Application of Process Analysis and Optimization tools in Hit-to-Lead and Lead Optimization phases of Drug Discovery at EPP, NIBR

By

Tatiana Yglesias
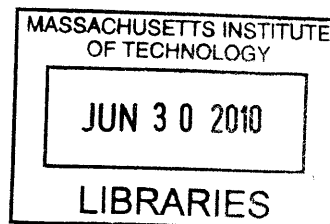
Licenciatura[1] of Science in Chemical Engineering, Universidad de Costa Rica, 2005

Submitted to the MIT Sloan School of Management and the Department of Chemical Engineering

in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration

AND

Master of Science in Chemical Engineering

In conjunction with the Leaders for Global Operations Program at the

Massachusetts Institute of Technology

June 2010

Signature of Author_____

$\cup O$ May 7, 2010

Chemical Engineering, MIT Sloan School of Management

Certified by_____

⏝. Christopher Love, Thesis Supervisor

Assistant Professor, Department of Chemical Engineering

Certified by_____

Roy Welsch, Thesis Supervisor

Professor of Statistics and Management Science, MIT Sloan School of Management

Accepted by_____

William Deen, Professor of Chemical Engineering

Chairman, Committee for Graduate Students, Chemical Engineering

Accepted by_____

Debbie Berechman

Executive Director of MBA Program, MIT Sloan School of Management

---

[1] Intermediate degree between a B.S. and a M.S., Thesis and Internship completed for this degree

This page has been intentionally left blank

Application of Process Analysis and Optimization tools in Hit-to-Lead and Lead Optimization phases of Drug Discovery at EPP, NIBR

By
Tatiana Yglesias
Submitted to the MIT Sloan School of Management and the Department of Chemical Engineering Division on May 7, 2010 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Chemical Engineering

ABSTRACT

Given that research is based on innovation, it has been believed that its activities can only be optimized with equipment upgrade, increment in personnel scientific knowledge, development of new analytical software and/or changing the areas of study. After realizing the limited results achieved with these approaches, lab representatives started to notice the opportunity of introducing process optimization tools, such as Lean and Six Sigma, which showed success in manufacturing environments,.

This project analyzes the interrelation between process and results, providing a clear explanation of cause and effect conditions, and a concise list of areas for improvement. Specifically, the document defines a measurement system using process maps and key performance indicators (KPIs). With this, the document describes the current state through historic trends, provides a complete data and root cause analysis for current state description, and provides a process capability study for the available indicators.

Implementation of the steps mentioned above show how focus in lab turnaround times have been deviating attention from more impactful improvements, which can greatly affect overall drug discovery duration. Also, the analysis identifies that constant technology changes caused constant adaptation of process procedures, which generated non-value added activities. These non-value added activities today occupy about 50% of a lab associate's time. Lastly, historic data evaluation shows that root cause statistical analysis is limited by the presence of a combination of special and common cause variations.

Some of the project recommendations include: incorporation of chemist's knowledge about compound potency, integration of equipment and software information, change in booking system, incorporation of assay and plate criteria, definition of standard procedures for specific activities, and integration of assay development and data submission tools. Overall, these changes can lead to a 50% reduction in the profiling times greater than 60 days, decrease of 62% and 60% in Compound Manager (CM) and Compound Profiler (CP) non-value added times respectively, 30% decrease in CM and CP total duration per assay plate, and increase in profiling time stability and predictability. Despite the fact that timing and scale of available resources will impact the realized benefits, the proposed framework gives EPP the opportunity to assess the improvements by their effect and alignment with goals.

Thesis Supervisor: J. Christopher Love
Title: Assistant Professor, Department of Chemical Engineering

Thesis supervisor: Roy Welsch
Title: Professor of Statistics and Management Science, MIT Sloan School of Management

This page has been intentionally left blank

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

Food and Drug Administration (FDA): The FDA is an agency inside the United States Department of Health and Human Services, which has the task of protecting and promoting public health. For this reason, one of the main activities is to regulate and supervise products related to human health like medical devices, pharmaceutical and biopharmaceutical drugs, and dietary supplements.

New molecular entity (NME): A NME, also called New Chemical Entity (NCE), is as a unique compound that has not been previously approved by the FDA. NMEs are usually developed by companies in early drug discovery stage, have been synthesized and will be employed and tested during Clinical Trials.

Hit-to-Lead (H2L): Is the phase of drug discovery where many compounds are tested to find the best combination of structure and activity for a particular target. At the end of the H2L phase, a compound known as hit candidate is nominated to continue further tests.

Lead Optimization (LO): After the hit compound has been approved, the LO phase seeks to optimize the chemistry of the compound such that activity, pharmacokinetics and pharmacodynamics are maximized.

Lead compound: Compound selected at the end of LO phase, which optimally modulates the activity of a receptor or other target protein. A successful lead compound becomes a drug candidate for further development.

The National Institutes of Health (NIH): The NIH, as the FDA, is an agency of the United States Department of Health and Human Services and is the primary agency of the United States government responsible for biomedical and health-related research. Its mission is to promote global health by sharing new knowledge in the biomedical arena.

Profiling assay and AC50 value: Also known as dose-respond curve, potency or inhibition assay, this test seeks to measure the binding occurred between the compound and the target in evaluation. Specifically, an assay is a procedure for detecting the presence, estimating the concentration, and determining the biological activity of a compound-target combination[2]. Assays are based on measurable parameters that enable the evaluation of differences between samples and controls. This measurement is given by the AC50 value, which represents the concentration of substance that provides 50% inhibition.

---

[2] (http://www.dddmag.com/Glossary.aspx)

Target: Is a DNA, RNA, or protein that is involved in a disease process and is a suitable for therapeutic compound development. A target defines the type of disease research efforts and is the basis to develop new compounds.

Compound: From a chemistry perspective, a compound is as a pure chemical substance consisting of two or more different chemical elements that can be separated into simpler substances by chemical reactions. In drug discovery, a compound is the chemical entity that is tested towards certain target and will be optimized further if its activity looks promising.

High Throughput Screening (HTS): Automated profiling process that uses a large number of assays to identify active compounds. HTS analyzes large numbers of compounds, resulting in a less costly and faster process.

Positive Control and Negative Control: These are wells in an assay plate that are used as boundary control. They set the range for expected IC50 values. Positive control confirms that the experiment conditions have a positive result, even if none of the compounds result active. Positive controls are assay wells that contain only enzyme, so, these wells show maximum enzyme activity. In the other hand, negative control defines the lowest possible, given that these wells contain no enzyme.

Reference compound: Is a compound with known activity for a specific target. These compounds are used in enzymatic assays as quality control. They allow analysis of trends and differences, not only across assays but also in different repetitions of the same assay.

1. INTRODUCTION

1.1 Project Drivers: Current Challenges in Drug Discovery

The pharmaceutical industry is facing a difficult period with increasing R&D expenditures, Food and Drug Administration (FDA) changing towards more complex new drug acceptance criteria and fierce competition among all the players. Increase in regulator's requirements, combined with the challenges to discover new targets and/or novel medical entities (NMEs), forces drug manufacturers to focus on time-to-market and cost allocation. While time-to-market allows companies to start investment recovery early on, adequate cost allocation ensures financial stability after products loose their patents. When evaluating the complete pharmaceutical process, drug discovery stands out with the most duration and need for resources. Because of this, optimization of Research and Development (R&D) can impact overall company results[3].

In response to the current pharmaceutical industry status, Novartis has shifted research strategy to focus on discovery projects that address the patients with greatest medical needs while enabling the development of a powerful scientific knowledge. Consequently, Novartis Institute for BioMedical Research (NIBR) integrated Scientific Centers for particular Disease Areas and formed Expertise Platforms committed to technologies and procedures of specific scientific fields. At the Expertise Platform Proteases (EPP), where this project is developed, Biology (BIO), Medicinal Chemistry (MCH) and Structural Sciences (STS) teams focus on discovery and optimization of novel compounds to deliver highly selective drugs which inhibit proteases, one of the main target classes in the human genome. These teams also work in collaboration with all Disease Areas for projects developed in and out the platform.

Beside organizational changes, Novartis has been implementing new ideas in both, manufacturing and R&D. For manufacturing, Novartis followed the initiative of other companies in adopting process improvement tools to achieve cost reduction and quality enhancement while stabilizing the company's profits. In the case of research, Novartis has been incorporating innovative technologies and diverse automation, such as High Throughput Screening (HTS) and combinatorial chemistry. Despite of observing some benefits, soon after the first technological upgrades were in place, R&D costs started far outweighing the observed gains. At this point, it became evident that sustainable benefits could only be

---

[3] (Steven Paul, 2010)

achieved by a change in the status-quo, creating an expectation of the possible outcomes if process optimization tools were to be applied[4].

1.2 Problem Statement

The journey to adopt and implement process optimization tools in research has been a challenge for drug manufacturers. Difficulty rises when focusing on upgrading time consuming and costly discovery processes with extremely high failure rates. Risks in drug research rely on the unpredictability of the compound's behavior towards certain target and are limited by the scientific knowledge. In addition to this "natural" variability, lack of reproducibility and repeatability of results adds to the variability of defining a therapeutic candidate, further increasing research uncertainty. As a result, many improvements can be obtained by the implementation of standards and the use of common methods among laboratories, which can decrease process variability but also facilitate improvement of equivalent tasks and decision-making processes. Lastly, standardization can also facilitate comparability of results across research labs.

EPP has been transforming operations within the Biology Unit (EPP/BIO), obtaining significant reduction of assay turnaround time. Despite these efforts, hurdles on critical project resolutions, a demanding work environment, and constant change in laboratory methods have created limited communication among players, causing internal discomfort with EPP's overall results.

Consequently, this project comes as a first approach for specific process analysis to deeply evaluate EPP's procedures within Hit-to-Lead (H2L) and Lead Optimization (LO) Phases of Drug Discovery, define impactful areas of improvement based on process analysis, identify key stakeholders, and evaluate the alignment of expectations and capabilities.

1.3 Thesis Overview

The document has the following structure:

Chapter 1 describes the drivers of the project and the general outline of the document.

Chapter 2 illustrates the pharmaceutical industry's dynamics and limitations, along with the specifics of NIBR and its role in Novartis drug discovery process.

Chapter 3 states the hypothesis and the project approach.

---

[4] (Pollok, 2005)

Chapter 4 shows the details of EPP's process analysis and explains the method for defining the proposed measurement system.

Chapter 5 presents a thorough analysis of the data for each one of the indicators of the proposed measurement system. The analysis is explained individually for each indicator defining first the current state using historic data evaluation, then listing and evaluating possible root causes, and finally realizing a study of the process capability.

Chapter 6 provides a summary of the results, lists specific recommendations and their impact on EPP's performance, presents internal challenges that can limit the implementation of the proposed solutions, and suggests future areas of study.

2. BACKGROUND

2.1. The Pharmaceutical Industry

In a general, a company in the pharmaceutical industry discovers, develops, manufactures, and markets a variety of medicinal goods, from life saving to health improvement products. Given its nature, the pharmaceutical industry is subject to a large variety of laws and regulations regarding the patenting, testing and marketing of drugs. These laws and regulations depend not only on the type of drug but also on the specifics of application, patient focus and location (selling and manufacturing)[5,6].

The complete process from discovery to manufacturing of a pharmaceutical product can take anywhere from 8 to 12 years, with an overall average cost of $800 to $1,200 million[7]. The specific duration and costs depend on an infinite number of variables such as the target, complexity of the pathways, properties of the compounds in evaluation, number of patients, and even location of clinical trials. A diagram with the phases and an average duration is shown in Figure 2. In the new drug application (NDA) step, the drug is approved and further phases include manufacturing, marketing and distribution.
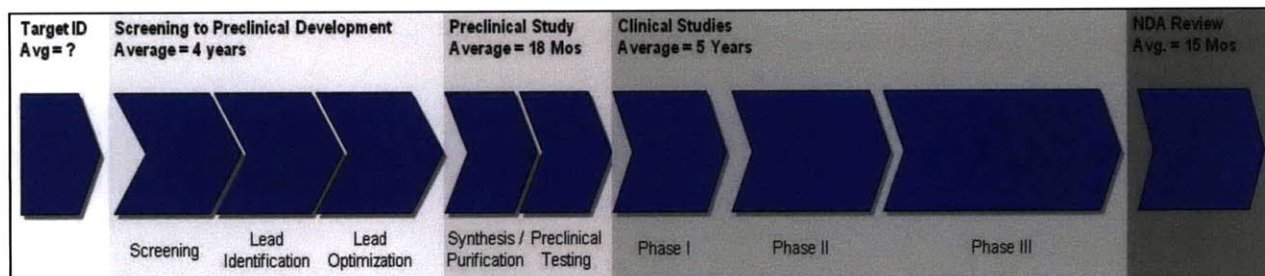
Figure 2. Classification of pharmaceutical industry phases, from discovery to market[8]

From all the value chain of drug manufacturing, the most costly and risky portion is discovery, including clinical trials. Table 1 presents some numbers for costs and probabilities of success for each phase of drug discovery. In a more comprehensive manner, Figure 3 compares the average risk (related to the probability of success) and the value creation over the total time from discovery to market (assuming total duration of 20 years).

---

[5] (United States Department of Labor, last modified: December 17, 2009)
[6] (World Health Organization (WHO), 2004)
[7] (Fee March 01, 2007)
[8] (Adapted from Fee, March 01, 2007)

Table 1. Average industry costs and probabilities of success for drug compounds[9]

	Cost (M)	Probability (%)
Phase I	$2-10	70
Phase II	$5-50	25-35
Phase III	$30-$100s	25-35
Aggregate	$50-800	4-10
Median	$250	8
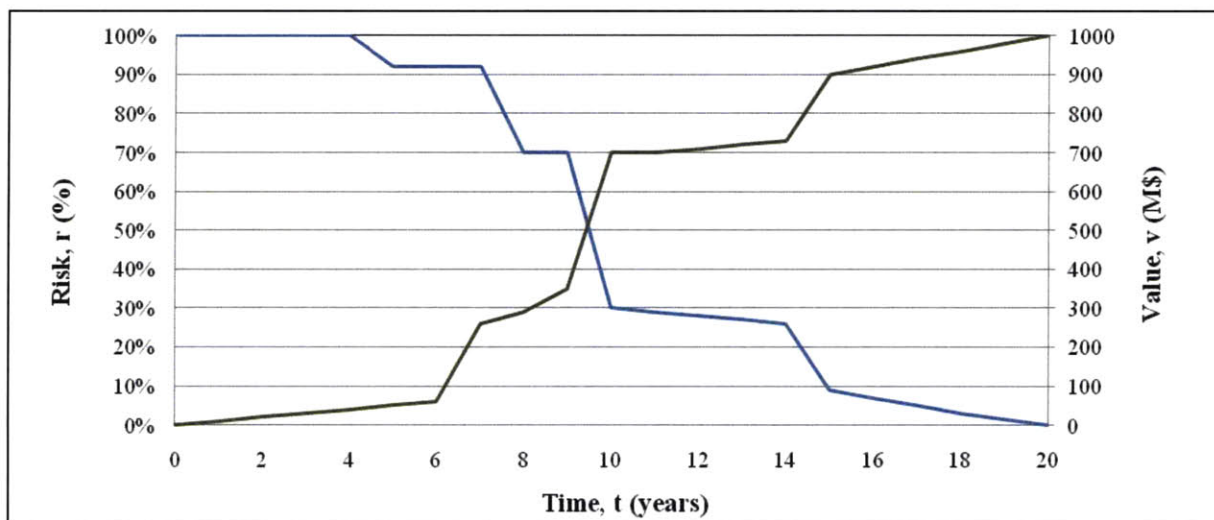
Figure 3. Cost and risk average values for pharmaceutical companies for 2009[10]

In Figure 2, clinical trials end after about 10 years, causing the steep decrease in risk and increase in costs at that time. After that step, additional costs represent manufacturing, marketing and other costs related to making the drug available in the market. Uncertainties after this stage only depend on selling and manufacturing constraints, but are negligible when compared to a compound starting tests or a non-patented drug. Given the combination of risk and costs, value creation peaks up when drugs have sufficient studies to demonstrate effectiveness to regulators.

