab_TestsFTE", "DevPerTest", "RecDev",
"CAPAsOver", "ReTestsComp", "ProdSco", "QualSco", "ServSco",
        "EffectSco", "CostSco", "LabScore"

LRPerf <- lm(Last_Insp_Lab ~ LabScore
```

```
    , data = Enabler_Insp_Lab)
summary(LRPerf)
```

 # 6. LGR

6.1. Enabler

```{r}
# Overall Enablers
LGROverall <- glm(formula = Last_Insp_outcome ~ Overall
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGROverall)
confint.default(LGROverall)
1-pchisq(79.019-78.465,57-56)
wald.test(b = coef(LGROverall), Sigma = vcov(LGROverall), Terms = 1)

LGROverallContr <- glm(formula = Last_Insp_outcome ~ Overall + LC + Size + SizeQC
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGROverallContr)
1-pchisq(76.082-65.892,54-50)

# Category Enablers
LGRCat <- glm(formula = Last_Insp_outcome ~ #Enabler_Overall +
      MaintQualSys + PlanSteer + MgmtSys
    , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRCat)
confint.default(LGRCat)
1-pchisq(79.019-76.906,57-54)
```

```
wald.test(b = coef(LGRCat), Sigma = vcov(LGRCat), Terms = 1:3)

LGRCatContr <- glm(formula = Last_Insp_outcome ~ #Enabler_Overall +
      MaintQualSys + PlanSteer + MgmtSys
      + LC + Size + SizeQC
   , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRCatContr)
1-pchisq(76.082-64.636,54-48)

# Sub-category Enablers
LGRSub <- glm(formula = Last_Insp_outcome ~ #Enabler_Overall +
      PrevMaint + Tech + FiveS + ProcMgmt + Setup + Pull + Layout + PlanAdh
+ VisMgmt + LeadCult + CI + FuncIntQual
   , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)
summary(LGRSub)
confint.default(LGRSub)
1-pchisq(77.561-68.198,55-43)
wald.test(b = coef(LGRSub), Sigma = vcov(LGRSub), Terms = 1:12)

#LGR <- train(Last_Insp_outcome ~ #Enabler_Overall +
#          Enabler_TPM + Enabler_TQM + Enabler_JIT + Enabler_EMS +
Enabler_BE
#          , data = Enabler_Insp_Scale, method="glm", trControl = ctrl,
family="binomial")
```

6.2. Performance

```{r}
# Various iOOS rates Logistic regression with Inspection outcome
```

```
#"iOOS_Proportion","iOOS_RATE", "iOOS_Proportion_DP",
"iOOS_RATE_DP", "iOOS_Proportion_Stab", "iOOS_RATE_Stab"

LGRPerf <- glm(formula = Last_Insp_outcome ~ iOOS_RATE_Stab

   , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)

summary(LGRPerf)

confint.default(LGRPerf)

1-pchisq(73.304-73.053,52-51)

wald.test(b = coef(LGRPerf), Sigma = vcov(LGRPerf), Terms = 1)


# Various Performance KPIs with Inspection outcome

#"InvestPerTest","InvestPerTest_DP", "InvestPerTest_Stab", "ATS",
"ATS_DP", "ATS_Stab","ATL","ATL_DP", "ATL_Stab", "CycleTime",
"CycleTime_DP", "CycleTime_Stab", "RFT","RFT_DP", "RFT_Stab",
"TestsFTE", "DP_TestsFTE", "Stab_TestsFTE", "DevPerTest", "RecDev",
"CAPAsOver", "ReTestsComp", "ProdSco", "QualSco", "ServSco",
        "EffectSco", "CostSco", "LabScore"

LGRPerf2 <- glm(formula = Last_Insp_outcome ~ LabScore

   , family = "binomial", data = Enabler_Insp_Scale, na.action = na.omit)

summary(LGRPerf2)

confint.default(LGRPerf2)

1-pchisq(73.304-73.053,52-51)

wald.test(b = coef(LGRPerf2), Sigma = vcov(LGRPerf2), Terms = 1)
```

A.4 R Code: Analyses for Chapter 4

The following code is provided as a R notebook with commentary and an interactive table of contents

---

---

title: "Company case study data preparation and ML"

output: html_notebook

 # 1. Setup

Install necessary packages:

```{r}
#Only run this once unless there are problems with the file
rm(list=ls())
remove.packages(c("ggplot2", "data.table","tidyverse", "Rcpp", "dplyr"))
install.packages('Rcpp', dependencies = TRUE)
install.packages('broom', dependencies = TRUE, type="binary")
install.packages('ggplot2', dependencies = TRUE)
install.packages('data.table', dependencies = TRUE)
install.packages("tidyverse", dependencies = TRUE)
install.packages("dplyr", dependencies = TRUE)
#install.packages("scorecard")
```

Load packages

```{r}
library(Rcpp)
# load packages
#install.packages(readxl)
```

```
library(readxl)
#install.packages("xlsx")
library(xlsx)
#install.packages("tidyverse")
library(tidyverse)
#install.packages("reshape2")
library(reshape2)
#install.packages("zoo")
library(zoo)
library(ggplot2)
#install.packages("ggcorrplot")
library(ggcorrplot)
library(tibble)
library(dplyr)
library(rpart)
library(rpart.plot)
library(caret)
library(rattle)
library(magrittr)
library(pROC)
#install.packages("gee")
library(gee)
#install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)
library(scorecard)
```

 # 2. Data preparation

Next, working directory and input filename are defined.

```{r}
rm(list=ls())
setwd("C:/R_Daten/ Company case study ")
INPUT_FILE <- "C:/R_Daten/ Company case study /Data/ Company case study R input.xlsx"
```

Data import

```{r}
quality <- read_xlsx(INPUT_FILE, skip = 1)
```

Create panel in long format by separating values based on the timecode, transform timecode and delete empty rows

```{r}
quality_long <- gather(quality,
                       key = "year_month",
                       value = "value",
                       `201301`:`201905`)
quality_long$Data_availability <- NULL
quality_long$Indicator_group <- NULL
# quality_long2 <- distinct(quality_long) #check for dublicates
quality_long$year_month <- as.yearmon(quality_long$year_month,
                                      format = "%Y%m")
quality_long2 <- na.omit(quality_long)
```

Create a wide table and selecting subsets with inspection and/or audit that month

```{r}
quality_wide <- spread(quality_long2, key = "Indicators", value = "value")

cols = c(5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76) #, 77, 78, 79, 80, 81, 82, 83, 84, 85


quality_wide[, c(5:78)] <- sapply(quality_wide[, c(5:78)], as.numeric)

#summary(quality_wide)

#summary(quality_wide$Inspection_Score_Normalized)

quality_wide$Current_Inspection <- ifelse(is.na(quality_wide$Inspection_Score_Normalized),quality_wide$EURO_Inspection_Score_Normalized,quality_wide$Inspection_Score_Normalized)

quality_wide_Insp <- quality_wide[!is.na(quality_wide$Current_Inspection),]

quality_wide_Insp <- quality_wide_Insp[!1<(quality_wide_Insp$Current_Inspection),]