Along the years, pharmaceutical companies have been trying to discover novel compounds for unique and un-studied target pathways. Given mentioned costs and uncertainty, along with an increase in prerequisites to receiving a patent, investments in research have been increasing with disproportional results in new drug discovery. Most of the current critiques regarding the increase R&D expenditures are based on the decrease in the number of registered NMEs. Figure 4 shows a graph with the trends of R&D expenditures, approval of NMEs, and other metrics that characterize the trends of pharmaceutical industry from 1980 to 2010. As Figure 4 shows, in the 1980's and 1990's, companies spent close to 17% of their

---

[9] (Fee March 01, 2007)

[10] (Modified from United States Department of Labor, last modified: December 17, 2009)

revenues in R&D. These days, that number has gone up to almost 20% while the number of NDA presents a steady (if not a decrease) trend[11]. Even when some experts argue that current focus has been on improving current approved therapeutics and/or increasing the number of disease treated by established indications, is still unclear how sustainable is spending trend in pharma R&D.
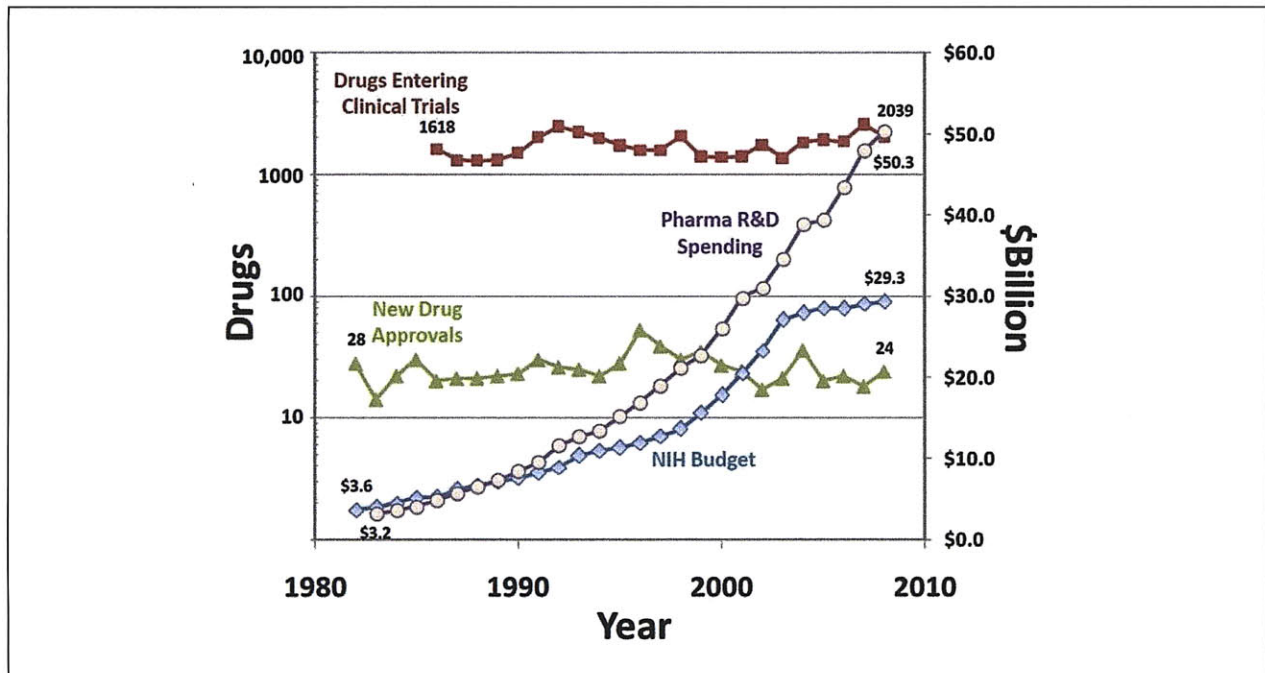
Figure 4. Diagram of average pharmaceutical industry dynamics over the last 30 years [12]

2.2. Novartis AG

Novartis is a multinational pharmaceutical company with headquarters located in Basel, Switzerland. It was created in 1996 through the merger of Ciba-Geigy and Sandoz. In 2009, Novartis was ranked number one in revenues and number three in sales ($44, 267 millions in 2009)[13]. Novartis has been growing not only in the pharmaceutical markets but also in the areas of vaccines, generics and consumer health, reaching almost 100,000 full-time employees around the world.

Novartis Pharmaceuticals has a portfolio including several disease areas such as Cardiovascular and Metabolism, Oncology, Neuroscience and Ophthalmics, Respiratory, and Immunology and Infectious

---

[11] (Food and Drug Administration Last updated: August 12, 2009)
[12] (Taken from www.veomed.com, Medical Student Visual Learning Resource)
[13] (Novartis Company Website)

Diseases. In this collection of areas, Novartis has more than 50 marketed products, many of which are leaders in their respective therapeutic areas[14].

To maintain the mission of discovering and developing innovative products, during 2008 Novartis increased R&D investments by 12 percent to USD 7.2 billion. This is one of the highest numbers in the industry relative to sales (17.4 percent)[15]. Figure 5 shows the increment of Novartis R&D expenditures over time, from 2004 to 2008.
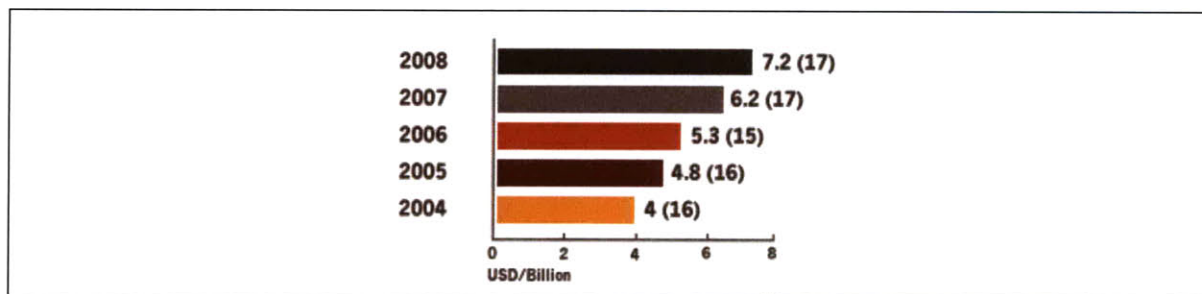
Figure 5. Trend of investment with respect to sales for Novartis[16]

The Novartis Institutes for BioMedical Research (NIBR) is the global pharmaceutical unit within Novartis, with about 5,000 associates distributed in 4 countries (US, Switzerland, UK and China). In order to boost discovery, NIBR has focused on developing key competencies in distinct areas, by combining automation, robotics, computational science, biology and chemistry into diverse expertise units. The objective with this approach is to have a strong scientific knowledge in a specific arena that collaborates with any discovery project across research.

Three main divisions were developed during the implementation of mentioned approach for improvement: Discovery Science, Clinical Sciences and Disease Areas. Clinical Sciences bridge bench science and clinical medicine by optimizing the performance of a new drug through drug metabolism and pharmacokinetics, and by employing translational medicine to interpret biology into medicine[17].

The Disease Areas group applies its knowledge to specific categories of therapeutic field. Some of the areas involved are Autoimmunity, Transplantation and Inflammatory Disease (ATI), Cardiovascular and Metabolic Diseases, Gastrointestinal Disease, Infectious Diseases, Oncology , Ophthalmology, and Respiratory Diseases.

---

[14] (Novartis Company Website)
[15] (Novartis Company Website)
[16] (From Novartis Company Website)
[17] (Novartis Company Website)

Lastly, Discovery Sciences focus mainly on developing technological expertise to create a common approach towards similar disease mechanisms. Some of the areas covered by Discovery Sciences are biologics, biomarker development, imaging, metabolisms, pharmacokinetics, preclinical safety, and proteomic chemistry. This last unit, known as the Center for Proteomic Chemistry (CPC), brings together biology and chemistry to study the interactions of small chemical molecules with biological macromolecules, generally proteins. The CPC group studies compound-target interactions by biophysical, protein structure, and cheminformatics approaches. Validated hits with the desired properties are passed to Global Discovery Chemistry (GDC) for further optimization[18].

To create more specialization and support to each disease area, CPC has established three platforms; two dedicated to target families, Expertise Platform Kinases (EPK) and Expertise Platform Proteases (EPP), and a group dedicated to natural products research. With this structure, CPC's capabilities include high-throughput screening, preclinical safety profiling, characterization of protein structure, discovery of natural products discovery, and research of protease and kinase targets[19].

EPP plays a vital role in NIBR given its focus on one of the main targets in the human genome. In general, EPP is involved in the discovery and optimization of compounds to deliver highly selective drugs that inhibit proteases. As shown in Figure 6, all areas are involved in each of the projects developed compounds in a shorter time. Complementing Figure 6 within the unit, creating a teamwork environment Figure 9 presents the involvement of each group throughout research and describes teamwork among units.

---

[18] (Novartis Company Website)
[19] (Novartis Company Website)
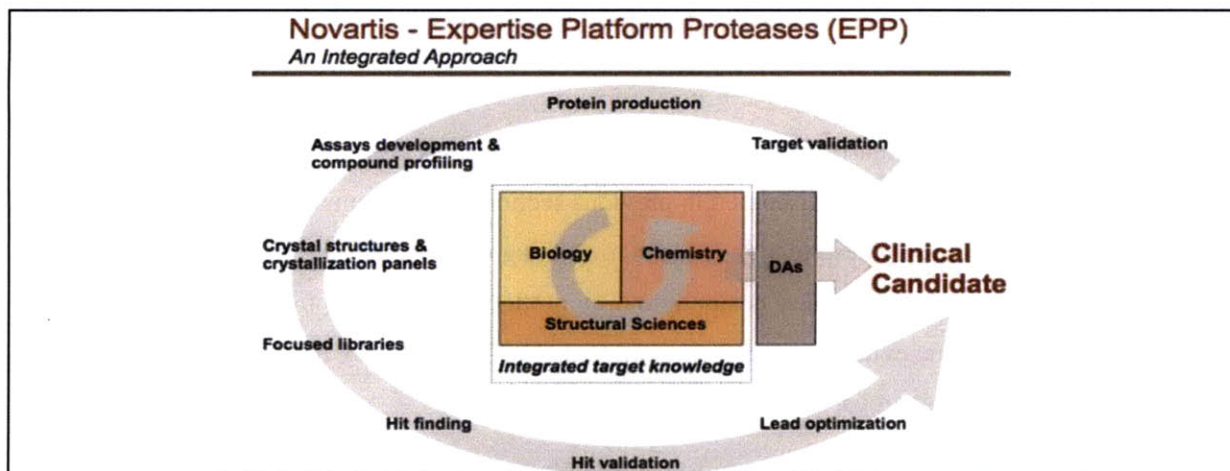
Figure 6. EPP diagram showing the integrated work between units[20]
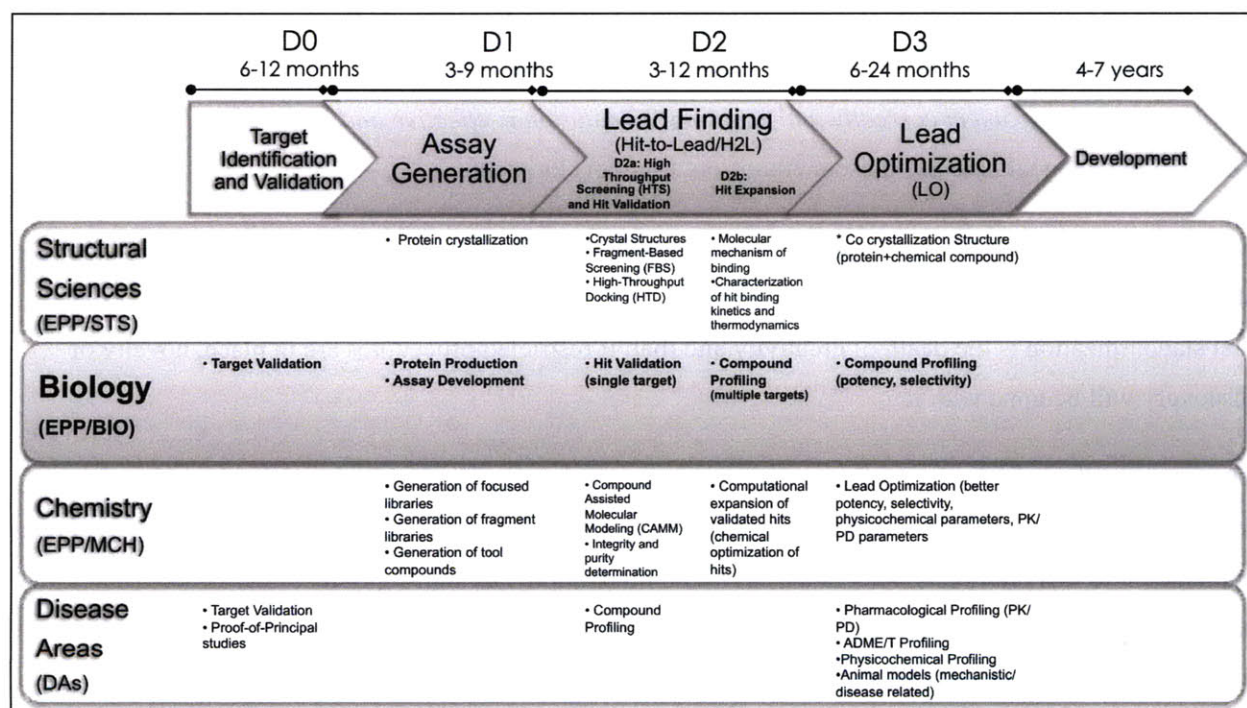
Figure 7. Phases of Novartis R&D and involvement of groups within EPP[21]

2.3. Recent initiative for drug discovery optimization

Starting about a decade ago, companies in all types of industries began implementing process optimization tools to control and systematically decrease costs while achieving better quality in the final product, service, or any metric used to evaluate customer satisfaction. In general, any tool used for

---

[20] (Modified from EPP's internal documentation)
[21] (Modified from EPP's internal documentation)

process improvement aims to find activities that deviate resources from creating value towards the final needed result.

Margins that characterize the pharmaceutical business have sometimes been observed as a limitation to encourage process optimization. As discussed in previous sections, when companies started to observe the decrease in new drug discovery despite high investments, along with an increase in customer awareness, and the shift of the FDA towards quality design into drug products, the need to focus on the way processes take place became evident. At this moment, tools like Lean and Six Sigma started to be implemented in pharma operations.

Novartis Pharmaceuticals started their internal approach towards process optimization by introducing IQP, which stands for innovation, quality and productivity. IQP is Novartis initiative to implement Lean, Six Sigma and other methodologies in their manufacturing facilities around the world. This approach "*is different to the [s,c] traditional management styles focusing on short term 'quick fix' solutions rather than on identifying the problem correctly and ensuring the solution is effective and sustainable long-term*[22]". Today, Novartis goal is to use these methods on drug development.

But even when advantages are obvious, the innovative and unpredictable nature of research has been stopping drug discovery from adopting these tools. Most of the scientists that work in research believe that standardization is the death of creativity and that, if restricted procedures are in place, novelty of discovery will be gone with it.

To overcome these fears, experts in process improvement have studied and piloted these tools in research laboratories. Interestingly, pilots revealed a strong correlation between innovation and standardization. An explanation to this relation is that results start with innovation, but their reproducibility comes from standardizing that innovation. In this way, innovation is required to identify targets and compounds, to understand scientific pathways, and to develop assays, but standardization is necessary to ensure consistency in the tests, to specify further assays, and to warranty reproducibility of inhibition down the road. Also, standardization forms the basis for further, process improvement, and innovation enables such improvement. In this way, process optimization can form part of the dynamic research loop[23].

Accordingly, application of process improvement tools in pharmaceutical R&D focuses on identification of common processes, which can be further optimized by reducing waste in the system, finding benefits of process control, and observing opportunities for standardization. One of the most impactful results of

---

[22] (Next Generation Pharmaceutical, 2010)
[23] (Goodman, 2010)

process improvement in drug discovery is the reduction of variation in potency results, leading to an increase in predictability and reproducibility of compound's data. Because Lean-Six Sigma provides a structured and data driven path for improvements, this tool has high potential in the highly scientific environment of R&D, generating also an increase on workers' engagement[24].

---

[24] (Wood, 2006)

3. HYPOTHESIS AND PROJECT OVERVIEW

3.1. Hypothesis

Critical differences between common manufacturing processes and laboratory procedures have created the paradigm that research cannot be optimized, at least without giving away some of the key benefits that process improvement brings or limiting science. Even when bringing drugs to market is surrounded by unpredictability, limitations faced by drug discovery R&D are not unique and can be compared with other industries where process optimization has been achieved. Understanding the needs of all the players in the H2L and LO loop, how these needs fit into EPP's overall strategy, and where improvements need to occur, is key for ensuring concrete long-lasting benefits through process analysis.

In order to create a framework to evaluate EPP's operations, this study divides the study in three categories: quality, effectiveness and efficiency. These categories allow a thorough evaluation of the current processes and facilitate evaluation of improvement areas.

The current need for streamlining processes within EPP comes from years of profiling upgrade attempts, focus on narrow indicators for measurement of performance, lack of process analysis, and need for implementing methodological procedures. Known tools such as Lean, Six Sigma or Novartis' IQP could offer a great impact in EPP's organizational performance, without requiring major investments or drastic modifications to actual procedures.

3.2. Project Scope and Approach

Even when H2L and LO specifics differ for every disease target and scientific study arena, general methodologies are comparable. Based on the above-mentioned limitations for process improvement, the aim of this project is to closely look at EPP's procedures, understand its dynamics, and create a foundation for the improvements application in other labs within research. As a result, this project applies an integrated approach, incorporating decision-making methods in daily activities, while creating a measurement system that reflects and aligns overall research mission. With this, internal performance can be tracked, operations within different labs can be compared, and best practices can be shared, thus resulting in a detailed breakdown of strengths and weaknesses, to frame final recommendations. Figure 8 shows the main area of study derived from the hypothesis and outlines general study objectives.
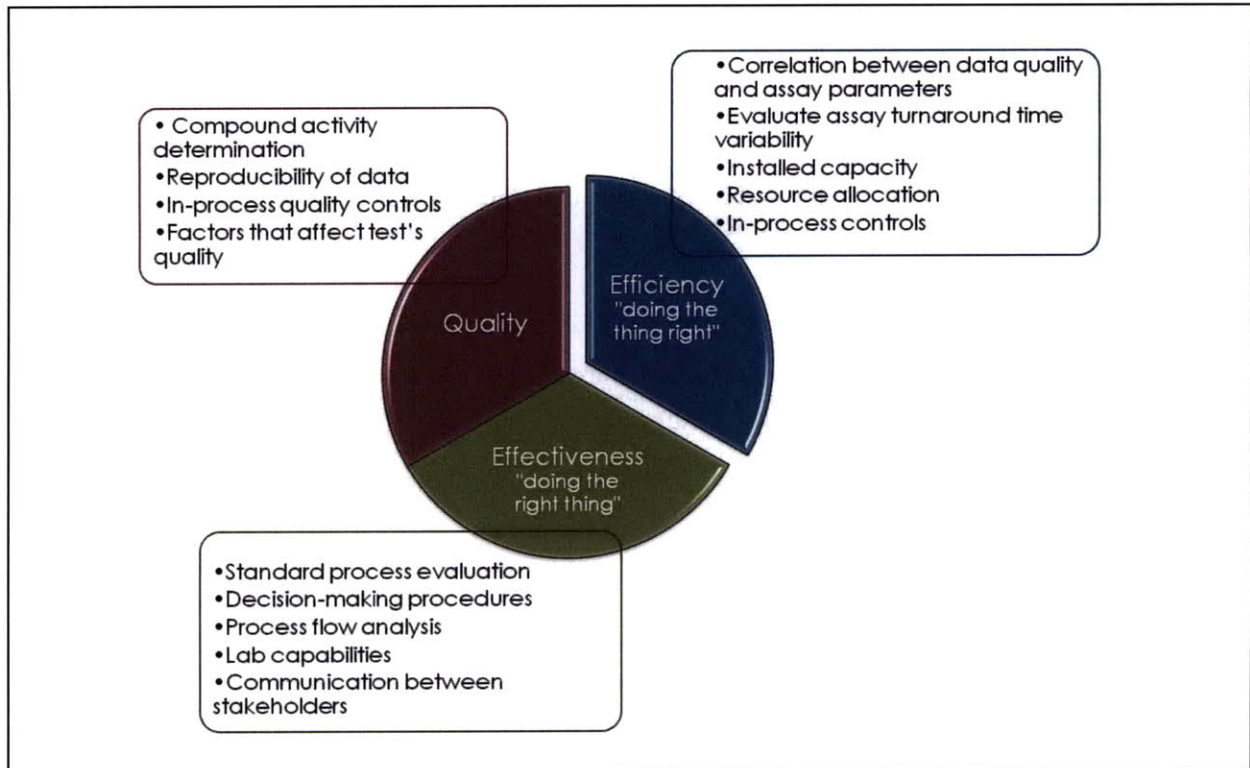
Figure 8. Diagram of the three main areas covered by the project

The project approach can be divided into two main routes: Process Analysis and Data Analysis. The first seeks to define a detailed description of the current activities while the second provides an understanding of historical results and observed process behavior. With the interconnection between procedures and resulting data, root causes can be revealed and important areas for improvement can be established. Figure 9 broadly describes the mentioned routes and the tools to be utilized in each case.
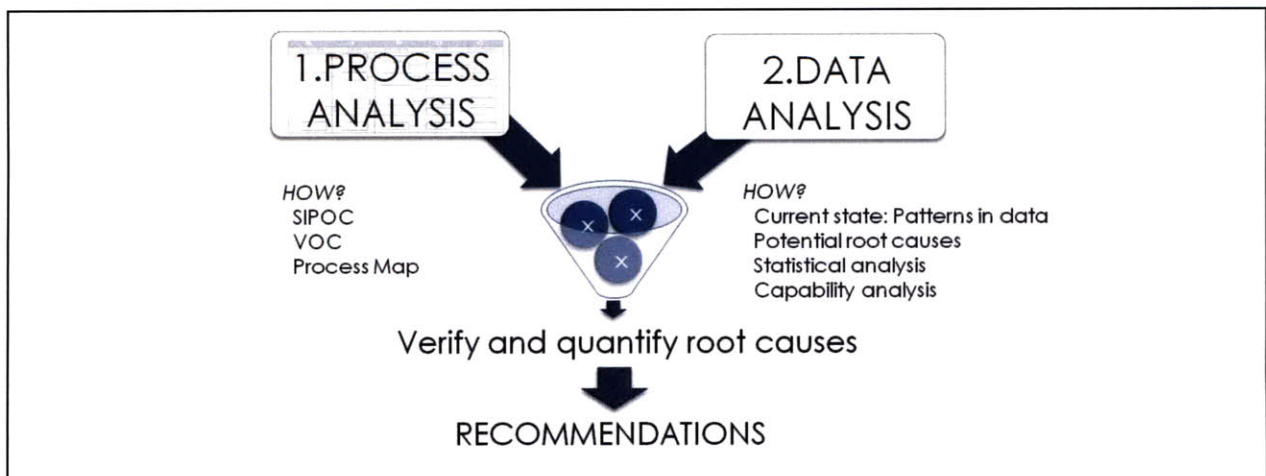
Figure 9. Diagram of project approach showing the two main stream focuses

4. PROCESS ANALYSIS

Although Novartis is strategically separated in functional units, i.e. research, development, marketing, and finance, significant activities and decision-making processes occur in their interface. Collaborations, cross-unit teamwork, and knowledge-based assessments underline a continuous flow of information across departmental borders.

Given that information is the most important outcome in every activity within drug discovery, the aim of this chapter is to apply a process-based view to EPP's activities; understand the processes, fully examine the transfer of information, and define the role of process flow, organizational structure and information systems towards efficient research metrics[25]. If compared to Novartis' IQP program, this chapter covers the Scope and Seek phases.

4.1. Process Mapping

The first step towards identifying opportunities is to completely understand the process steps and their dynamics. Process mapping is a visual representation of workflow activities, and is applied to generate a real view of how steps take place and stakeholders interact.

Although different types of maps and diagrams have been developed and used while optimizing processes, it has been concluded that the outcome is independent of the method used, if real process knowledge and critical judgment are developed. In other words, the real impact is achieved when the process analysis creates a deep understanding of current gaps and shows clear opportunities for improvement. For this reason, process mapping for EPP is developed in three steps, gradually adding more information to create an accurate picture of the current flow. With an accurate current picture, research dynamics are compared to the optimal state and gaps are further assessed.

The first representation is shown in Figure 10, corresponding to the H2L and LO loop diagram when the Chemistry Lab produces new compounds. The diagram resembles what is known as High-Level Process Map or Relationship Map, which, by definition, shows the interface between internal and external process stakeholders[26].

---

[25] (Rosenfield, 2008)
[26] (Rosenfield, 2008)
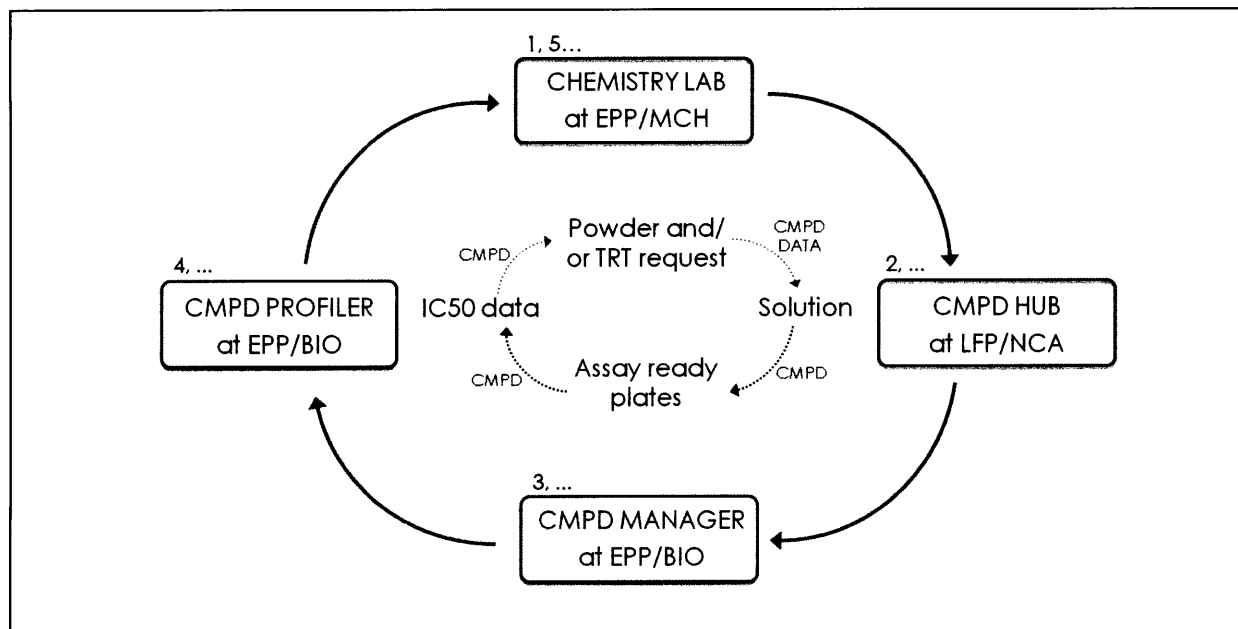
Figure 10. General description of H2L and LO loop in EPP and outcomes of each step

Let's assume that chemistry has synthesized a new chemical compound intended to develop a novel therapeutic. As shown, the loop starts when MCH produce the new compound that needs to be tested against a particular disease target. The compound, most of the times produced as a powder, is delivered to the NCA and required experiments are entered in the ordering system called *Test Request Tool (TRT)*.

Once in the NCA, a copy of the compound is produced and a vial with a standard dilution is delivered to each laboratory specified in the chemist's request. As of December 2009, the NCA was delivering one vial for each test requested by the chemists, even if the tests were performed in the same laboratory.

Each vial that arrives to the EPP/Biology laboratory is picked up by the *Compound Manager (CM)*, who sorts the compounds and prepares the plates with compound's dose-response curves (DRCs). The outcome of this process is a plate containing a serial dilution of the compounds to be tested, including the reference compounds that correspond to each experiment.

When ready, the plates are distributed to the *Compound Profilers (CPs)*, responsible for determining compound inhibition (expressed as IC50 value) towards the target(s) requested by the chemist. After experimental data is generated, results (IC50 values) are provided to the chemists.

As the final step, chemists analyze compound data by relating it with the compound's structure and comparing obtained behavior with related compounds. This compound relation is critical for knowledge

building and will define the next compound to be tested. After this step, new chemical structures are defined and the loop starts over again. Iteration will continue with the framework showed in Figure 10, until a candidate that satisfies decisive pre-defined criteria is discovered.

The next tool for process analysis is the Cross-Functional Map (comparable to a Deployment Flowchart). This map shows the process flow and identifies who is in charge of each activity and when they take place[27]. Cross-Functional Map for EPP's discovery process is shown in Figure 11.
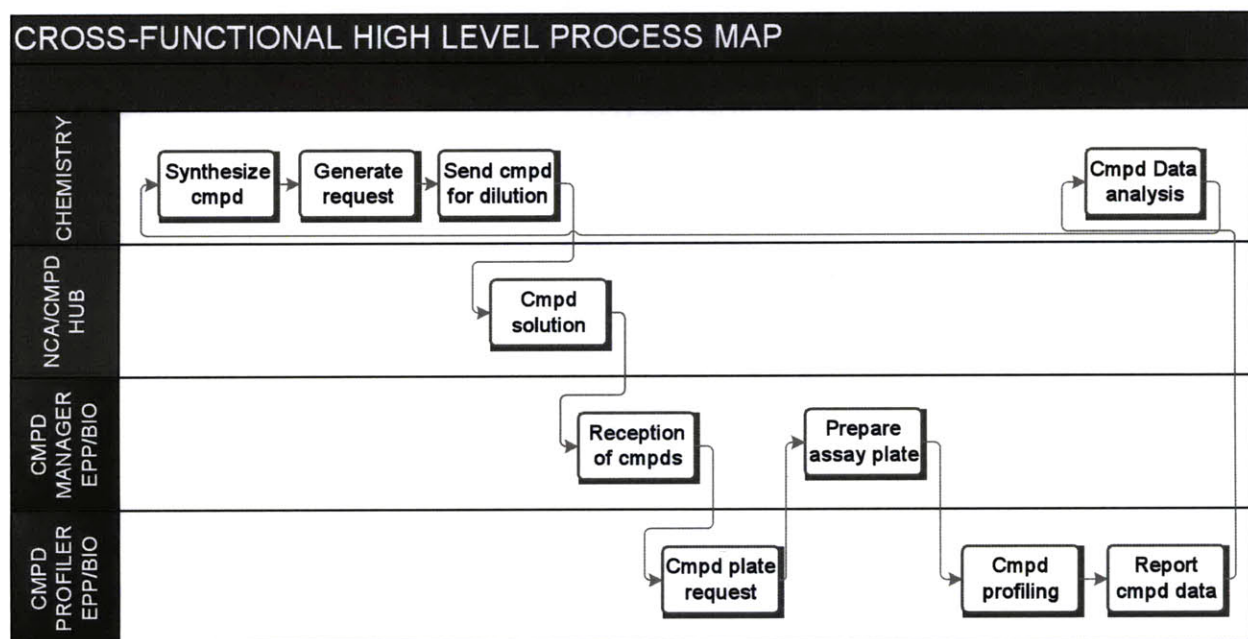
Figure 11. High Level Cross-functional Process Map for H2L and LO phases in EPP

The interest in applying this tool is to facilitate discovering who is performing which task, when, where and how long. From the figure above, it is observed:

1. Chemistry starts and closes one loop, defines the number of iterations and relies completely on EPP/Biology's output to define the next steps.

2. CM's activities are constrained by CP's scheduling and are the ones with the key process of tracking compound position in each well of the assay plate. This compound position is linked to final results and the correctness of upcoming steps relies on the correct track of this position.

3. CPs has the responsibility of performing the assay for each compound under the conditions stated by the protocol, and submitting the final data to the chemists. Systematic and detailed analysis of the potency obtained is critical to ensure data reliability. This data will impact directly the decision-making process of

the medicinal chemists.

Now, building on the process knowledge acquired until this point, the SIPOC map is also completed and shown in Figure 12. SIPOC stands for Suppliers-Inputs-Process-Outcome and Customers and not only describes general activities as in the Cross-Functional map presented in Figure 6, but also shows what are the inputs needed for the process, what do each step generates, and how every output affects and determines the performance of the following activity.

S	I	P	O	C
SUPPLIER	**INPUT**	**PROCESS**	**OUTPUT**	**CUSTOMER**
1 Chemical companies or internal chemical inventory	Chemicals for compound synthesis	Synthesis of compounds → Generation of cmpd data and assay request	1-Cmpd powder 2-TRT request	Hub in NCA
2 Chemistry	Vial with compound powder and a request in TRT	Dilute the compound based on the request	Vial with compound dilution	EPP /Biology lab
3 Compound Hub (NCA)	Compound solution	Compound dilution and copies for all the panels *based on CP request*	Compound dilution in plate for profiling	Compound profiler in biology lab
4 Compound manager in Biology lab	TRT request from chemist	*Write request to CM for cmpds that will be tested*	*Excel sheet with cmpd request and info about assay*	Compound profiler
	Plate with compounds to be tested	Prepare substrate and enzyme plates and profile compounds	IC50 value of every compound	MCH lab that submitted the compound
5 Compound profiler in Biology lab	IC50 data	Data analysis and design of new compounds	Synthesis path for new compounds	Chemistry lab associates

Figure 12. SIPOC Map for Hit-to-Lead and Lead optimization phases of drug discovery in EPP

The main reason for utilizing SIPOC maps is that they help understanding the relationship and dependence between the EPP platform and other groups involved in the discovery loop under evaluation. Consequently, the SIPOC map provides a framework to observe specific activities in EPP and relate them to the up and downstream procedures. In this way, the SIPOC map sketches what are the up and downstream needs for every stakeholder and help to recognize what will be the impact in downstream

stakeholder's procedures if any internal process in changed. In the same way, it also allows to look at the changes that have to occur in the upstream processes to achieve improvement on internal procedures.

From this plot, it is observed again the disruption CP's scheduling has in CM's process flow: CMs will generate plate copies only if the compounds were previously requested by the CP responsible of the assay. As presented, CMs will consider scheduled CP's assays after realizing all the preparation of the assay plates. If the profiling is not scheduled, prepared assay plates will be stored until the request is generated.

At this point, information regarding stakeholder's interaction through the discovery loop and general procedures in the lab has been obtained. Now, attention is shifted towards expansion of the EPP/Biology lab's activities to fully understand the current state and its dynamics. For this, a Detailed Process Map is developed.

By definition, this type of map allows process evaluation and potential root cause identification[28]. Figure 13 and Figure 14 show the Medium Detailed Maps for CM and CP's summarized activities. Appendix A through D show all steps inherent of each summarized activity and Appendix E shows the shape's key used in the maps.

Particular observations from the diagrams are:

▪ Technology has been upgraded and processes changed mainly as a way to further decrease compound profiling times. Limitations arise when continuity of changes blend with lack of analysis of their impact giving a wrong perception of real obtained benefits. In addition, time for the learning curve to stabilize has not been provided, resulting in the observed pile of steps that limit fluent workflow and divert personnel from standard methods.