#summary(quality_wide_Insp)

quality_wide_Insp$Inspection_Score_Normalized <- NULL

quality_wide_Insp$EURO_Inspection_Score_Normalized <- NULL

write.xlsx(quality_wide_Insp, "C:/R_Daten/Sanofi/Data/quality_data_wide_Insp.xlsx", sheetName = "Sheet1",
        col.names = TRUE, row.names = TRUE, append = FALSE)


quality_wide_Audit <- quality_wide[!is.na(quality_wide$Audit_Score_Normalized),]

#summary(quality_wide_Audit)
```

```
quality_wide_current <- quality_wide[quality_wide$year_month == "Apr 2019",]


write.xlsx(quality_wide_Audit,
"C:/R_Daten/Sanofi/Data/quality_data_wide_Audit.xlsx", sheetName = "Sheet1",
          col.names = TRUE, row.names = TRUE, append = FALSE)
```

 # 3. Missing data

Build ML models with identical test and training datasets

First, we code the missing variables

```{r}
quality_wide_ML <- quality_wide_Insp %>%
  mutate(
    Time_Since_Insp_missing = ifelse(is.na(Time_Since_Last_Insp), 1, 0),
    Time_Since_Last_Insp = ifelse(is.na(Time_Since_Last_Insp), 50, Time_Since_Last_Insp))
quality_wide_ML <- quality_wide_ML %>%
    mutate(
    Prev_Insp_missing = ifelse(is.na(Previous_Inspection_Score_Normalized), 1, 0),
    Previous_Inspection_Score_Normalized = ifelse(is.na(Previous_Inspection_Score_Normalized), 0, Previous_Inspection_Score_Normalized))
quality_wide_ML <- quality_wide_ML %>%
  mutate(
    Time_Since_Aud_missing = ifelse(is.na(Time_Since_Last_Audit), 1, 0),
    Time_Since_Last_Audit = ifelse(is.na(Time_Since_Last_Audit), 25, Time_Since_Last_Audit))
quality_wide_ML <- quality_wide_ML %>%
```

```
  mutate(
    Prev_Aud_missing = ifelse(is.na(Previous_Audit_Score_Normalized), 1, 0),
    Previous_Audit_Score_Normalized = ifelse(is.na(Previous_Audit_Score_Normalized), 0, Previous_Audit_Score_Normalized))
quality_wide_ML <- quality_wide_ML %>%
  mutate(
    Rejected_batches_missing = ifelse(is.na(Rejected_batches), 1, 0),
    Rejected_batches = ifelse(is.na(Rejected_batches), mean(Rejected_batches, na.rm = TRUE), Rejected_batches))
quality_wide_ML <- quality_wide_ML %>%
  mutate(
    Right_first_time_missing = ifelse(is.na(Right_first_time), 1, 0),
    Right_first_time = ifelse(is.na(Right_first_time), mean(Right_first_time, na.rm = TRUE), Right_first_time))
quality_wide_ML <- quality_wide_ML %>%
  mutate(
    recalculated_Complaints_product_missing = ifelse(is.na(recalculated_Complaints_product), 1, 0),
    recalculated_Complaints_product = ifelse(is.na(recalculated_Complaints_product),mean(recalculated_Complaints_product, na.rm = TRUE), Right_first_time))
quality_wide_ML <- quality_wide_ML %>%
  mutate(
    Six_Month_Avg._Rejected_batches_missing = ifelse(is.na(Six_Month_Avg._Rejected_batches), 1, 0),
    Six_Month_Avg._Rejected_batches = ifelse(is.na(Six_Month_Avg._Rejected_batches), mean(Six_Month_Avg._Rejected_batches, na.rm = TRUE), Six_Month_Avg._Rejected_batches))
quality_wide_ML <- quality_wide_ML %>%
```

```
  mutate(
    Six_Month_Avg._Right_first_time_missing = ifelse(is.na(Six_Month_Avg._Right_first_time), 1, 0),
    Six_Month_Avg._Right_first_time = ifelse(is.na(Six_Month_Avg._Right_first_time), mean(Six_Month_Avg._Right_first_time, na.rm = TRUE), Six_Month_Avg._Right_first_time))

summary(quality_wide_ML)
quality_wide_ML_op <- quality_wide_ML[!(quality_wide_ML$Rejected_batches_missing==1),]
summary(quality_wide_ML_op$Current_Inspection)
sd(quality_wide_ML_op$Current_Inspection, na.rm = TRUE)
p <- ggplot(quality_wide_ML_op, aes(x=Current_Inspection)) +
  geom_histogram(color="black", fill="white", binwidth = 0.1)
p
```

 # 4. ML models - Replication

4.1. Logistic Regression model 1 using only last inspection and time since

This chapter applies models to identical test and training datasets, therefore just attempting to replicate the original target values.

```{r}
lr_mod <- glm(Current_Inspection ~ Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
              # + Prev_Insp_missing + Time_Since_Insp_missing
              , data = quality_wide_ML, family = binomial)
```

```
quality_wide_ML$Insp_LogPred <- predict(lr_mod, quality_wide_ML, type = "response")

### Evaluate performance

ggplot(data = quality_wide_ML, aes(x = Insp_LogPred, y = Current_Inspection)) +
  geom_point(color='blue') +
  geom_smooth(method = "lm", se = FALSE)

correlations <- cor(quality_wide_ML[c("Current_Inspection", "Insp_LogPred", "Previous_Inspection_Score_Normalized",
                                      "Previous_Audit_Score_Normalized")],
                    use = "everything")

ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())
```

4.2. Logistic Regression model 2 using only last inspection, audit and time since

```{r}
lr_mod2 <- glm(Current_Inspection ~ Previous_Inspection_Score_Normalized + Time_Since_Last_Insp +
                 #Prev_Insp_missing + Time_Since_Insp_missing +
                 Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                 #+ Time_Since_Aud_missing + Prev_Aud_missing
                 ,data = quality_wide_ML, family = binomial)

quality_wide_ML$Insp_LogPred2 <- predict(lr_mod2, quality_wide_ML, type = "response")
```

Evaluate performance

```
dat_plot <- melt(quality_wide_ML %>% select(Current_Inspection, Insp_LogPred, Insp_LogPred2), id.vars = "Current_Inspection")


ggplot(quality_wide_ML, aes(Current_Inspection, y = Prediction, color = variable)) +
  geom_point(aes(y = Insp_LogPred, col = "Insp_LogPred")) +
  geom_point(aes(y = Insp_LogPred2, col = "Insp_LogPred2")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred, col = "Insp_LogPred")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred2, col = "Insp_LogPred2"))


correlations <- cor(quality_wide_ML[c("Current_Inspection", "Insp_LogPred", "Insp_LogPred2", "Previous_Inspection_Score_Normalized",
                                      "Previous_Audit_Score_Normalized")],
                    use = "everything")


ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())
```

4.3. Logistic Regression model 3 using last inspection and time since; RFT and Rejected batches

```{r}
```

```
lr_mod3 <- glm(Current_Inspection ~ Previous_Inspection_Score_Normalized
+ Time_Since_Last_Insp +
            #Prev_Insp_missing + Time_Since_Insp_missing +
            Previous_Audit_Score_Normalized + Time_Since_Last_Audit +
            #Time_Since_Aud_missing + Prev_Aud_missing +
            Rejected_batches + Right_first_time
            #+ Rejected_batches_missing + Right_first_time_missing
            ,data = quality_wide_ML, family = binomial)
quality_wide_ML$Insp_LogPred3 <- predict(lr_mod3, quality_wide_ML, type =
"response")