▪ Despite of having great information technology databases for the company as a whole, the EPP Unit has mostly developed their own communication systems. Because of convenience and straight-forward use, Excel has been widely used as the tool for data storage and transfer of information between individuals in the lab. Satisfactory short-term benefits were observed until the amount of data started to reveal Excel's limited capabilities when used as a database. Additionally, since Excel files and macros are developed internally, responsibility for support, modification, and replacement relies on the same lab resources, creating also delays in needed upgrades.

---

[28] (Rosenfield, 2008)

- From process flow observation it is clear how the number of requests received in a particular day, the total number of compounds tested, and the CP's scheduling procedure are approached as independent steps. Therefore, even when the lab capacity allows for a smooth process flow, lack of coordination creates oscillation of turnaround time and unsteady workload among lab associates.
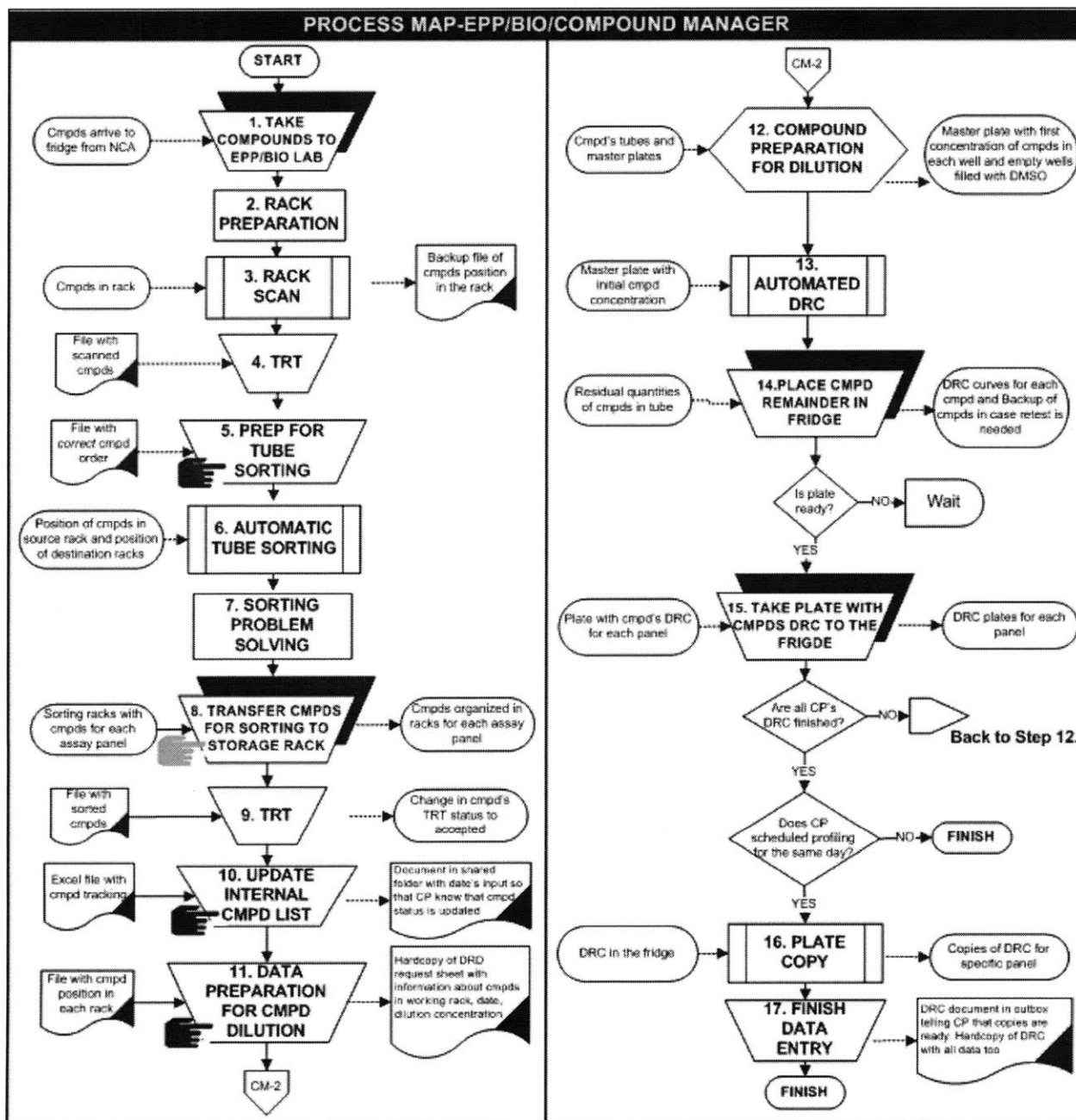
Figure 13. Medium Detail Process Map for Compound Manager
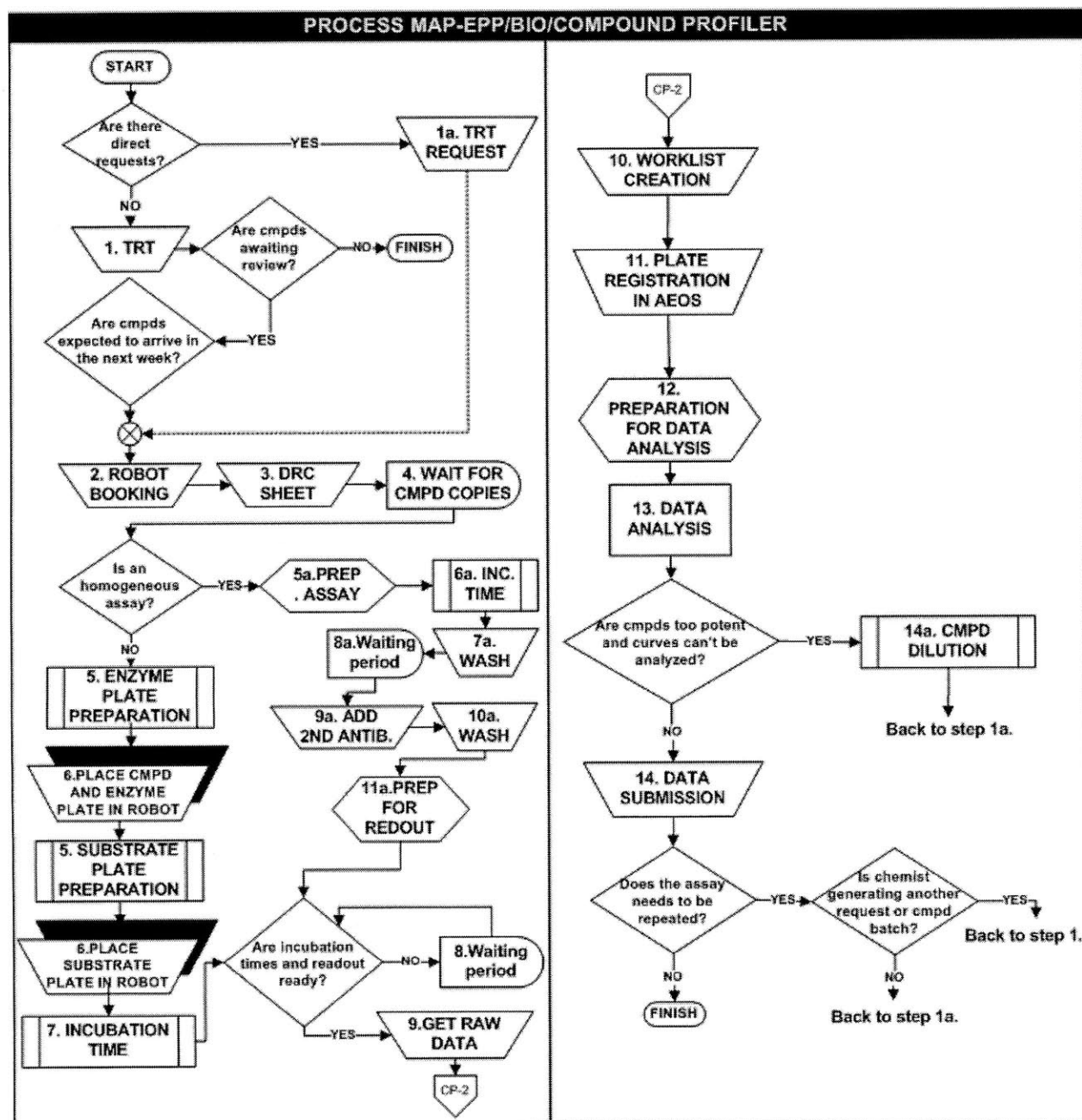
Figure 14. Medium Detail Process Map for Compound Profiler

▪ Focus on achieving low turnaround times seems to be the main limitation for process upgrade. Lack of a complete set of performance indicators that align stakeholder's incentives towards a common objective, is created by the individualized focus of daily work. Divided efforts decrease teamwork, not only among internal laboratory groups but mostly between key players along the H2L and LO iteration. This generates the observed unnecessary file storage (*just in case* backup), addition of non-value added activities to speed up the process, and non-standard procedures. Unclear bottlenecks and milestones are the bottom-line result.

After analyzing the current flow at the EPP/Biology labs, it is evident that root-cause and value-added activities need to be addressed. Knowing the lack of performance indicators and observing its effect on profiling methodologies, the next phase concentrates on defining a common framework for project evaluation that connects all different incentives into a common measurement system.

4.2. Voice of the Customer and KPI identification

It has been a topic of great discussion among crucial stakeholders whether the term customer is applicable to the drug discovery, and if the use of this expression will have an adverse impact on achieved teamwork. Even when collaboration is reinforced through all projects in EPP, at the end of the day chemists have the task to pull all the data together and will be the main judges when deciding future compounds for analysis. Also, chemistry is the final stakeholder in the analyzed drug discovery loop. Therefore, in this study, chemistry is considered the final customer of the EPP/Biology laboratory and all the tools that Lean, Six Sigma, and IQP provide for customer satisfaction are applied to the chemists.

The first tool to be utilized is Voice of the Customer (VOC). VOC is a set of tools that permits specific understanding of what the customer wants and translates it into measurable indicators, often called Critical-to-Quality (CTQ) variables[29]. In general, VOC is used to understand the customer's (chemist) needs when performing their internal processes, and define the gaps between the current and the ideal delivery. In addition, VOC is utilized to evaluate the lab's flexibility to adapt to the changing environment in the chemistry unit[30]. Lastly, to ensure sustainability in the future, when VOC is applied in this project, indicators are evaluated against EPP Unit's strategy requirements. In this way, when selected, KPIs are not a collection of ideal chemist's outcomes, but a list of measurable and impactful indicators that will help EPP achieve the platform's strategic mission.

VOC determination for EPP is based on several interviews with specific groups, including project team heads, key chemistry laboratory heads, and lab associates. This method is commonly referred to as *proactive data collection[31]*.

During the discussions, three main observations always came to the table:

1. Why in some circumstances if one particular compound is tested at different periods of time the IC50 result can vary significantly?

[29] (IQP Novartis website)
[30] (Army Business Transformation Knowledge Center, 2009)
[31] (IQP Novartis website)

2. Why when *several compounds* are requested for *the same profiling assay* on occasion results are submitted with considerably different times?

3. Why when *one profiling assay* is requested for *several compounds* on occasion results are submitted with considerably different times?

Complete evaluation of chemist's concerns and their classification based on the approach are shown in Table 2. As observed from the summary, all the issues listed in every category (quality, effectiveness, and efficiency) have an impact on basically the same results, like profiling yield, turnaround time, and data quality. This aspect shows hoe meaningful results can be achieved if improvements are focused on these specific areas. Also, this frames the basis for the measurement system.

The next step compares all the issues expressed by the chemists with every activity graphed in the process maps. This procedure helps finding the causes that are impacting, directly or indirectly, the results perceived by the chemists. These causes, known as *Customer Key Issues*, reflect critical specifications that characterize the alignment between EPP's laboratory deliverables and chemist's decision-making process. This step helps decide where improvement efforts should be focused, what the key drivers for *customer satisfaction* are, and what the baseline for a holistic measurement system is[32].

The first concern listed is the difference in inhibition values. The principal observation here is the lack of a specific threshold that defines an acceptance range for a compound value. Because of unpredictability of compound's potency and uncertainty in discovery, many argue that a rigid threshold cannot be stated. With the aim of explaining the importance of setting an acceptance range for compound's potency, here is a possible scenario: Let's assume that one of the teams is focused on Project A, and that compounds are separated in groups based on their scaffolds. Now, assume that today compounds from every group are requested for profiling. One of these compounds is compound C, from group 1. After d days, results are submitted and the compound's C potency value is X.

After several compound iterations and structure analysis, the project team selects the scaffold from group 2 to continue in further discovery phases. Following numerous tests, finally the project reaches D3 phase (lead optimization), when selectivity and efficacy of compounds in animals starts to have the most significance. Results in this phase turned out not to be as promising as expected, and the project team had to go back to the first generated data to select a second series compound. Let's assume that the second best group is group 1, and already tested compound C is now selected to be the one in further research. Compound C is requested by the chemists for confirmatory tests and the result is an AC50 15 times

---

[32] (Army Business Transformation Knowledge Center, 2009)

higher than. At this point, debate about assay accuracy and test repetition begins.

Table 2. Relation of VOC issues with observations from the Process Maps

ISSUE	IMPACT	RESULT
QUALITY		
1. Chemists have different definition	- Difficulty for Bio Lab to set KPI -Change in project team set different expectations from lab	-Repetition by MCH - Dissimilar threshold's selection for active cmpds[33] and unequal criteria in the lab
2. Unequal criteria used in the lab	-Heterogeneous data quality -Misalignment of required principles -Change in project team/profiler gives data variation	-Subjective curve fitting -Repetition by MCH/profiler
3. No report of minor profiling modifications	-Decrease in assay data reliability - Cmpd selection can be done in different/ incomparable conditions	-Repetition by MCH -Incorrect thresholds for active cmpds
EFFECTIVENESS		
1. Upgrading in technology not in process analysis	-Investments for improving possible constraints that can be addressed by process variations - Processes are modified to match machine procedures →decreases process standardization - Lack of real bottleneck determination	-Overcomplicating simple processes - Excel sheets and macros become a "need" for profiling - Overcapacity in the lab - Unclear installed capacity and capabilities
2. Process modification is performed individually	-Missing opportunity for best practices - Process owned by individuals → reduces team work and - Limited empowerment for improvement	-Difficult interactions among the lab -Nonstandard procedures - Non-optimized processes - Uncertainty in lab capacity/capabilities
EFFICIENCY		
1. Expected IC50 range not shared	- Assay repetition for active cmpds	-Increase in data delivery time -Imprecise lab turnaround time -Decreases plate optimization → min # of cmpds used in the repeated plate -Assay time variable/unpredictable
2. No centralized info/limited automation	-Excel sheets with macros developed to communicate and get data - Copy/paste and manual record is needed - Time in the lab is focused on process management	-Data backup in excel files → nonstandard procedure traceability difficulties -Overcomplicating internal processes/need of "macro expertise" - Influence in lab interactions -Turnaround time variability/predictability
3. Non-standard processes	-Difficulty to keep track of lab activities/process improvement - Any variation in responsibilities impact process and discrepancy on expectations	-Turnaround time variability/predictability -Increase process complexity → cmpd manager, data for chemist

In order to define which data point is the accurate one, compound C is tested a third time, so that, if the

---

[33] Cmpd=chemical compound

third test shows an IC50 of 10X, the team will go with the fact that the compound has an inhibition in the 10X to15X range. But, if the third assay result shows an IC50 of 5X, the project team will define the real value of compound C to be in the X to 5X range.

The example's value is to reveal the importance of potency data correctness: an incorrect data point will lead not only to wrong decisions, but also to an increase in the non-value added compound discovery time, and, as a result, in an increment of research costs. Therefore, it is recommended that a set of tools is defined to evaluate data quality and reveal accuracy of profiling processes. One useful method is the definition of an assay's repeatability range through statistical analysis of the reference compound's data.

Another observation when applying the VOC tool refers to turnaround time variability and its impact in chemistry processes speed. If a chemist wants to test two compounds that belong to the same project, most of the time they will pause project activities until both results are received. The same situation occurs when chemists need the same compound to be tested in several assays for selectivity analysis. Most of the time, no substantial progress occurs until all data is gathered together. Because chemists support their decisions in laboratory results, variability in the time lab results are submitted causes variability in the time chemists provide the next series of compounds, amplifying further variability in the overall project length.

Bringing all back to the VOC tool, it can be said that discovery time length is described mainly by two variables: length of one iteration and number of iterations. If phases after D3 remain unchanged, time-to-market can be decreased with improvements in discovery, when the number of iteration loops and/or the time it takes for a single loop to be completed are reduced. Length of iteration is defined by the speed each of the individual tasks can provide, while the number of iterations is affected by the assay reproducibility and repeatability. These are the results of the VOC and the baseline for KPI determination. Figure 15 shows a diagram of the discussed results.
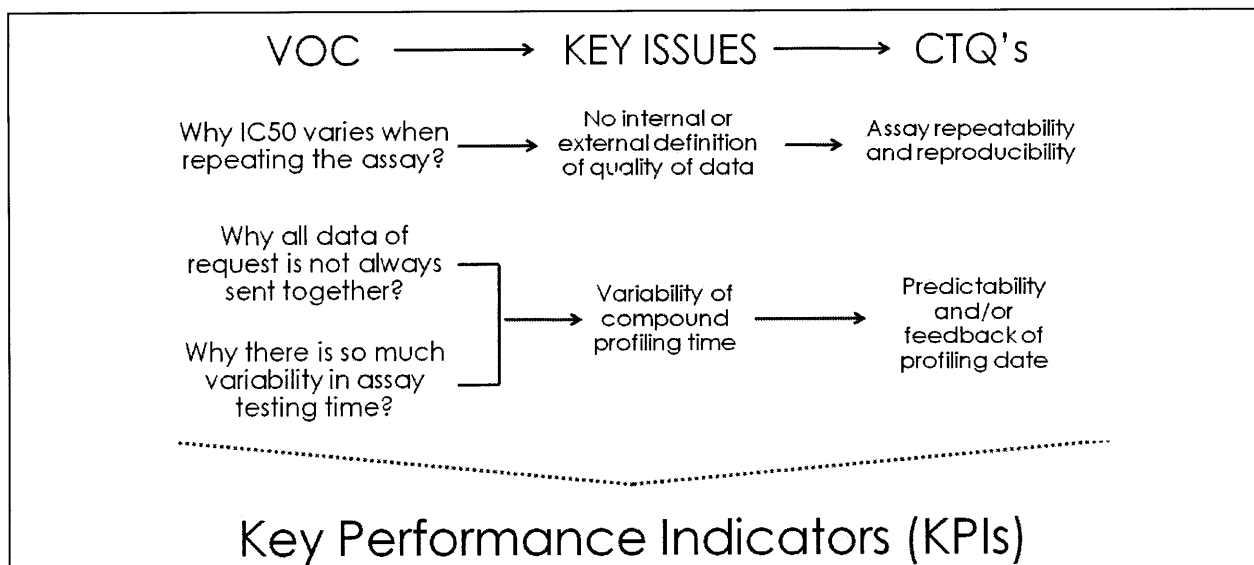
Figure 15. Voice of the customer (VOC)

4.3. Measurement System determination

Key Performance Indicators (KPIs) are specifications or measurements critical for business success, characterized for being quantifiable, measurable and result-oriented. At a corporate level, KPIs align any process with overall company goals, while at a unit level they reflect the objectives that describe the activities of a particular unit[34]. The main advantages of having performance indicators is that they allow process owners to monitor day-to-day tasks, promptly observe unexpected results, and take immediate actions towards process constraints, having an instant and continuous influence on personnel actions[35].

The aim of this section is to define a set of variables that allows laboratory associates to focus on simple day-to-day goals while having a direct impact in overall platform performance. Tying realized VOC analysis back to Novartis corporate strategy -deliver novel therapeutics in the shortest time possible- KPIs for EPP should combine profiling speed with data reliability in order to create a comprehensive measurement system.

The EPP Biology lab is currently using a modification of a program that was developed for High Throughput Screening (HTS). Because of the number of compounds profiled and the nature of the process, HTS deals with large number of datasets and provides many variables to expedite observation of potent compounds. A total of 100 available indicators are displayed for the associate. Because EPP has a more dedicated approach, more than 50 out of the 100 available indicators are not presently defined in the

---

[34] (IBM)

[35] (Drug discovery: are productivity metrics inhibiting motivation and creativity?, November, 2008)

system or calculated for the user. This large selection of variables has promoted the use of different quality standards for evaluating assay results, creating an actual obstacle for standard quality procedures. Thus, measurement system definition looks at the set of variables presented by the software and analyses them from a data quality and robustness perspective, not depending on the current usage and/or acceptance among the personnel.

While evaluating the list of metrics displayed by the software, attention needs to be focused on the detail of the measurement. As commented by Boutellier and Ullman[36] and shown in Figure 16, the level of detail an indicator has is directly related to the level of motivation it creates in the individuals. As observed, motivation is highest when the level of detail in the indicator has a medium level. This means that associates are highly motivated if they have clear goals that show their efforts but still have room for innovation, risk taking and perform their own process management.
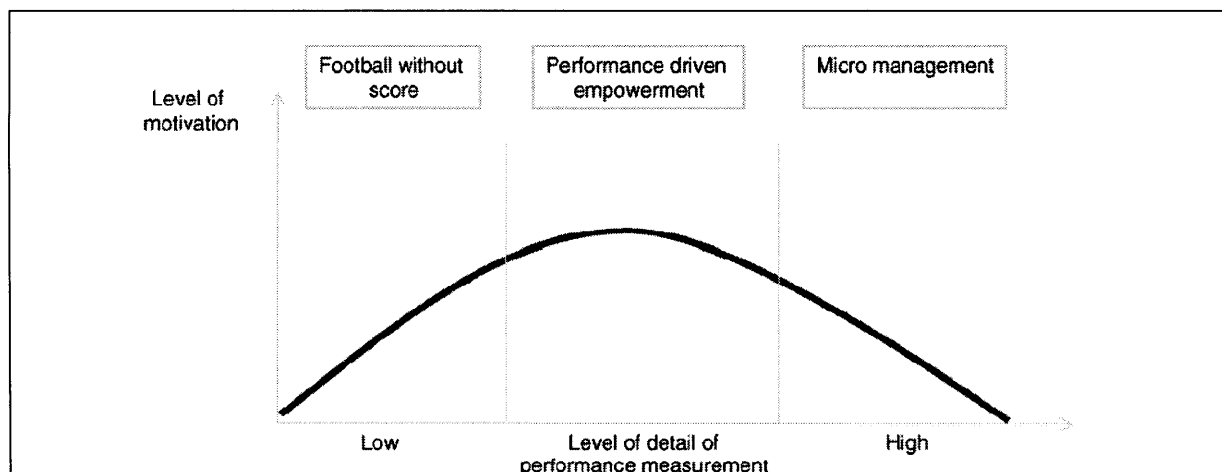
Figure 16. Representation of the level of motivation and the level of detail the KPIs should have, given the type of process is being analyzed[37]

The same authors state that core activities performed in rational drug discovery, scaffold analysis and DMPK, can be categorized in a transition phase from a product to process innovation. Consequently, they ensure that, if the indicator used to measure performance has an extreme level of detail, there is limited room for an individual's initiative and the process arrangement can kill innovation. Boutellier and Ullman's recommendation in this case is to develop a measurement system that combines process control with individual's autonomy enabling also a higher level of motivation.

---

[36] (Boutellier, A case study of lean drug discovery: from project driven research to innovation studios and process factories June, 2008)
[37] (Adapted from Boutellier, 2008)

Along the same lines, another case study analyzed by Boutellier and Ullman suggests that streamlining activities achieve better results when the organization is process oriented. The four step approach presented by the authors and shown in Figure 17 mirrors EPP's unit goals. This provides the second hint towards defining EPP's KPIs: if the aim is to construct reliable operations, selected indicators should combine profiling data quality with specific speed metrics.
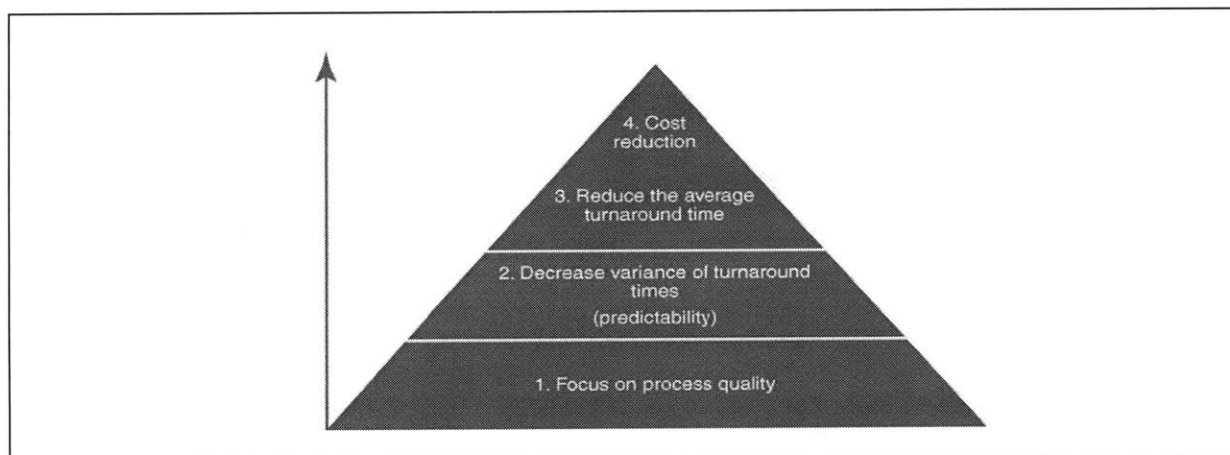
Figure 17. Representation of the steps to create a process oriented organization[38]

Using the above information, an iterative procedure is conducted to create a robust and outlier resistant statistical method that can be monitored for quick identification of experimental problems[39]:

1. Interview the Medicinal Chemists to review CTQ's outcome and define which measurements are understood by the team.

2. Review selected measures and evaluate how each of them explains defined CTQs.

3. Assess ability of selected variables to link different stakeholder's incentives and consider its robustness and significance in drug research activities.

4. Examine the variables with lab associates and lab heads to find applicability of indicators in day-to-day operations.

5. Take chosen variables to chemists for a re-check, obtaining defined in step 2. Steps 2-5 are repeated until general agreement defines ultimate indicators to be used.

6. After a series of interviews with all stakeholders, initial goals were created for each one of the selected KPIs.

---

[38] (Adapted from Boutellier, 2008)
[39] (Gubler, February, 2007)

Final selected KPIs are separated into two major categories: speed and quality. Figure 18 outlines each selected indicator into these categories, showing the variables selected for each category and the basic framework forthcoming analysis will be based on.
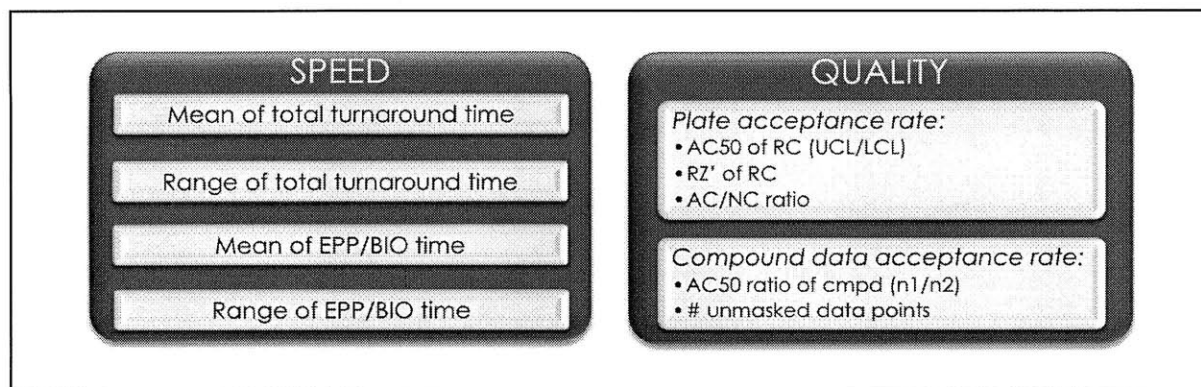
Figure 18. KPIs selected for EPP's process analysis and performance measurement

A more detailed description of the KPIs is presented in Table 3. In here, definition of each indicator, units of measure, and calculation procedure are expressed. In addition, each independent variable is identified by a letter and a number. The codifications will be used in data analysis when relating the outcome of the process, KPIs, with input variables, X (using the analogy to a function, where y is the result and x the input: $y = f(x)$). It is important to mention that the goals from Table 3 illustrate the ideal state (mostly for measurements of quality that hasn't been tracked this way before) and refer to First Pass Yield (FPY) for indicators different than time. In other words, a 100% of Reference Compound AC50 means that a 100% of the data submitted to chemists should have reproducibility of Reference Compounds data (second column), but internally, the goal is that 95% of Reference Compound's AC50 is within the defined control limits. If not, the assay should be repeated in order to achieve the goal of 100% of delivered data with acceptable Reference Compound AC50.

Table 3. Description of Key Performance Indicators

KPI	DESCRIPTION	UNITS**	CALCULATION	GOALS Internal /	
Turnaround time	*Time elapsed from beginning to end of the activities and measures the total length of a cycle. Total time starts when a chemist generates a request and finishes when assay data is published. For EPP time, cycle starts when the compound is received in the biology laboratory and finishes when the data is published*	*days*	-	-	
Mean total turnaround time (Y1)	These times refer to the average of the times in a particular timeframe	days	$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x$$	12 days	
Mean EPP turnaround time (Y2)				10 days	
Range total turnaround time (Y3)	Range is calculated by subtracting the highest time observed for a given period from the shortest time registered for that same period	days	$t_{max,i}$-$t_{min,i}$ where t=time, i=period eg.month	22 days	
Range EPP turnaround time (Y4)				20 days	
Plate acceptance rate	*Shows the trends for indicators of the plate as an unit, guiding the decision of repeating the profiling of all the compounds in that particular plate*	%	$\prod_{i=5}^{7}(\%Y_i)$	90%	98%
AC50 of reference compound (Y5)	AC50 value of the Reference Compound, which should be within the control limits, taking all the data generated from past assays. Reference compounds AC50 is previously defined and evaluated in assay development and repeatability and reproducibility are already known	%	*	96% 100% within 3 standard deviations of the expected value	
RZ' of reference compound (Y6)	Robust Z' gives some information about typical data scatter and high-low band separation, which is also relevant for the curve analysis. This outlier resistant indicator is defined as the screening window coefficient being a function of the ratio of the signal window and the signal range	dl	$$RZ' = 1 - \frac{3(mad_{max} + mad_{min})}{|median_{max} - median_{min}|}$$	95% 98% of compound's curve with RZ'≥0.6	
Active Control and Negative Control ratio (Y7)	Ratio of the AC and NC values to guarantee that the analysis is done under linear range	dl	$AC50_{AC}/AC50_{NC}$	99% 100% of compound's curve with a ratio of X	
Compound acceptance rate	*Relates specific values for a particular profiled compound that guide the decision of repeating the AC50 determination for that specific compound*	%	$\prod_{i=5}^{7}(\%Y_i)$	90%	100%
AC50 compound ratio (Y8)	Each compound AC50 is determined by duplicates (n1 and n2), so Y9 is the ratio of those two AC50 values	dl	$AC50C_{i,n1}/AC50C_{i,n2}$ where Ci=cmpd i and n1 and n2 are AC50 cmpd i measurements	95% 100% of the ratio between n1 and n2 are within 5	
Number unmasked data points (Y9)	Automatic calculation that considers the number of data points that are manually eliminated from the fitting curve and the spread of the data. Defines a contribution weight of each data point and then calculates the total number of data being used (not necessary an integer)	#	*	95% 100% of each compound's AC50 is defined by curves with more than 7 data points	

* variables determine by software
** dl=dimensionless

4.4. Understanding and defining the current state

Even with clearly defined metric systems, sustainability of changes can only be achieved when people become aware of the reasons for current behavior, so that limitations are surmounted. A broadly used tool for root cause evaluation is System Losses, where the process is scrutinized looking for 8 different types of wastes or non-value added activities[40]: defects or rework, overproduction, over-processing, waiting time, inventory, transportation, motion, and underutilization of people. Table 4 shows a summary of the wastes in evaluation and common causes for their existence. Often times, other wastes such as inflexibility and variability are also considered when realizing process evaluation.