### Evaluate performance
dat_plot <- melt(quality_wide_ML %>% select(Current_Inspection,
Insp_LogPred2, Insp_LogPred3), id.vars = "Current_Inspection")

ggplot(quality_wide_ML, aes(Current_Inspection, y = Prediction, color =
variable)) +
  geom_point(aes(y = Insp_LogPred2, col = "Insp_LogPred2")) +
  geom_point(aes(y = Insp_LogPred3, col = "Insp_LogPred3")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred2,
col = "Insp_LogPred2")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred3,
col = "Insp_LogPred3"))

correlations <- cor(quality_wide_ML[c("Current_Inspection", "Insp_LogPred",
"Insp_LogPred2", "Insp_LogPred3", "Previous_Inspection_Score_Normalized",
"Previous_Audit_Score_Normalized")], use = "everything")

ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
```

```
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())
```

4.4. Logistic Regression model 4 using last inspection and time since; RFT and Rejected batches plus missing variables

```{r}
lr_mod4 <- glm(Current_Inspection ~ Type_Detail + Previous_Inspection_Score_Normalized + Time_Since_Last_Insp +
                 Prev_Insp_missing + Time_Since_Insp_missing +
                 Previous_Audit_Score_Normalized + Time_Since_Last_Audit +
                 Time_Since_Aud_missing + Prev_Aud_missing +
                 Rejected_batches + Right_first_time +
                 Rejected_batches_missing + Right_first_time_missing
               ,data = quality_wide_ML, family = binomial)
quality_wide_ML$Insp_LogPred4 <- predict(lr_mod4, quality_wide_ML, type = "response")
```

Evaluate performance

```
dat_plot <- melt(quality_wide_ML %>% select(Current_Inspection, Insp_LogPred2, Insp_LogPred3, Insp_LogPred4), id.vars = "Current_Inspection")

ggplot(quality_wide_ML, aes(Current_Inspection, y = Prediction, color = variable)) +
  geom_point(aes(y = Insp_LogPred3, col = "Insp_LogPred3")) +
  geom_point(aes(y = Insp_LogPred4, col = "Insp_LogPred4")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred3, col = "Insp_LogPred3")) +
```

```
 geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred4, col = "Insp_LogPred4"))


correlations <- cor(quality_wide_ML[c("Current_Inspection", "Insp_LogPred", "Insp_LogPred2", "Insp_LogPred3", "Insp_LogPred4", "Previous_Inspection_Score_Normalized",
                                      "Previous_Audit_Score_Normalized")],
                    use = "everything")


ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())
```

4.5. Decision tree model 1 using last inspection and time since; RFT and Rejected batches

```{r}
RegDT = rpart(Current_Inspection ~ Type_Detail + Previous_Inspection_Score_Normalized + Time_Since_Last_Insp +
                #Prev_Insp_missing + Time_Since_Insp_missing +
                Previous_Audit_Score_Normalized + Time_Since_Last_Audit +
                #Time_Since_Aud_missing + Prev_Aud_missing +
                Rejected_batches + Right_first_time
              #+ Rejected_batches_missing + Right_first_time_missing
              ,data = quality_wide_ML)
RegDT
fancyRpartPlot(RegDT,cex=0.65)
```

```
quality_wide_ML$Insp_RDTPred <- predict(RegDT, quality_wide_ML)

ggplot(quality_wide_ML, aes(Current_Inspection, y = Prediction, color = variable)) +

  geom_point(aes(y = Insp_RDTPred, col = "Insp_RDTPred")) +

  geom_point(aes(y = Insp_LogPred4, col = "Insp_LogPred4")) +

  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_RDTPred, col = "Insp_RDTPred")) +

  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred4, col = "Insp_LogPred4"))

correlations <- cor(quality_wide_ML[c("Current_Inspection", "Insp_LogPred", "Insp_LogPred2", "Insp_LogPred3", "Insp_LogPred4", "Insp_RDTPred", "Previous_Inspection_Score_Normalized",

                                      "Previous_Audit_Score_Normalized")],

                    use = "everything")


ggplot(melt(correlations),

       aes(Var1, Var2, fill = value, label = round(value, 2))) +

  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +

  geom_tile() + geom_text() +

  theme(axis.title.x = element_blank(), axis.title.y = element_blank())


MAE.RegDT <- mean(abs(quality_wide_ML$Current_Inspection-quality_wide_ML$Insp_RDTPred))

MAE.RegDT
```

4.6. Decision tree model 2 using last inspection and time since; RFT and Rejected batches and missing variables

```{r}
```

```
ctrl <- trainControl(method = "repeatedcv", number = 3, repeats = 3, #no #5-fold CV
                     selectionFunction = "best" # select the best performer
)
RegDT2 <- train(Current_Inspection ~ Type_Detail + Previous_Inspection_Score_Normalized + Time_Since_Last_Insp +
                  #Prev_Insp_missing + Time_Since_Insp_missing +
                  Previous_Audit_Score_Normalized + Time_Since_Last_Audit +
                  #Time_Since_Aud_missing + Prev_Aud_missing +
                  Rejected_batches + Right_first_time
                #+ Rejected_batches_missing + Right_first_time_missing
                ,data = quality_wide_ML,
                method = "rpart",
                trControl = ctrl)
RegDT2
fancyRpartPlot(RegDT2$finalModel)
quality_wide_ML$Insp_RDTPred2 <- predict(RegDT2, quality_wide_ML)
rpart.plot(RegDT2$finalModel, box.palette="RdBu", shadow.col="gray", nn=TRUE)
quality_wide_ML$Insp_RDTPred2 <- predict(RegDT2, quality_wide_ML, type = "raw")
ggplot(quality_wide_ML, aes(Current_Inspection, y = Prediction, color = variable)) +
  geom_point(aes(y = Insp_RDTPred, col = "Insp_RDTPred")) +
  geom_point(aes(y = Insp_RDTPred2, col = "Insp_RDTPred2")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_RDTPred, col = "Insp_RDTPred")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_RDTPred2, col = "Insp_RDTPred2"))
```

```
correlations <- cor(quality_wide_ML[c("Current_Inspection", "Insp_LogPred", "Insp_LogPred2", "Insp_LogPred3", "Insp_LogPred4","Insp_RDTPred", "Insp_RDTPred2", "Previous_Inspection_Score_Normalized",
                                      "Previous_Audit_Score_Normalized")],
                    use = "everything")


ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())
MAE.RegDT2 <- mean(abs(quality_wide_ML$Current_Inspection-quality_wide_ML$Insp_RDTPred2))
MAE.RegDT2
```

A.4.1 Prediction models with numerical target score

 # 5. ML models - Inspection Score - 10fold CV

This chapter aims tp predict inspection outcomes in a 10-fold cross-validation mode.

First, we need to randomly split the data into training and test datasets. We choose a size ratio of 2/3 to 1/3

```{r}
smp_size <- floor(0.67 * nrow(quality_wide_ML))