Table 4. Type, definition and causes of the 8 waste types defined by Lean

WASTE	DEFINITION	COMMON CAUSES
Defects/ Rework	Work that contains an error or is not "right" and needs to be fixed	Processes are not correctly specified or people are not trained correctly
Over processing	Extra usage of activities to double check but not adding real value to the final outcome	Non standardize process makes it necessary to double check Over processing makes a backup in case something goes wrong Redundant information
Over-production	Redundant work or too much generation of a product too early in time	Multiple forms of the same information Quantities not clearly defined Process capacity is not known Poor communication between stakeholders
Waiting time	People waiting for something needed to continue or start their own activities	Process steps are not balanced Poor coordination between process parts Lack of understanding of the time it takes for each step Un responsiveness of scheduling systems
Inventory	More materials on-hand than the ones needed to perform any activity in a given time	Products or materials are ordered in more amount "just-in-case" Process is noisy or volatile so inventory is needed to deal with unpredictability of activities
Transportation Motion	Materials or people moving along the process	Layout is not correct Multiple unnecessary activities or materials are needed from multiple places Lack of standardization makes unclear what materials will be needed until process starts
Under-utilization of people	People are not utilized for improvement or value added activities	Associates do not have the authority needed to decide over process upgrades and direct decision making

Utilizing the process maps for the CM and the CP, each activity can be evaluated in terms of the types of wastes. Figure 19 and Figure 20 summarize the wastes discovered in each of the general activities. As presented, wastes illustrated in each box correspond to all the wastes observed for the sub-activities represent

---

[40] (Helping Science to succeed: improving processes in R&D, 2008)

by the main one. In this analysis, over-processing waste is used to explain processes that shouldn't be present or processes that are a repetition of some other process (i.e. saving documents already available or re-checking data).
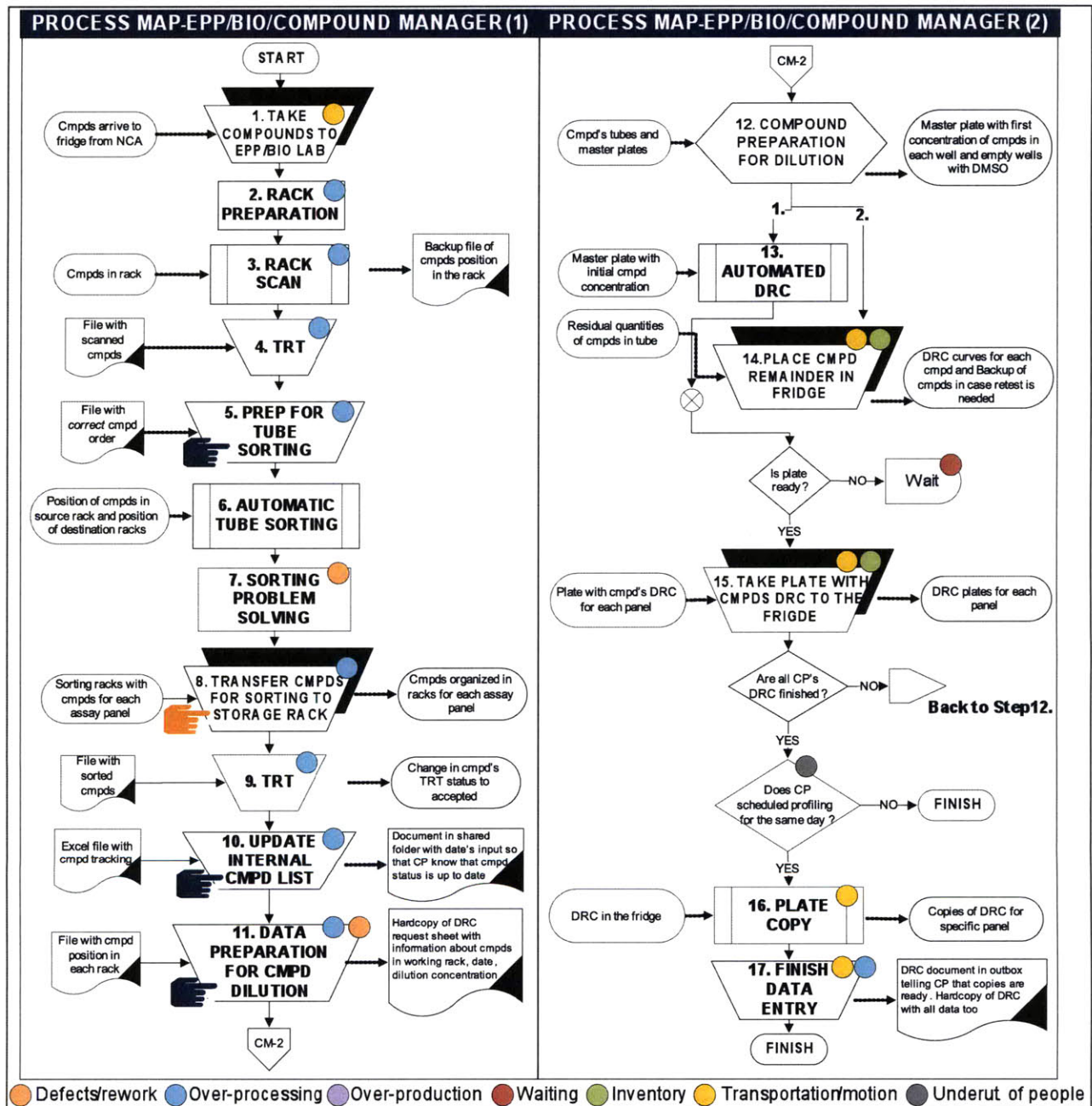
Figure 19. Types of wastes observed for each activity in Compound Manger's process
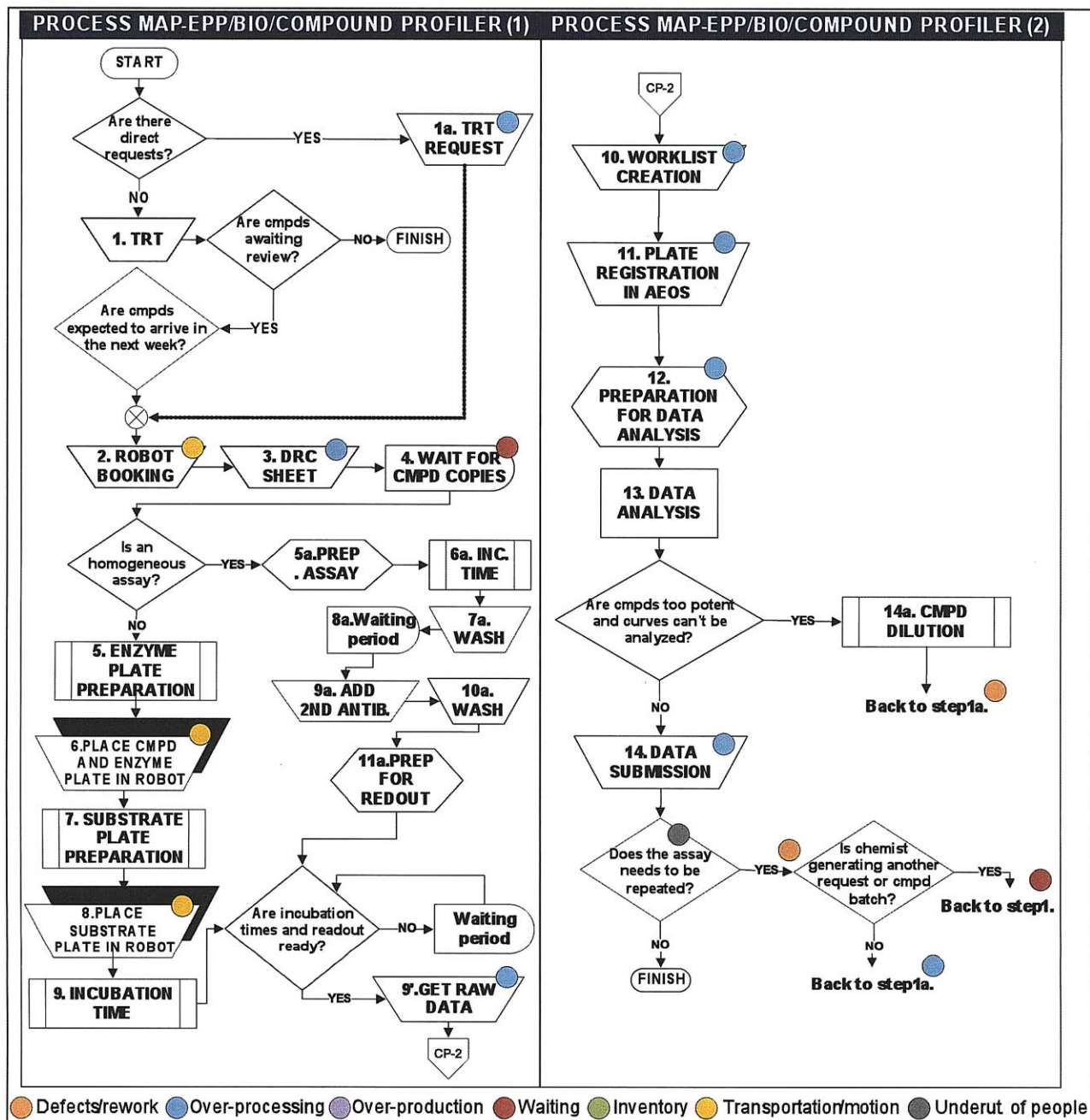
Figure 20. Types of wastes observed for each activity in Compound Profiler's process

To define the numeric impact of these non-value added activities in the overall profiling time, value-added analysis tool is implemented to the diagrams above. Steps that may seem or have been classified as system losses but are necessary to deliver the profiling data given current process, are classified as value added. Non-value added steps are definitive activities that are in place just as a response of lack of

process improvement and, if changed, will directly impact overall performance. The following diagram sketches process activities over time, splitting value added from non-value added ones for a clear study.

COMPOUND MANAGER

TIME (mins/plate of 14 cmpds)

Activity	Prob.	TOTAL	VA	NVA
1. Cmpds to Lab		10	5	5
2. Rack Prep.		10	-	10
3. Rack Scan		15	5	10
4.TRT		15	5	10
5. Prep. sorting		15	5	10
6. Auto sorting		5	5	-
7. Problem Solving	5%	0.8	-	0.8
8. Transfer cmpds		10	-	10
9. TRT		10	5	5
10.Update list		5	-	5
11. Data prep. dilution		15	-	15
12. Cmpd prep. dilution		15	15	-
13.Automated DRC		65	65	-
14.Remainder in fridge		5	-	5
15.Plate DRC to fridge		5	-	5
16. Plate copy	70%	10.5	7.0	3.5
17.Finish data entry		10	-	10
		221	117	104
			53%	47%

■ Non-value added   ▨ Value-added

COMPOUND PROFILER

TIME (mins/plate of 14 cmpds)

Activity	Prob.	TOTAL	VA	NVA
1.a. Direct request		5	-	5
1. TRT		5	-	5
2. Robot booking		5	5	-
3. DRC sheet		10	-	10
5. Enzyme prep	90%	13.5	14	-
6. Enzyme in robot	90%	4.5	5	-
7. Substrate prep	90%	13.5	14	-
8. Substrate in robot	90%	4.5	5	-
9. Incubation	90%	63.0	63	-
5a-11a. Het assay	10%	10.0	10	-
9'. Get raw data		5	-	5.0
10. Worklist		5	-	5.0
11. Plate reg		10	-	10.0
12. Prep. Analysis		10	-	10.0
13. Data analysis		15	15	-
14a. Cmpd repetition	20%	40.0	-	40.0
14. Data submission		10	-	10.0
Assay repetition?	15%	33.8	-	33.8
		263	129	134
			49%	51%

■ Non-value added   ▨ Value-added

Figure 21. Comparison of value and non-value added activities in Compound Manager and Compound Manager's current process

From the figures above, the main conclusions are:

1. Non-value added activities take almost half of the time needed for the CM and the CP to end one process cycle.

2. For the CM, system losses are primarily due to re-processing and the lack of a better communication method among associates.

3. For CP, system losses are mainly due to software and information technology infrastructures, the lack of associate empowerment to take decisions or review information about compound data reproducibility, and limited information flow from the chemists to better consider starting concentration when profiling.

Given that activities are not standard and procedures differ even for the same associate when performing sequential assays, time stamps show the most common pathway with an average length for any given compound. In the same way, the column of probabilities shows the proportion of the time that the activity occurs.

When performing the value-added study, it becomes clear that variability in profiling delivery times come from the lack of procedure standardization. Even with all automation already on place and software for data analysis, if the process has a high variability, profiling times won't be stable. One of the reasons for the lack of common work is the high proportion of non-value added activities for both, CP and CM. These activities deviate people's attention towards process improvement, generating instead a day-to-day focus in problem solving.

In the next chapters, data is deeply studied and related to current process observations to define root causes and main focus for immediate improvement.

5. DATA ANALYSIS

Now, based on the framework presented in Figure 9, Data Analysis is the next step after realizing the Process Analysis. This section of the project evaluates historic data from the process in order to understand trends, patterns, relationships, and, most importantly, the impact that each variable has on the overall observed result. When looking at a process and its resulting data, many variables seem to be responsible for the outcomes and a common mistake is to try to either control or change all at once. Even if several factors are involved in the process, patterns can be explained by finding the root causes that trigger the rest to behave in a certain way. Comprehension and control of these variables lead to continuous process improvement.

The first step for data analysis is to characterize the current state. To define a lab's performance and operations, assay demand is the first measure to evaluate. As shown in Figure 22, the number of assay requests has been changing throughout the months with no specific trend. As presented, the green bar corresponds to the number of tests and the red bar shows the number of requests that were rejected in a given month.
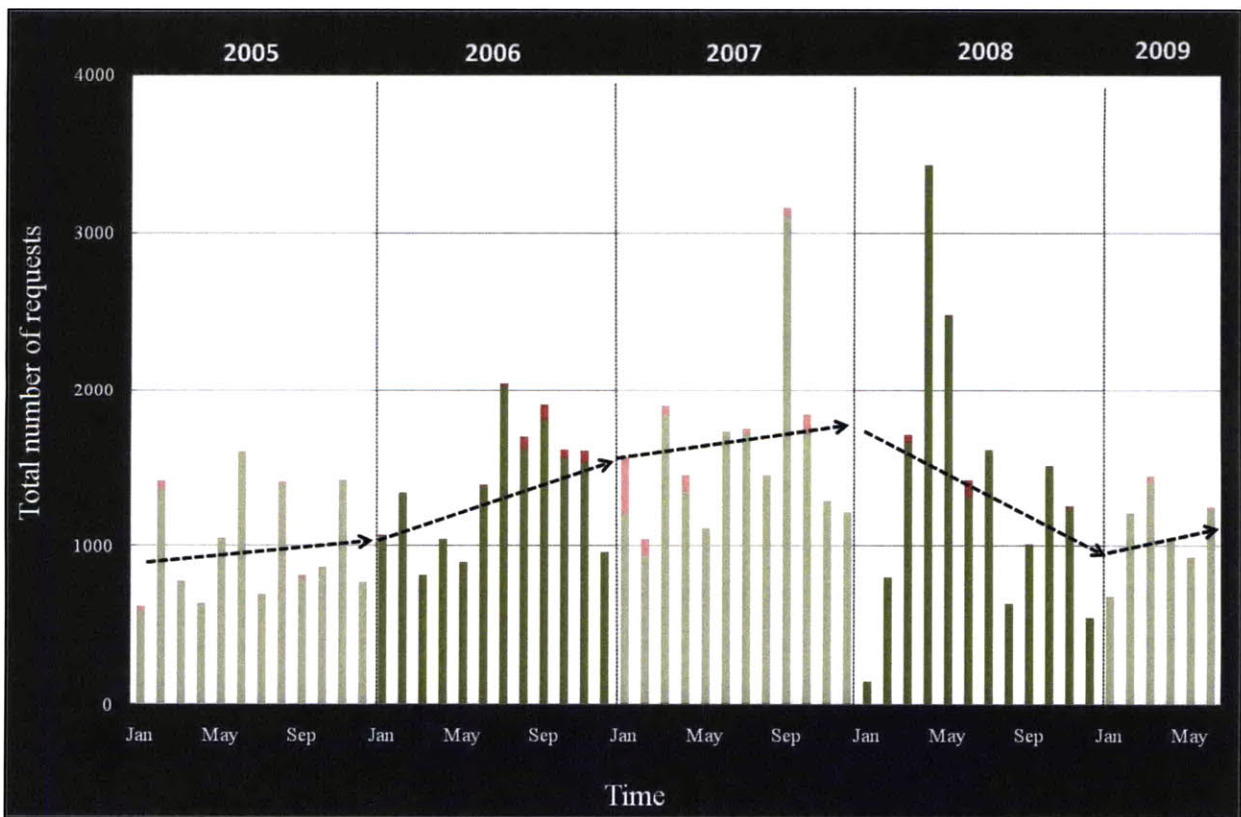
Figure 22. Total number of assay requests per month, from January 2005 to July 2009

When combining the total number of assays with the specific turnaround times for each period, trends presented in Figure 23 are obtained. Important to notice how overall speed is not directly correlated with the number of compounds tested. This fact is the main trigger for future data study. In the graph, left axis corresponds to turnaround days represented by the wide bars. Axis to the right corresponds to total compounds tested, which are represented by the skinny bars.
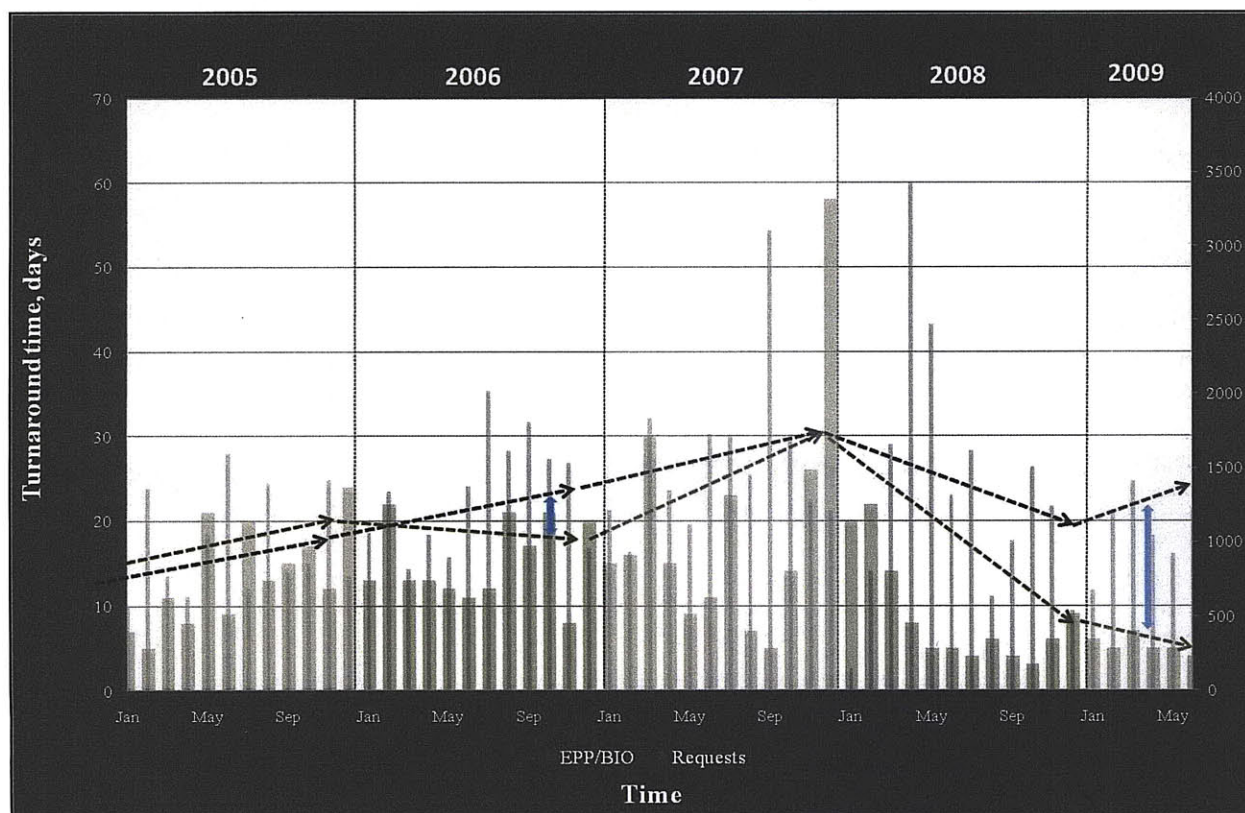
Figure 23. Combination of compound demand with turnaround times, from January 2005 to July 2009

Lastly, a summary of EPP/Biology's lab demand for 2009 is summarized in Table 5. When looking at this data and comparing it to the same analysis for previous years (not presented in this document), a key observation is that Compounds tested in different assays increase as the project advance in phase. This is a result of common practice in drug discovery, in order to test compound selectivity. Although, when observed carefully, data shows how the trend for EPP has been slightly shifted: in the years 2005 and 2006, tests for selectivity were done at later phases, mostly D3. After 2008, and as of October 2009, selectivity has been evaluated mostly in D2b. This can be observed as a positive trend that will, if pursued correctly, push forward only candidates that are selective enough, potentially decreasing the decision

making time in later phases.