## set the seed to make your partition reproducible
set.seed(123)
train_ind <- sample(seq_len(nrow(quality_wide_ML)), size = smp_size)
```

```
quality_wide_ML_train <- quality_wide_ML[train_ind, ]
quality_wide_ML_test <- quality_wide_ML[-train_ind, ]
boxplot(quality_wide_ML_op$Current_Inspection)
```

5.1. Linear Regression model 1 using only last inspection and time since

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 10 or 3-fold CV
  savePredictions = TRUE # needed to plot the predictions
)
lr_mod1 <- train(Current_Inspection ~ Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
                 #+ Prev_Insp_missing + Time_Since_Insp_missing
                 #+ Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                 #+ Time_Since_Aud_missing + Prev_Aud_missing
                , data = quality_wide_ML_op, method="lm", trControl = ctrl, family = binomial, na.action = na.pass)
lr_mod1
summary(lr_mod1)

quality_wide_ML_op$Insp_LRPred <- predict(lr_mod1, quality_wide_ML_op, type = "raw")
### Evaluate performance
ggplot(data = quality_wide_ML_op, aes(x = Insp_LRPred, y = Current_Inspection)) +
  geom_point(color='blue') +
  geom_smooth(method = "lm", se = FALSE)
```

```
correlations <- cor(quality_wide_ML_op[c("Current_Inspection",
"Insp_LRPred", "Previous_Inspection_Score_Normalized",
                                       "Previous_Audit_Score_Normalized")],
                    use = "everything")
ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())
```

5.2. Linear Regression model 2 using only last inspection, audit and time since

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 10 or 3-fold CV
  savePredictions = TRUE # needed to plot the predictions
)
lr_mod2 <- train(Current_Inspection ~ Previous_Inspection_Score_Normalized
+ Time_Since_Last_Insp
                 #+ Prev_Insp_missing + Time_Since_Insp_missing
                 + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                 #+ Time_Since_Aud_missing + Prev_Aud_missing
                , data = quality_wide_ML_op, method="lm", trControl = ctrl, family =
binomial, na.action = na.pass)
lr_mod2
summary(lr_mod2)
```

```
quality_wide_ML_op$Insp_LRPred <- predict(lr_mod2, quality_wide_ML_op, type = "raw")

### Evaluate performance

ggplot(data = quality_wide_ML_op, aes(x = Insp_LRPred, y = Current_Inspection)) +
  geom_point(color='blue') +
  geom_smooth(method = "lm", se = FALSE)

correlations <- cor(quality_wide_ML_op[c("Current_Inspection", "Insp_LRPred", "Previous_Inspection_Score_Normalized",
                                         "Previous_Audit_Score_Normalized")],
                    use = "everything")

ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())
```

5.3. Linear Regression model 3 using last inspection and time since; RFT and Rejected batches

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 10 or 3-fold CV
  savePredictions = TRUE # needed to plot the predictions
)

lr_mod4 <- train(Current_Inspection ~ Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
                 + Prev_Insp_missing + Time_Since_Insp_missing
                 + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
```

```
        + Time_Since_Aud_missing + Prev_Aud_missing
        + Rejected_batches + Right_first_time
        + Rejected_batches_missing + Right_first_time_missing
        , data = quality_wide_ML_op, method="lm", trControl = ctrl, family = binomial, na.action = na.pass)
lr_mod4
summary(lr_mod4)


quality_wide_ML_op$Insp_LRPred <- predict(lr_mod4, quality_wide_ML_op, type = "raw")
### Evaluate performance
ggplot(data = quality_wide_ML_op, aes(x = Insp_LRPred, y = Current_Inspection)) +
  geom_point(color='blue') +
  geom_smooth(method = "lm", se = FALSE)
correlations <- cor(quality_wide_ML_op[c("Current_Inspection", "Insp_LRPred", "Previous_Inspection_Score_Normalized",
                                         "Previous_Audit_Score_Normalized")],
                    use = "everything")
ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())


### Evaluate performance
dat_plot <- melt(quality_wide_ML_test %>% select(Current_Inspection, Insp_LogPred2, Insp_LogPred3), id.vars = "Current_Inspection")
```

```
ggplot(quality_wide_ML_test, aes(Current_Inspection, y = Prediction, color = variable)) +
  geom_point(aes(y = Insp_LogPred2, col = "Insp_LogPred2")) +
  geom_point(aes(y = Insp_LogPred3, col = "Insp_LogPred3")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred2, col = "Insp_LogPred2")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred3, col = "Insp_LogPred3"))


correlations <- cor(quality_wide_ML_test[c("Current_Inspection", "Insp_LogPred", "Insp_LogPred2", "Insp_LogPred3", "Previous_Inspection_Score_Normalized",
                                           "Previous_Audit_Score_Normalized")],
                    use = "everything")


ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())
```

5.4. Linear Regression model 4 using last inspection and time since; RFT and Rejected batches plus missing variables

```{r}
lr_mod4 <- glm(Current_Inspection ~ Type_Detail + Previous_Inspection_Score_Normalized + Time_Since_Last_Insp +
                 Prev_Insp_missing + Time_Since_Insp_missing +
                 Previous_Audit_Score_Normalized + Time_Since_Last_Audit +
                 Time_Since_Aud_missing + Prev_Aud_missing +
```

```
            Rejected_batches + Right_first_time +
            Rejected_batches_missing + Right_first_time_missing
          ,data = quality_wide_ML_train, family = binomial)
quality_wide_ML_test$Insp_LogPred4 <- predict(lr_mod4, quality_wide_ML_test, type = "response")

### Evaluate performance
dat_plot <- melt(quality_wide_ML_test %>% select(Current_Inspection, Insp_LogPred2, Insp_LogPred3, Insp_LogPred4), id.vars = "Current_Inspection")

ggplot(quality_wide_ML_test, aes(Current_Inspection, y = Prediction, color = variable)) +
  geom_point(aes(y = Insp_LogPred3, col = "Insp_LogPred3")) +
  geom_point(aes(y = Insp_LogPred4, col = "Insp_LogPred4")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred3, col = "Insp_LogPred3")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_LogPred4, col = "Insp_LogPred4"))

correlations <- cor(quality_wide_ML_test[c("Current_Inspection", "Insp_LogPred", "Insp_LogPred2", "Insp_LogPred3", "Insp_LogPred4", "Previous_Inspection_Score_Normalized",
                                           "Previous_Audit_Score_Normalized")],
                    use = "everything")

ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
```

```
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())
```

5.5. Decision tree model 1 using last inspection and time since; RFT and Rejected batches

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 10 or 3-fold CV
  savePredictions = TRUE # needed to plot the predictions
)
DT_mod <- train(Current_Inspection ~ Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
            #+ Prev_Insp_missing + Time_Since_Insp_missing
            #+ Previous_Audit_Score_Normalized + Time_Since_Last_Audit
            #+ Time_Since_Aud_missing + Prev_Aud_missing
            #+ Rejected_batches + Right_first_time
            #+ Rejected_batches_missing + Right_first_time_missing
            , data = quality_wide_ML_op, method="rpart", trControl = ctrl, tuneLength = 4, na.action = na.pass)
DT_mod
summary(DT_mod)
fancyRpartPlot(DT_mod$finalModel,cex=0.65)
```

5.6. Decision tree model 2 using last inspection and time since; RFT and Rejected batches and missing variables

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 10 or 3-fold CV
```

```
  savePredictions = TRUE # needed to plot the predictions
)
DT_mod2 <- train(Current_Inspection ~ Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
        #+ Prev_Insp_missing + Time_Since_Insp_missing
        + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
        #+ Time_Since_Aud_missing + Prev_Aud_missing
        #+ Rejected_batches + Right_first_time
        #+ Rejected_batches_missing + Right_first_time_missing
        , data = quality_wide_ML_op, method="rpart", trControl = ctrl, tuneLength = 10, na.action = na.pass)
DT_mod2
summary(DT_mod2)
fancyRpartPlot(DT_mod2$finalModel,cex=0.65)
```

5.7. Decision tree model 3 using last inspection and time since; RFT and Rejected batches, complaints and missing variables

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 10 or 3-fold CV
  savePredictions = TRUE # needed to plot the predictions
)
DT_mod3 <- train(Current_Inspection ~ Type_Detail + Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
        + Prev_Insp_missing + Time_Since_Insp_missing
        + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
        + Time_Since_Aud_missing + Prev_Aud_missing
        + Rejected_batches + Right_first_time
```

```
        + Rejected_batches_missing + Right_first_time_missing
        , data = quality_wide_ML_op, method="rpart", trControl = ctrl, tuneLength = 10, na.action = na.pass)
DT_mod3
summary(DT_mod3)
fancyRpartPlot(DT_mod3$finalModel,cex=0.65)



#fancyRpartPlot(RegDT3$finalModel)
#rpart.plot(RegDT3$finalModel, box.palette="RdBu", shadow.col="gray", nn=TRUE)
quality_wide_ML_test$Insp_RDTPred3 <- predict(RegDT3, quality_wide_ML_test)
ggplot(quality_wide_ML_test, aes(Current_Inspection, y = Prediction, color = variable)) +
  geom_point(aes(y = Insp_RDTPred3, col = "Insp_RDTPred3")) +
  geom_point(aes(y = Insp_RDTPred2, col = "Insp_RDTPred2")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_RDTPred3, col = "Insp_RDTPred3")) +
  geom_smooth(method= 'lm',aes(x=Current_Inspection, y = Insp_RDTPred2, col = "Insp_RDTPred2"))
correlations <- cor(quality_wide_ML_test[c("Current_Inspection", "Insp_LogPred", "Insp_LogPred2", "Insp_LogPred3", "Insp_LogPred4", "Insp_RDTPred", "Insp_RDTPred2", "Insp_RDTPred3", "Previous_Inspection_Score_Normalized",
                                           "Previous_Audit_Score_Normalized")],
                use = "everything")


ggplot(melt(correlations),
       aes(Var1, Var2, fill = value, label = round(value, 2))) +
```

```
  scale_fill_gradient2(low = "blue", high = "red", limit = c(-1, 1)) +
  geom_tile() + geom_text() +
  theme(axis.title.x = element_blank(), axis.title.y = element_blank())


MAE.RegDT3 <- mean(abs(quality_wide_ML_test$Current_Inspection-quality_wide_ML_test$Insp_RDTPred3))
MAE.RegDT3
```

5.8. Neuralnet using last inspection and time since; RFT and Rejected batches and missing variables

```{r}
tune.grid.neuralnet <- expand.grid(
  layer1 = 5, #10 #5 #3
  layer2 = 5, #10 #5 #3 #1
  layer3 = 1  #10 #1
)
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 10 or 3-fold CV
  savePredictions = TRUE # needed to plot the predictions
)
NN_mod <- train(Current_Inspection ~ Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
            #+ Prev_Insp_missing + Time_Since_Insp_missing
            + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
            #+ Time_Since_Aud_missing + Prev_Aud_missing
            + Rejected_batches + Right_first_time
            #+ Rejected_batches_missing + Right_first_time_missing
    , data = quality_wide_ML_op,    method = "neuralnet", linear.output = TRUE, tuneGrid = tune.grid.neuralnet,
```

```
    metric = "RMSE",
    trControl = ctrl)
NN_mod
summary(NN_mod)
plot(NN_mod$finalModel, rep = NULL, x.entry = NULL, x.out = NULL,
  radius = 0.15, arrow.length = 0.15, intercept = FALSE,
  intercept.factor = 0.4, information = TRUE, information.pos = 0.1,
  col.entry.synapse = "red", col.entry = "black",
  col.hidden = "black", col.hidden.synapse = "black",
  col.out = "green", col.out.synapse = "black",
  col.intercept = "blue", fontsize = 11, dimension = 10,
  show.weights = TRUE, file = NULL)
```

A.4.2 Prediction models with binary target

 # 6. ML models - BINARY - 10fold CV

In this chapter, inspection outcomes are transformed into a binary variable and then predicted in a 10-fold cross-validation mode.

First we create binary targets:

```{r}
quality_wide_bin <- quality_wide_ML %>%
  mutate(
    Current_Inspection = ifelse(Current_Inspection == 1, 1, 0))
quality_wide_bin$Current_Inspection <- factor(quality_wide_bin$Current_Inspection , levels = c(0, 1), labels = c("yes", "no"))
summary(quality_wide_bin$Current_Inspection)
```

```
quality_wide_bin_op <- quality_wide_bin[!(quality_wide_bin$Rejected_batches_missing==1),]
summary(quality_wide_bin_op$Current_Inspection)
df <- data.frame(value = c(65, 28),
                 Group = c("Concluded with findings", "Concluded without Findings"))
pie <- ggplot(df, aes(x="", y=value, fill=Group)) +
       geom_bar(width = 1, stat = "identity") +
       coord_polar("y", start=0)
pie
#smp_size_bin <- floor(0.67 * nrow(quality_wide_bin))
#smp_size_bin_op <- floor(0.67 * nrow(quality_wide_bin_op))
## set the seed to make your partition reproducible
#set.seed(123)
#train_ind <- sample(seq_len(nrow(quality_wide_bin)), size = smp_size_bin)
#quality_wide_bin_train <- quality_wide_bin[train_ind, ]
#quality_wide_bin_test <- quality_wide_bin[-train_ind, ]
#quality_wide_bin__op_train <- quality_wide_bin_op[train_ind, ]
#quality_wide_bin_op_test <- quality_wide_bin_op[-train_ind, ]
```

6.1. Logistic regression models

6.1.1. Only compliance history

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 3-fold CV
  #repeats = 10,
  selectionFunction = "best", # select the best performer
  classProbs = TRUE, # requested the predicted probs (for ROC)
```

```
  summaryFunction = twoClassSummary, # needed to produce the ROC/AUC measures
  savePredictions = TRUE # needed to plot the ROC curves
)
logReg_mod1 <- train(Current_Inspection ~ Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
            + Prev_Insp_missing + Time_Since_Insp_missing
            + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
            + Time_Since_Aud_missing + Prev_Aud_missing
           , data = quality_wide_bin, method="glm", trControl = ctrl, family="binomial")
logReg_mod1
roc.logReg_mod1 <- roc(predictor = logReg_mod1$pred$yes, response = logReg_mod1$pred$obs)
auc.logReg_mod1 <- round(auc(roc.logReg_mod1), 3)
auc.logReg_mod1

# plot the OneR ROC curve and set up the identity line and axes
roc.logReg_mod1_roc <- plot(roc.logReg_mod1,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "Logistic Regression *AI prediction, (10-fold CV)"
)
```