Table 5. Summary of selected data for EPP during the year 2009 from January to October

Phase of the project	Total tests (#)	Total cmpds (#)	Cmpds tested in several assays (#)	Average time (days/test)	Average synthesis time (days/comp)	Average tests (assays/month)
D0	409	200	87	4.95	0.59	34
D1	370	282	52	2.74	0.65	31
D2a	1,275	584	221	7.40	0.19	106
D2b	4,687	2,370	829	7.65	0.05	391
D2b-D3	299	1,160	43	6.31	0.80	25
D3	1,909	1,112	357	6.31	0.13	159
D3-CSP	313	162	20	8.07	0.77	26
D4	9	9	0	8.89	26.67	1
DevSupport	6	6	0	9.50	40.00	1
ReferenceCpd	16	14	2	11.63	0.00	1
TOTAL	**9,293**	**4,857**		**7.34**	**6.98**	**774**

The purpose of the next sections is to evaluate each one of the Ys defined in the metric system from Table 3 and identify specific root causes (Xs) that control the overall process behavior. Given the variety of tools and combination of measures used in the process of finding root causes, all following analysis is based on specific guidelines and tools defined by Novartis IQP. The goal is to provide for each variable a comprehensive evaluation of the current state, a list of identified root causes and initiatives for improvement.

5.1. Turnaround time: Mean and range values

The first indicator in the KPI list is assay time, defined by the average duration and the range of durations for a given period. Response variables 1 to 4 (Y1, Y2, Y3 and Y4 as named in Table 3) represent the different measurements for profiling time and spread. This combination allows the Biology lab to track the specific variation of the internal profiling time without forgetting that the time the chemists perceive is the total time starting with the filing of the request until receiving compound results. To provide a structured analysis, this section is further divided in 4 main sections: Current state definition, list of possible root causes, statistical analysis, and process capability analysis.

5.2.1. Current state definition

When studying historic turnaround times, one of the key challenges is to determine which of all the possible independent variables from the system is/are the one(s) influencing the observed results. To overcome this challenge, this section evaluates historic trends of total and EPP/Biology lab's times from an analytical and statistical perspective.

An indicator used in process improvement to describe the current results in terms of the desired state is *Defect per million opportunities (DPMO)*, calculated by the formula below:

$$DPMO = \frac{1,000,000 \times \text{number of defects}}{\text{number of units} \times \text{number of opportunities per unit}}$$

Equation 1

The DPMO value shows the total number of defects or nonconformities that an outcome has, related to the number of chances the process has to generate a defect. This parameter differs from the calculation of number of parts defectives or incorrect outcomes when the indicator has more than one possible defect or a defect can be generated in several activities of the process.

In the case of turnaround time, *number of defects* is the number of instances that the time measured is greater than the agreed one, and the *number of units* is the number of requests that are being considered. The *number of opportunities per unit* in this case is 1 given that there is only 1 chance of "getting a unit wrong" (an assay is either on time or not). The time used as the parameter to define if the request is on-time or not is 10 days for EPP/Biology and 12 days for the total turnaround time. These two times are the ones already accepted as the threshold for the EPP associates. Table 6 shows the summary of the results taking historic data from 2007 to the third quarter of 2009, also showing the percentage of errors (DPMO÷10,000).

Table 6. DPMOs for Total and EPP/Biology lab turnaround time in 2007, 2008 and until Q3 2009

		2007				2008				2009		
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3
EPP Biology	N° of defects	766	1,095	925	730	1,175	979	89	258	287	113	127
	N° of opport.	3,481	3,994	4,661	2,748	2,467	5,166	3,045	3,019	2,833	2,705	2,535
	DPMO	220,052	274,161	198,455	265,648	476,287	189,508	29,228	85,459	101,306	41,774	50,099
	Percentage w/ defects	22.0%	27.4%	19.8%	26.6%	47.6%	19.0%	2.9%	8.5%	10.1%	4.2%	5.0%
Total time	N° of defects	791	1,395	1,053	780	1,315	1,001	133	385	390	142	201
	N° of opport.	3,481	3,994	4,661	2,748	2,467	5,166	3,045	3,019	2,833	2,705	2,535
	DPMO	227,234	349,274	225,917	283,843	533,036	193,767	43,678	127,526	137,663	52,495	79,290
	Percentage w/ defects	22.7%	34.9%	22.6%	28.4%	53.3%	19.4%	4.4%	12.8%	13.8%	5.2%	7.9%

When realizing previous study, some data was not included. This data will not be considered in the rest of the analysis:

1. *Times with zero days in total and EPP/Biology time:* Given current process flow, the turnaround time for assay profiling in the Biology lab has to be greater than zero. In the same way, if the time in the lab has to be greater than zero, the total turnaround time should also be greater than zero. Data points with zero turnaround time correspond to requests that did not go through the processes established in the maps (Figure 13 and Figure 14). Given that those specific activities are the ones under evaluation, these data points will not give any important insight about the process, although can confuse obtained results.

The reason why it is possible to have in the system zero days is because every compound has to be manually transferred from one period to another (via TRT). At the end of the process, the software has the constraint that if the compound is not "in the last step" in the system, the data cannot be evaluated or submitted. Requests that were not released in the appropriate moment, in order to evaluate the data they are virtually promoted through steps to close the loop. As commented before, these data points provide no real evidence of the process length or dynamics.

In contrast, requests with a zero value for the process in the NCA correspond are considered in the study. These cases correspond to requests that are generated internally by the biology lab when problems in the process arise. These problems require the compound to be tested again. In some cases, the lab has still enough amount to perform the test repetition or the chemists provide another vial with a sample of the same batch in order to decrease profiling time. In either case, the lab does not need material from the compound hub, reflecting a zero time in this activity. Given that the processes in analysis are the ones in EPP/Biology, these data points have information that can be useful to evaluate. Later in the project, this dataset is analyzed aside from the rest of the data, in order to observe trends or specific dynamics when compound repetition is present.

2. *Requests with a total turnaround time greater than 90 days:* as agreed internally with EPP's associates, any request with a time greater than 90 days correspond to data that was not updated in the system in the correct time and was virtually moved through the departments without reflecting the real duration in any phase. As discussed with the lab heads and the project heads, no project will wait more than 90 days for a potency result. This argument demonstrates how database reports are not being tracked and raises a flag of data possibly not being reliable.

3. *Rejected requests:* this type of requests is stopped at any point of the process because they present some error, which could be a typo, a compound requested that is not available, a compound quantity not

available, etc. Any time reported in the database for received requests is not certain to provide a real insight about profiling times.

From a first observation to Table 6, it can be observed how from 2007 to 2008 there has been an increase in the number of requests that were delivered under the expected timeframe (10 for EPP/Biology and 12 for total time). To corroborate that observation, a time series plot of the mean turnaround values over time is presented in Figure 24.
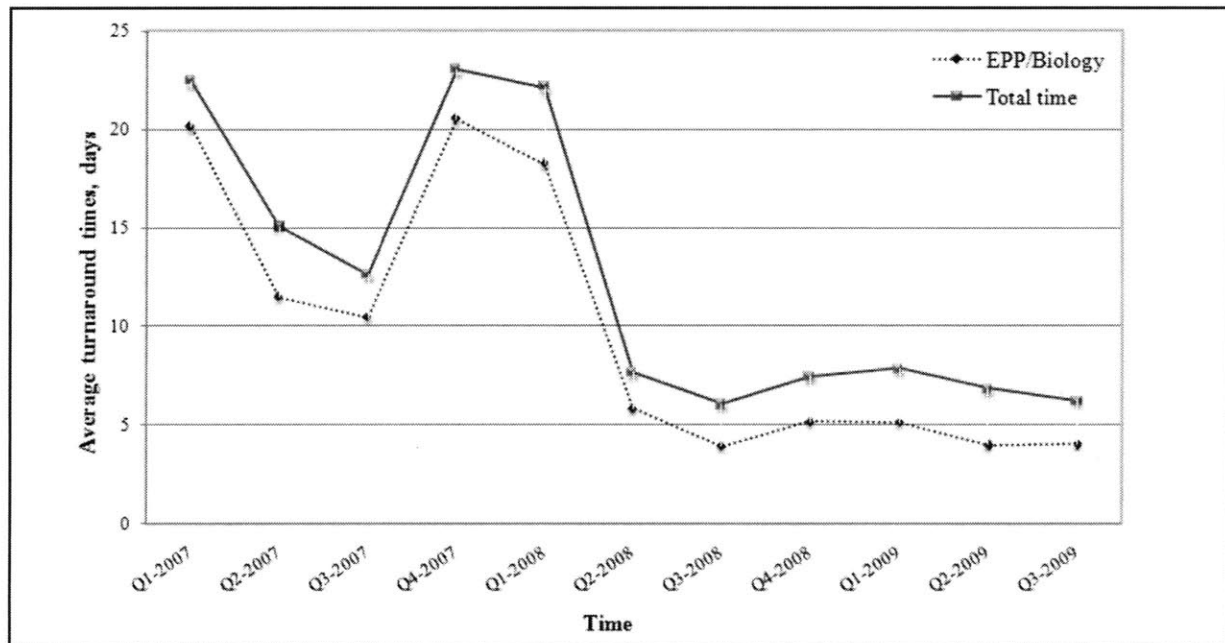
Figure 24. Time series plot for the length of requests in EPP/Biology lab and total time, from January 2007 to the third quarter of 2009

From this figure, the observation can be sustained and further statistical analysis is to be performed to define of the decrease in turnaround times over time is significant. To accomplish this study, a Mood's Median test is performed and the outcome is presented in Figure 25, assuming that the data from each population is an independent random sample and the population distributions have the same shape[41].

Mood's Median test was selected given that the data is not normally distributed and Mood's Median test is more robust against outliers and extreme values[42], which, as mentioned before, are present in these datasets. In the analysis, the average time for each quarter is compared, using data from Q1-2007 to Q3-

---

[41] (Rice, 2007)
[42] (Rice, 2007)

2009. The null hypothesis to be proved is that there is no significant difference between the average times of the quarters.

```
┌────────────────────────────────────────────────────┬────────────────────────────────────────────────────┐
│ Mood Median Test: Time in EPP/Biology Lab versus Q-Year │ Mood Median Test: Time in total loop versus Q-Year   │
│                                                    │                                                    │
│ Mood median test for Y2-TIME IN EPP/BIO            │ Mood median test for Y3-TOTAL TIME                 │
│ Chi-Square = 3548.79   DF = 10   P = 0.000         │ Chi-Square = 3216.46   DF = 10   P = 0.000         │
│                                                    │                                                    │
│                          Individual 95.0% CIs      │                          Individual 95.0% CIs      │
│ Q-Year  N<=   N>  Median Q3-Q1 --+---------+---------+---------+---- │ Q-Year  N<=   N>  Median Q3-Q1 ------+---------+---------+---------+-------- │
│ Q1-2007 1709 2274  5.00  8.00            *         │ Q1-2007 1776 2207  7.0   8.0        *---)           │
│ Q2-2007 1752 2435  6.00  9.00              *       │ Q2-2007 1232 2955  9.0  10.0             *--)       │
│ Q3-2007 4189 2087  2.00  6.00 *|                   │ Q3-2007 3726 2550  5.0  10.0     *                 │
│ Q4-2007 1317 1713  5.00 10.00            *         │ Q4-2007 1199 1831  8.0  10.0         (---*          │
│ Q1-2008  644 1966 10.00 11.00                  *---)│ Q1-2008  579 2031 14.0  15.0                  *--) │
│ Q2-2008 4856 2362  2.00  6.00 *                    │ Q2-2008 4561 2657  4.0   8.0 *                     │
│ Q3-2008 2436  820  2.00  4.00 *                    │ Q3-2008 2256 1000  5.0   4.0 (---*                 │
│ Q4-2008 2422  860  3.00  4.00    *                 │ Q4-2008 2155 1127  5.0   5.0    *                 │
│ Q1-2009 2131 1155  3.00  5.00    *                 │ Q1-2009 1742 1544  6.0   6.0        *             │
│ Q2-2009 1982 1208  3.00  5.00    *                 │ Q2-2009 1566 1624  7.0   5.0    (--*              │
│ Q3-2009 1729 1068  4.00  3.00       *              │ Q3-2009 1547 1250  6.0   4.0    *                 │
│                          --+---------+---------+---------+---- │                          ------+---------+---------+---------+-------- │
│                          2.5       5.0       7.5      10.0   │                              6.0       9.0      12.0           │
│                                                    │                                                    │
│ Overall median = 4.00                              │ Overall median = 6.0                               │
└────────────────────────────────────────────────────┴────────────────────────────────────────────────────┘
```

Figure 25. Mood's Median test result for Total and EPP/Biology times from Q1-2007 to Q3-2009

Two main conclusions can be derived from this figure:

1. The null hypothesis is rejected, concluding that there is a significant difference between the average turnaround times for the quarters. As observed in Figure 25, the quarters that present a significant difference for EPP/Biology are Q1, Q2 and Q4 of 2007 and Q1 of 2008. From these, Q1-2008 can be neglected, given that in January 2008 the personnel was changing buildings, so that there were almost no request profiled during this month. This month can be driving the average value down. One could evaluate the quarter with no January data, but this is not realized in this project in order to maintain the data form the database.

In the case of 2007 it can be observed how the efforts the lab has been promoting have generated a significant reduction, decreasing the average turnaround time of the lab from 5 days in 2007 to a value close to 3 days in the following year (if Q1 of 2008 is not considered).

2. The spread of the data and the length of the total turnaround times can be explained by the time in the EPP/Biology lab. It is concluded that the quarters that have significant difference correspond to the same ones that have a significant difference EPP/Biology lab time: Q1, Q2 and Q4 of 2007 and Q1 of 2008. This result shows how impactful it is to improve EPP/Biology operations. Given this trend, future analysis only consider EPP/Biology times, based on the conclusion that this time is the one that drives the total compound request profiling time.