6.1.1b. Only compliance history - only records with KPIs

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 3-fold CV
  #repeats = 10,
```

```
  selectionFunction = "best", # select the best performer
  classProbs = TRUE, # requested the predicted probs (for ROC)
  summaryFunction = twoClassSummary, # needed to produce the ROC/AUC measures
  savePredictions = TRUE # needed to plot the ROC curves
)
logReg_mod1 <- train(Current_Inspection ~ Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
            #+ Prev_Insp_missing + Time_Since_Insp_missing
            + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
            #+ Time_Since_Aud_missing + Prev_Aud_missing
           , data = quality_wide_bin_op, method="glm", trControl = ctrl, family="binomial")
logReg_mod1
roc.logReg_mod1 <- roc(predictor = logReg_mod1$pred$yes, response = logReg_mod1$pred$obs)
round(auc(roc.logReg_mod1), 3)
summary(logReg_mod1)
# plot the OneR ROC curve and set up the identity line and axes
roc.logReg_mod1_roc <- plot(roc.logReg_mod1,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "Logistic Regression *AI prediction, (10-fold CV)"
)
confusionMatrix(logReg_mod1$pred[order(logReg_mod1$pred$rowIndex),2], logReg_mod1$pred$obs)
```

6.1.2. Only compliance history and production type

```{r}
```

```
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 3-fold CV
  #repeats = 10,
  selectionFunction = "best", # select the best performer
  classProbs = TRUE, # requested the predicted probs (for ROC)
  summaryFunction = twoClassSummary, # needed to produce the ROC/AUC measures
  savePredictions = TRUE # needed to plot the ROC curves
)
logReg_mod2 <- train(Current_Inspection ~ Type_Detail + Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
                + Prev_Insp_missing + Time_Since_Insp_missing
                + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                + Time_Since_Aud_missing + Prev_Aud_missing
        , data = quality_wide_bin, method="glm", trControl = ctrl, family="binomial")
logReg_mod2
roc.logReg_mod2 <- roc(predictor = logReg_mod2$pred$yes, response = logReg_mod2$pred$obs)
auc.logReg_mod2 <- round(auc(roc.logReg_mod2), 3)
auc.logReg_mod2

# plot the OneR ROC curve and set up the identity line and axes
roc.logReg_mod2_roc <- plot(roc.logReg_mod2,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "Logistic Regression *AI prediction, (10-fold CV)"
)
```

6.1.2b. Only compliance history and production type - only records with KPIs

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 3-fold CV
  #repeats = 10,
  selectionFunction = "best", # select the best performer
  classProbs = TRUE, # requested the predicted probs (for ROC)
  summaryFunction = twoClassSummary, # needed to produce the ROC/AUC measures
  savePredictions = TRUE # needed to plot the ROC curves
)
logReg_mod2 <- train(Current_Inspection ~ Type_Detail + Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
                     #+ Prev_Insp_missing + Time_Since_Insp_missing
                     + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                     #+ Time_Since_Aud_missing + Prev_Aud_missing
       , data = quality_wide_bin_op, method="glm", trControl = ctrl, family="binomial")
logReg_mod2
roc.logReg_mod2 <- roc(predictor = logReg_mod2$pred$yes, response = logReg_mod2$pred$obs)
round(auc(roc.logReg_mod2), 3)
summary(logReg_mod2)
# plot the OneR ROC curve and set up the identity line and axes
roc.logReg_mod1_roc <- plot(roc.logReg_mod2,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "Logistic Regression *AI prediction, (10-fold CV)"
```

```
)
confusionMatrix(logReg_mod2$pred[order(logReg_mod2$pred$rowIndex),2],
logReg_mod2$pred$obs)
```

6.1.3. Compliance history, production type and Operational KPIs

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 3-fold CV
  #repeats = 10,
  selectionFunction = "best", # select the best performer
  classProbs = TRUE, # requested the predicted probs (for ROC)
  summaryFunction = twoClassSummary, # needed to produce the ROC/AUC measures
  savePredictions = TRUE # needed to plot the ROC curves
)
logReg_mod3 <- train(Current_Inspection ~ Type_Detail + Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
                     + Prev_Insp_missing + Time_Since_Insp_missing
                     + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                     + Time_Since_Aud_missing + Prev_Aud_missing
                     + Rejected_batches + Right_first_time
                     + Rejected_batches_missing + Right_first_time_missing
                     + recalculated_Complaints_product + recalculated_Complaints_product_missing
       , data = quality_wide_bin, method="glm", trControl = ctrl, family="binomial")
logReg_mod3
roc.logReg_mod3 <- roc(predictor = logReg_mod3$pred$yes, response = logReg_mod3$pred$obs)
auc.logReg_mod3 <- round(auc(roc.logReg_mod3), 3)
```

```
auc.logReg_mod3

# plot the OneR ROC curve and set up the identity line and axes
roc.logReg_mod3_roc <- plot(roc.logReg_mod3,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "Logistic Regression *AI prediction, (10-fold CV)"
)

# Build a confusion matrix
confusionMatrix(logReg_mod3$pred[order(logReg_mod3$pred$rowIndex),2],
logReg_mod3$pred$obs)
```

6.1.3b. Compliance history, production type and Operational KPIs - only records with KPIs

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 3-fold CV
  #repeats = 10,
  selectionFunction = "best", # select the best performer
  classProbs = TRUE, # requested the predicted probs (for ROC)
  summaryFunction = twoClassSummary, # needed to produce the ROC/AUC
measures
  savePredictions = TRUE # needed to plot the ROC curves
)
logReg_mod3 <- train(Current_Inspection ~ Type_Detail +
                       Previous_Inspection_Score_Normalized                    +
Time_Since_Last_Insp
```

```
        + Prev_Insp_missing + Time_Since_Insp_missing
        + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
        + Time_Since_Aud_missing + Prev_Aud_missing
        + Rejected_batches + Right_first_time
        + Rejected_batches_missing + Right_first_time_missing
        #+ recalculated_Complaints_product + recalculated_Complaints_product_missing
        #+ Six_Month_Avg._Rejected_batches + Six_Month_Avg._Right_first_time
        #+ Six_Month_Avg._Rejected_batches_missing + Six_Month_Avg._Right_first_time_missing
    , data = quality_wide_bin_op, method="glm", trControl = ctrl, family="binomial")
#logReg_mod3
roc.logReg_mod3 <- roc(predictor = logReg_mod3$pred$yes, response = logReg_mod3$pred$obs)

summary(logReg_mod3)
# plot the OneR ROC curve and set up the identity line and axes
roc.logReg_mod1_roc <- plot(roc.logReg_mod3,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "Logistic Regression *AI prediction, (10-fold CV)"
)
confusionMatrix(logReg_mod3$pred[order(logReg_mod3$pred$rowIndex),2], logReg_mod3$pred$obs)
```

6.1b Current risk predictions with Log Reg

Data preparation

First, we code the missing variables

```{r}
quality_wide_prediction <- quality_wide_current %>%
  mutate(
    Time_Since_Insp_missing = ifelse(is.na(Time_Since_Last_Insp), 1, 0),
    Time_Since_Last_Insp = ifelse(is.na(Time_Since_Last_Insp), 50, Time_Since_Last_Insp))
quality_wide_prediction <- quality_wide_prediction %>%
    mutate(
    Prev_Insp_missing = ifelse(is.na(Previous_Inspection_Score_Normalized), 1, 0),
    Previous_Inspection_Score_Normalized = ifelse(is.na(Previous_Inspection_Score_Normalized), 0, Previous_Inspection_Score_Normalized))
quality_wide_prediction <- quality_wide_prediction %>%
  mutate(
    Time_Since_Aud_missing = ifelse(is.na(Time_Since_Last_Audit), 1, 0),
    Time_Since_Last_Audit = ifelse(is.na(Time_Since_Last_Audit), 25, Time_Since_Last_Audit))
quality_wide_prediction <- quality_wide_prediction %>%
  mutate(
    Prev_Aud_missing = ifelse(is.na(Previous_Audit_Score_Normalized), 1, 0),
    Previous_Audit_Score_Normalized = ifelse(is.na(Previous_Audit_Score_Normalized), 0, Previous_Audit_Score_Normalized))
quality_wide_prediction <- quality_wide_prediction %>%
  mutate(
    Rejected_batches_missing = ifelse(is.na(Rejected_batches), 1, 0),
    Rejected_batches = ifelse(is.na(Rejected_batches), mean(Rejected_batches, na.rm = TRUE), Rejected_batches))
```

```
quality_wide_prediction <- quality_wide_prediction %>%
  mutate(
    Right_first_time_missing = ifelse(is.na(Right_first_time), 1, 0),
    Right_first_time = ifelse(is.na(Right_first_time), mean(Right_first_time, na.rm = TRUE), Right_first_time))
quality_wide_prediction <- quality_wide_prediction %>%
  mutate(
    recalculated_Complaints_product_missing = ifelse(is.na(recalculated_Complaints_product), 1, 0),
    recalculated_Complaints_product = ifelse(is.na(recalculated_Complaints_product),mean(recalculated_Complaints_product, na.rm = TRUE), Right_first_time))
quality_wide_prediction <- quality_wide_prediction %>%
  mutate(
    Six_Month_Avg._Rejected_batches_missing = ifelse(is.na(Six_Month_Avg._Rejected_batches), 1, 0),
    Six_Month_Avg._Rejected_batches = ifelse(is.na(Six_Month_Avg._Rejected_batches), mean(Six_Month_Avg._Rejected_batches, na.rm = TRUE), Six_Month_Avg._Rejected_batches))
quality_wide_prediction <- quality_wide_prediction %>%
  mutate(
    Six_Month_Avg._Right_first_time_missing = ifelse(is.na(Six_Month_Avg._Right_first_time), 1, 0),
    Six_Month_Avg._Right_first_time = ifelse(is.na(Six_Month_Avg._Right_first_time), mean(Six_Month_Avg._Right_first_time, na.rm = TRUE), Six_Month_Avg._Right_first_time))

summary(quality_wide_prediction)
```

```
```

Application of model from 6.1.3b.:

```{r}
quality_wide_prediction$result <- predict(logReg_mod3, quality_wide_prediction, type = "prob")
#quality_wide_prediction[89] <- NULL
quality_wide_prediction_extr <- quality_wide_prediction[c(3,89)]
colnames(quality_wide_prediction_extr)[2] <- "result"
#quality_wide_prediction %>%
#  rename(
#    action_indicated = result.yes
#   ,no_action_indicated = result.no
#    )
ggplot(data=quality_wide_prediction, aes(x=Site, y=result)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=len), vjust=-0.3, size=3.5)
```

6.2. Decision tree models

6.2.3. Compliance history, production type and Operational KPIs

```{r}
trctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 3,
                       selectionFunction = "best", # select the best performer
  classProbs = TRUE, # requested the predicted probs (for ROC)
  summaryFunction = twoClassSummary, # needed to produce the ROC/AUC measures
  savePredictions = TRUE # needed to plot the ROC curves
  )
```

```
set.seed(3333)
DTbin3 <- train(Current_Inspection ~ Type_Detail +
                  Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
                #+ Prev_Insp_missing + Time_Since_Insp_missing
                + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                #+ Time_Since_Aud_missing + Prev_Aud_missing
                + Rejected_batches + Right_first_time
                #+ Rejected_batches_missing + Right_first_time_missing
                #+ recalculated_Complaints_product +
recalculated_Complaints_product_missing
        , data = quality_wide_bin_op, method = "rpart",
                  parms = list(split = "information"),
                  trControl=trctrl,
                  tuneLength = 10)
roc.DTbin3 <- roc(predictor = DTbin3$pred$yes, response = DTbin3$pred$obs)
auc.DTbin3 <- round(auc(roc.DTbin3), 3)
auc.DTbin3
fancyRpartPlot(DTbin3$finalModel,cex=0.65)
DTbin3
# plot the OneR ROC curve and set up the identity line and axes
roc.DTbin3 <- plot(roc.DTbin3,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "Decision Tree *AI prediction, (10-fold CV)"
)

```
```
```

6.3. Random Forest

6.3.1. Create Random Forest for binary prediction and ROC curve with n-fold CV

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 3-fold CV
  selectionFunction = "best", # select the best performer
  classProbs = TRUE, # requested the predicted probs (for ROC)
  summaryFunction = twoClassSummary, # needed to produce the ROC/AUC measures
  savePredictions = TRUE # needed to plot the ROC curves
)

grid <- expand.grid(.splitrule = "gini",
                    .mtry = 1:6,
                    .min.node.size = c(5, 8, 10, 15, 20))

RF_mod <- train(Current_Inspection ~ #Type_Detail +
                 Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
                #+ Prev_Insp_missing + Time_Since_Insp_missing
                + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                #+ Time_Since_Aud_missing + Prev_Aud_missing
                + Rejected_batches + Right_first_time
                #+ Rejected_batches_missing + Right_first_time_missing
                #+ recalculated_Complaints_product + recalculated_Complaints_product_missing
        , data = quality_wide_bin_op,
                method = "ranger",
                trControl = ctrl,
```

```
            tuneGrid = grid,
            metric = "ROC",
            num.trees = 10,
            importance = 'impurity')
RF_mod

importance <- as.data.frame(varImp(RF_mod)$importance)
importance <- as.data.frame(cbind(rownames(importance), importance, row.names=NULL))
names(importance)[1] <- "factor"
names(importance)[2] <- "f_importance"
importance


importance_plot <- ggplot(data=importance, aes(x=reorder(factor,f_importance), y=f_importance)) +
  geom_bar(stat="identity", position="dodge") +
  ylab("Variable Importance") +
  xlab("") +
  scale_fill_gradient(low="red", high="blue") +
  coord_flip()
importance_plot

# save the predicted probabilities and actual values for OneR
roc.RF_mod <- roc(predictor = RF_mod$pred$yes, response = RF_mod$pred$obs)
auc.rf_mod <- round(auc(roc.RF_mod), 3)
auc.rf_mod
```

```
# plot the OneR ROC curve and set up the identity line and axes
roc.rf_mod_roc <- plot(roc.RF_mod,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "Random Forest *AI prediction, (10-fold CV)"
)
```

6.3.2. Create Random Forest for binary prediction and ROC curve with separate training an testing

```{r}
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 3, # 3-fold CV
  selectionFunction = "best", # select the best performer
  classProbs = TRUE, # requested the predicted probs (for ROC)
  summaryFunction = twoClassSummary, # needed to produce the ROC/AUC measures
  savePredictions = TRUE # needed to plot the ROC curves
)

grid <- expand.grid(.splitrule = "gini",
                    .mtry = 2:4,
                    .min.node.size = c(10, 20))

RF_mod <- train(Current_Inspection ~ Type_Detail + Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
                + Prev_Insp_missing + Time_Since_Insp_missing
                + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                + Time_Since_Aud_missing + Prev_Aud_missing
```

```
            #+ Rejected_batches + Right_first_time
            #+ Rejected_batches_missing + Right_first_time_missing
            #+ recalculated_Complaints_product + recalculated_Complaints_product_missing
      , data = quality_wide_bin_op,
            method = "ranger",
            trControl = ctrl,
            tuneGrid = grid,
            metric = "ROC",
            num.trees = 50)

# save the predicted probabilities and actual values for OneR
roc.RF_mod_train <- roc(predictor = RF_mod$pred$yes, response = RF_mod$pred$obs)
auc.rf_mod_train <- round(auc(roc.RF_mod_train), 3)
auc.rf_mod_train

# plot the OneR ROC curve and set up the identity line and axes
roc.rf_mod_roc <- plot(roc.RF_mod_train,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "RF xAI prediction, (3-fold CV) - training"
)

# Test predicitions and evaluation
quality_wide_bin_train$prob <- predict(RF_mod, quality_wide_bin_test)

roc.RF_mod_test <- roc(predictor = quality_wide_bin_train$prob, response = quality_wide_bin_train$Current_Inspection)
```

```
auc.rf_mod_test <- round(auc(roc.RF_mod_test), 3)
auc.rf_mod_test
roc.rf_mod_roc_test <- plot(roc.RF_mod_test,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "RF xAI prediction, (3-fold CV) - test"
)
```

6.4. kNN using last inspection and time since; RFT and Rejected batches and missing variables

```{r}
ctrl <- trainControl(method="repeatedcv",repeats = 10,classProbs=TRUE,summaryFunction = twoClassSummary)

kNN_mod <- train(Current_Inspection ~ Type_Detail +
                   Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
                 + Prev_Insp_missing + Time_Since_Insp_missing
                 + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                 + Time_Since_Aud_missing + Prev_Aud_missing
                 + Rejected_batches + Right_first_time
                 + Rejected_batches_missing + Right_first_time_missing
    , data = quality_wide_bin_op,   method = "knn", trControl = ctrl, preProcess = c("center","scale"), tuneGrid = expand.grid(k = 1:25))
kNN_mod

confusionMatrix(kNN_mod)

roc.kNN_mod <- roc(predictor = kNN_mod$pred$yes, response = kNN_mod$pred$obs)
```

```
auc.kNN_mod <- round(auc(roc.kNN_mod), 3)
auc.kNN_mod
```

6.5. Neuralnet using last inspection and time since; RFT and Rejected batches and missing variables

```{r}
nnetGrid <-  expand.grid(size = seq(from = 1, to = 10, by = 1),
                         decay = seq(from = 0.1, to = 0.5, by = 0.1))
ctrl <- trainControl( # create a caret trainControl object
  method = "cv", number = 10, # 10 or 3-fold CV
  savePredictions = TRUE, # needed to plot the predictions
  classProbs = TRUE
)
NN_mod <- train(Current_Inspection ~ Type_Detail +
                  Previous_Inspection_Score_Normalized + Time_Since_Last_Insp
                + Prev_Insp_missing + Time_Since_Insp_missing
                + Previous_Audit_Score_Normalized + Time_Since_Last_Audit
                + Time_Since_Aud_missing + Prev_Aud_missing
                + Rejected_batches + Right_first_time
                + Rejected_batches_missing + Right_first_time_missing
    , data = quality_wide_bin_op,     method = "nnet", linear.output = TRUE, tuneGrid = nnetGrid ,
    metric = "ROC",
    trControl = ctrl)
NN_mod
summary(NN_mod)
# save the predicted probabilities and actual values for OneR
roc.NN_mod <- roc(predictor = NN_mod$pred$yes, response = NN_mod$pred$obs)
```

```
auc.NN_mod <- round(auc(roc.NN_mod), 3)
auc.NN_mod

# plot the OneR ROC curve and set up the identity line and axes
roc.NN_mod <- plot(roc.NN_mod,
  identity.lty = 2,  # plot the identity line (diagonal) as a dashed line
  legacy.axes = TRUE, # x-axis will be the false positive rate (1 - Specificity)
  main = "NNet *AI prediction, (10-fold CV)"
)
plot(NN_mod$finalModel, rep = NULL, x.entry = NULL, x.out = NULL,
  radius = 0.15, arrow.length = 0.2, intercept = TRUE,
  intercept.factor = 0.4, information = TRUE, information.pos = 0.1,
  col.entry.synapse = "black", col.entry = "black",
  col.hidden = "black", col.hidden.synapse = "black",
  col.out = "black", col.out.synapse = "black",
  col.intercept = "blue", fontsize = 12, dimension = 8,
  show.weights = TRUE, file = NULL)
```

Curriculum Vitae

Name	**Steffen Eich**
Date/place of birth	April 29, 1992 in Langenfeld (Rhld.), Germany
Nationality	German

Work Experience

	Georg Fischer Piping Systems, Schaffhausen
Since 09/2021	Global Technology Manager
	M-Industrie AG (Migros Group), Zurich
11/2020– 08/2021	Center of Excellence, Sales & Operations Planning
	Institute of Technology Management, University of St.Gallen
08/2017 – 10/2020	Research Associate in Operational Excellence
	Laboratory for Machine Tools & Production Engineering (WZL), RWTH Aachen
08/2013 –08/2015	Student Employee in R&D / Production Management
	BMW Group, Munich
09/2014 – 03/2015	Intern/Bachelor's Candidate in Product Modularization

Education

	University of St.Gallen (HSG)
08/2017– 02/2022	**Doctor of Philosophy** in Management focusing on Business Innovation
	Executive School, University of St.Gallen (HSG)
09/2019 – 07/2021	**CAS** in Didactics for Higher Education
	Massachusetts Institute of Technology, Cambridge (MA)
10/2016 – 03/2017	Scholar, Consortium for Engineering Program Excellence
	清华大学 **Tsinghua University**, Beijing
09/2015 – 07/2016	**M.Sc.** in Power Engineering & Engineering Thermophysics
	RWTH Aachen University
04/2015 – 07/2017	**M.Sc. RWTH** in Production Engineering
10/2011 – 03/2015	**B.Sc. RWTH** in Mechanical Engineering
08/2002 – 07/2011	**Konrad-Adenauer-Gymnasium**, Langenfeld
	Abitur (A-levels)