With the aim of characterizing the current behavior, basic statistics are calculated for total and EPP/Biology lab times, and the results are shown in Figure 26, Figure 27 and Figure 28. Once again, it can be observed that the outliers present in the total request time perceived by the chemists, correspond to

outliers generated in the Biology Lab, so that, by controlling the time internally in the lab, significant results will follow.
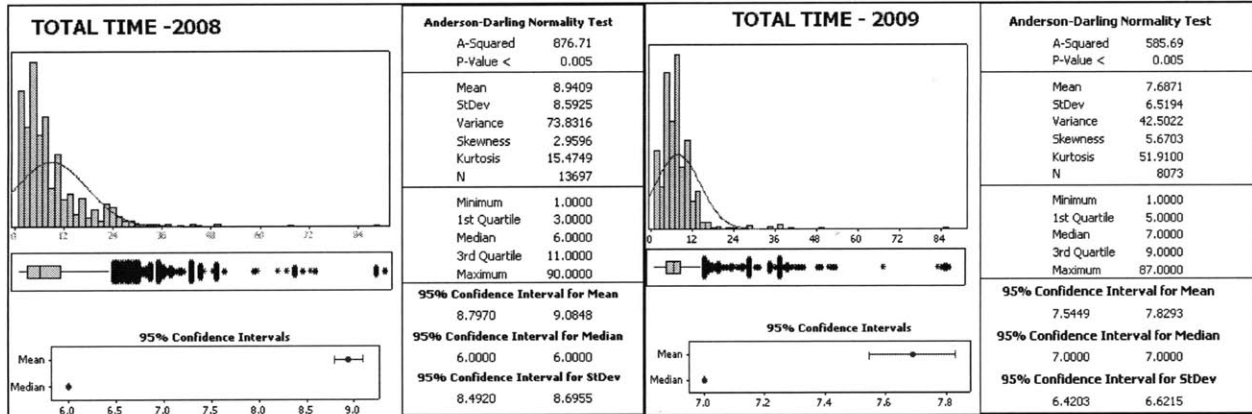
Figure 26. Basic statistics for the total turnaround time, from January 2008 to October 2009
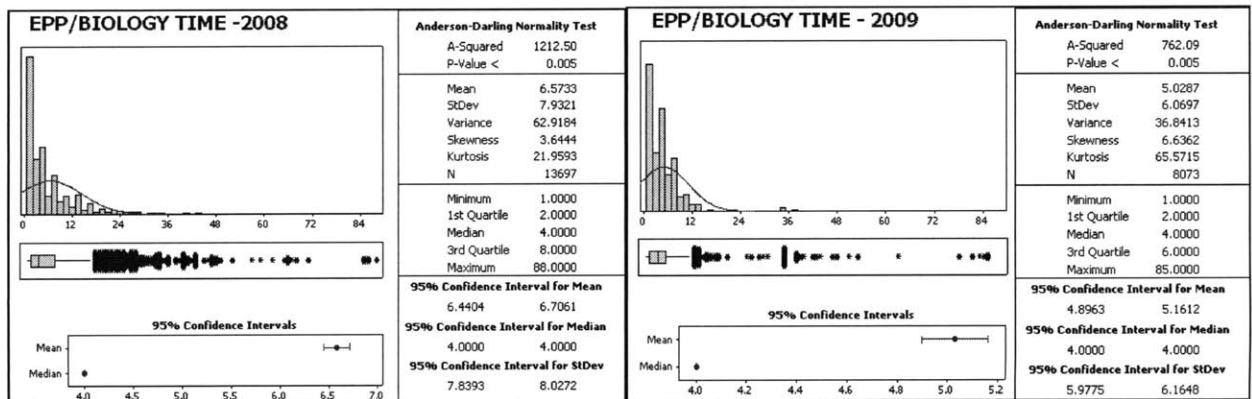
Figure 27. Basic statistics for the time in EPP/Biology lab, from January 2008 to October 2009
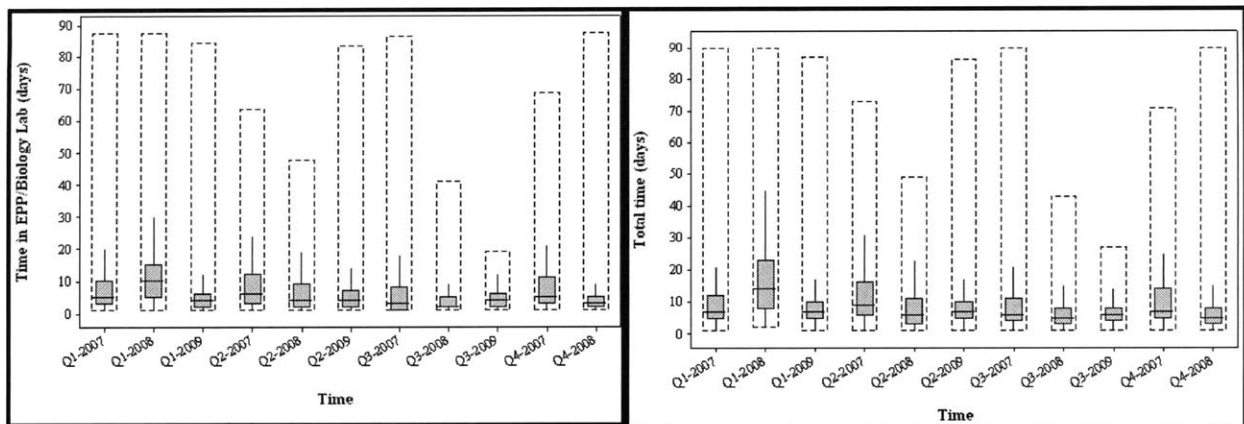
Figure 28. Box plot of profiling time in EPP/Biology Lab and total time for each quarter from 2007 to the third quarter of 2009

Finally, it is important to observe from the figures above how outliers are still present even when the data was cleaned before evaluation. This situation shows how the variation of the turnaround times is critical to the final profiling time, and a strict threshold for outlier definition ca not be established from the current values in database. This phenomenon is observed from the difference between mean and median values presented in Figure 26 and Figure 27, and clearly represented in Figure 28, which shows the box plot surrounded by a dotted line representing the total range.

In conclusion, from the analysis of the turnaround times it can be said that the profiling duration in the EPP/Biology controls the total duration of the requests turnaround times, not only in length but also in spread and variation. Even when average times have been decreasing over time total spread has not, which is one of the reasons leading to the large turnaround time difference mentioned by the chemists when the research for this project was starting. As a result, focus on EPP/Biology improvement can lead to an improvement of the overall turnaround compound test duration.

5.2.2. List of possible root causes

The brainstorming of root causes is formalized by the *5 Whys* procedure, resulting in the Fishbone diagram displayed in Appendix F. If the brainstormed causes are related to the current variables stored in the database, the list of possible root causes can be defined as:

- *Type of project:* it is perceived that projects can have different profiling times, depending on a particular priority, number of compounds produced, or even interest of the personnel (mostly lab heads). To analyze this situation, the type of project will be evaluated using type of screening, type of profile, type of protocol, and target for each specific compound requested.
- *Internal vs. external requests:* as discussed before when evaluating the profiling times in the lab, internal requests are observed to have a minor turnaround time than normal requests. In order to define if this difference is significant, a statistical analysis is performed to compare internal versus external compound requests. Some of these requests can even be for the same compounds, if the trigger for the internal requests is the need for a profiling repetition. For this analysis, requests are separated into EPP/Biology internal requests, EPP/MCH for requests from the chemistry lab within EPP and others, which will correspond to any requestor out of EPP projects.

- *Phase of projects:* another variable that can affect the turnaround time is the phase of the project. It is understood that projects in some stages are more important than others and may have priority when profiling. To evaluate this situation, requests are divided in the phase established in the system when the

compound is submitted. It can be the case that a project, overall, is in a phase but a compound is submitted to an earlier stage of the same project. This is present when second generation of compound is developed in order to have backups for future analysis. This separation of compounds by phase will reveal is the compounds that are moving towards later phases are really prioritized in the profiling process.

- *Location:* there is the argument that having a centralized profiling process in Switzerland for other locations in the world introduces variability and longer lead times to the total length of compound profiling. For EPP some projects are driven by the chemistry in Massachusetts, USA, and the profiling is realized in Basel, Switzerland. One of the thoughts is that these data points should not be considered when calculating the total turnaround time, given that the biology lab cannot control shipping times. This scenario will statistical evaluated analyzing the profiling times for each location.

5.2.3. Statistical Analysis

With the intention of reducing the impact of outliers in the analysis, only data from the second quarter of 2008 to the third quarter of 2009 is considered. Also, each variable will be evaluated towards EPP/Biology profiling time, referring to the conclusion stated before that outliers and total process performance are described by EPP/Biology lab's behavior. Lastly, Minitab is employed for the statistical analysis and Mood's Median Test is used if not stated otherwise.

The results obtained for the Mood's Median Test are presented in Appendix G. As observed, the screening types with codes As2, Se16 and Se4 are the ones that present a higher variability. Even when the turnaround times for these screening types can be affecting the biology lab performance, only Se4 and Se5 can be said to be statistically different, given that the confidence intervals for all the rest overlap at least at the lower values.

The second statistical analysis is with the protocol type, which is separated only in two categories: primary and secondary. Primary screening is the first screening that a compound needs and is realized to obtain a rough potency value. The only question to be answer with this analysis is if the compound is active or not towards certain target. A secondary screening is necessary when comparing between different scaffolds or compounds within the same family, so the exact values become critical.

Observing the data Appendix G, clearly secondary screenings have significantly higher profiling times than primary screenings. From a first thought, this difference can be perceived as coherent with the process given that secondary screenings can be understood as needing more time to obtain specific potency values. This conclusion is not correct. Going back to the process maps, it is observed that the

activities are the same for any compound, independent on any variable that is been analyzed here.

Instead, a root cause analysis shows that longer times in secondary assays are given by the plate optimization. With the current layout in the biology lab, each assay plate has a maximum of 14 compounds. In primary screening, many compounds are profiled to observe trends among scaffolds and choose from a broader selection of possibilities. In the other hand, secondary screening is performed under a more strict compound structure, having a less number of compounds produced per unit of time. If the usage of assay plates is optimized, secondary compound will be waiting longer for more compounds to come in order to fill the plate, which promoted longer lead times to this profiling type.

The last indicator is profile code that reflects the target type and enzyme present in the assay. For this analysis, it is concluded that there is remarkable different between values, not only in the average times but also in the spread of those periods. As shown in the fishbone diagram, most of the differences in profiling speed come from the lack of standard procedures among lab technicians, which is reflected in the profile code too. This result shows, once again, the effect of non-standard profiling schedules for all the projects and assay panels.

Interesting results are obtained when assessing the impact of internal versus external in Biology lab's profiling times. As shown in Appendix G, requests that are generated internally by the lab associates have a significantly lower turnaround time than requests from EPP/Chemistry lab or other Disease Areas. This behavior is generated by the fact that internal requests are due to either internal error so or a particular compound need. In both cases, is in the best interest of the direct associate to quickly analyze the compound. In either case, the compounds are treated preferentially in the process, resulting in a significantly lower profiling time. Even when times in NCA are zero for these compounds (meaning that at least they have 2 days less) the difference is still significant and even more than 2 days, probably showing the minimum achievable time in the internal process of the lab.

Continuing with the analysis of internal versus external requests, time lengths for requests that come from other Disease Areas are significantly higher than those from internal requests, and have much more variation than internal or MCH's ones. This trend points an area for improvement: scheduling is currently not showing standard basis, but a preference mode. Further analysis, out of the scope of this project, should be realized in order to observe the factors affecting this mode of operation.

With all above analysis, it can be concluded that both, internal and non-specific EPP project requests are shifting the measured turnaround times towards un-real values. In order to have a more standardized and controlled process, systematic procedures need to be in place so that any request is delivered within

specific timeframes, in order to push overall drug discovery length when referring to protease targets. Finally, when implementing a metric system, times should be separated to observe this type of trends and dynamics in the lab, to have real and sustainable benefits.

Now profiling time will be evaluated towards the project phase. Following the same procedure with Moods Median test, it is observed the higher the phase the more spread the turnaround times have. It is interesting how; particularly lead candidate, D3 and D4 phases have significantly longer turnaround times. Concern may rise also when observing the length and spread of compounds for projects in D4. Deep analysis of root causes and interviews with stakeholders reveal how compounds that go into more advanced phases are usually more potent than the current process is set for. Because of this, most of the compounds that come into D3 and D4 phases are profiled and then, after observing that they are so potent that IC50 cannot be determined from the graphs, the compound is further diluted and the IC50 determination is repeated. This procedure not only slows the entire assay speed, but also consumes materials and introduces the repetition that was analyzed in previous sections and defined in the process maps. This situation is also considered in the value-added analysis shown in Figure 21, as the probability of compound repetition. Further recommendations will address this limitation of the current flow of activities.

Lastly, location is evaluated in profiling times. From Minitab's output it is concluded that the work with the US is indeed increasing the total turnaround times with no difference for the profiling time within the lab. As noticed from the same Appendix, average values for the US increase from three to nine days when shipping periods are considered. Even when this may seem as a disadvantage, the total number of compounds profiled in Switzerland is significantly higher compared to the rest of the world, so that their impact is not really substantial. If cross-country projects are to be increased, this is the best time to optimize Switzerland's main procedure to ensure a steady and predictable profiling timeframe for any country, considering shipping times.

 The final analysis is derived from the interviews with all stakeholders, who show concern when thinking on implementing changes to the process, given that they support about 170 programs Novartis wide. The following Table shows a Pareto analysis of all the programs.

Table 7. Pareto evaluation of the programs supported by EPP over the last 3 years

PROGRAM	Cumulative % of annual requests	Average of NCA time (days)	Average of EPP/Biology time (days)	Average total time (days)
e12	18.2%	4.2	15.9	20.1
a19	31.0%	2.7	16.0	18.8
a26	42.7%	4.5	10.2	14.8
e4	50.5%	2.7	10.8	13.5
c13	56.3%	2.9	5.0	7.8
d12	61.9%	2.7	4.3	7.0
c14	67.3%	2.9	12.3	15.2
f6	70.5%	3.2	14.8	17.9
d2	73.5%	2.7	8.7	11.4
x	75.9%	2.9	15.8	18.8
c10	78.2%	3.5	11.0	14.5
f	80.3%	4.8	10.2	15.0

It can be concluded that from the 172 programs supported by EPP/Biology lab, only 18 programs account for 80% of all the activities. Also, it can be observed that the time in EPP/Biology for 9 out of the 18 programs has an average equal or higher than 10 days, which means that, even if on average the total profiling times on the biology lab have an average of 6 days, half of the programs that account for the 80% of the requests will perceive an unacceptable turnaround time (i.e. higher than 12 days). This shows again how process improvement can be obtained just by focusing on the core activities in the lab, and when standardized and controlled processes are in place, all current outliers or non-common assays will follow the optimal trends too.

5.2.4. Process Capability study

This study compared the process to a specific goal and its consistency around the average performance. This evaluation is employed to assess the ability of a process in meeting expectations and observe changes that have to occur in order to improve it. In the same way, when tracked, process capability study is an accurate and easy way to study the impact that changes have in the overall KPI in this study.

The indicators used to define the capability of a process are Cp, CPk, Pp and Ppk. The equations to calculate these indicators are:

$$Cp = \frac{USL - LSL}{6s}$$

Equation 2

$$Cpk = Min(\frac{USL - \bar{x}}{3s}, \frac{\bar{x} - LSL}{3s})$$

Equation 3

$$Pp = \frac{USL - LSL}{6s}$$

Equation 4

$$Ppk = Min(\frac{USL - \bar{x}}{3s}, \frac{\bar{x} - LSL}{3s})$$

Equation 5

While Cp and CPk, known as capability indicators, show the ability of the process to meet goals and its trend around the mean, Pp and Ppk extrapolate the behavior of the system in the future based on the observed noise and trend in the present. The latter are known as process performance measures. In other words, Cpk gives information about the capacity of the process to meet certain requirements in the present, using a calculated standard deviation from a short period (usually 50 data points in a shorter period of time, i.e. days), while Ppk extrapolates current noise to forthcoming scenarios (the same 50 data points but larger timeframes, i.e. months)[43]. Both are commonly used depending on the aim of the study.

A big advantage of this type of analysis is that it permits the specific study of the noise and stability of a process in a really simple manner. The following figure shows the different scenarios of process stability.
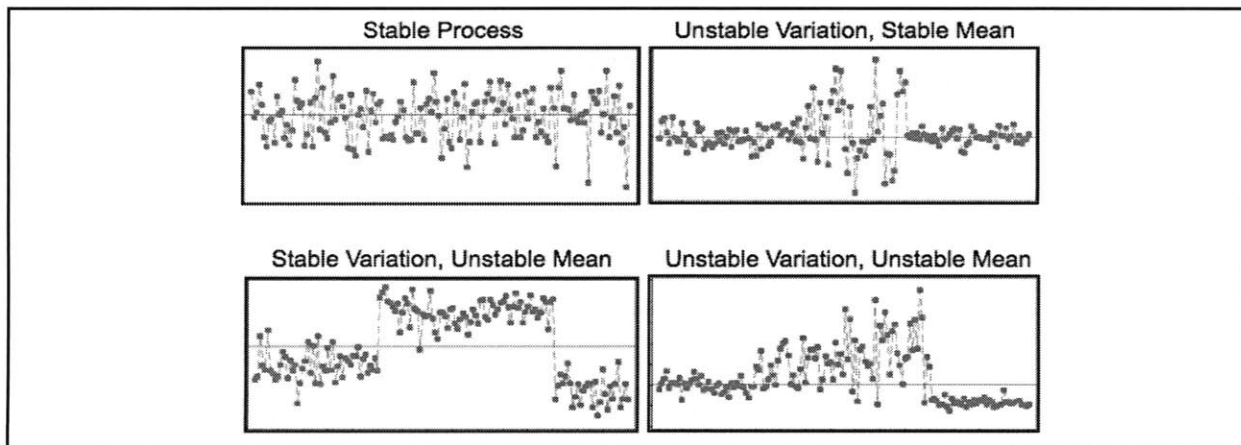
Figure 29. Examples of process stability trends[44]

In the same lines, by looking at the trends in a Control Chart, common and special cause variation can be separated and actions can be focused on the real limitations of the process. Knowing the specific process noise that can come from equipment, methods, purity of compounds, etc, actions can be taken when process outcomes have un-expected and out-of-control patterns in the results. The final goal of this

---

[43] (Novartis Technical Operations)
[44] (Adapted from IQP internal presentation: *Process Stability*)

analysis is to achieve a state were the process is in control and any abnormal behavior can be observed and addressed promptly.

Process capability study for turnaround times is realized using Minitab's non-normal data analysis tools, utilizing data points from Q1-2008 to Q3-2009. The complete output of Minitab's Capability Analysis is presented in Figure 30 for EPP/Biology times.
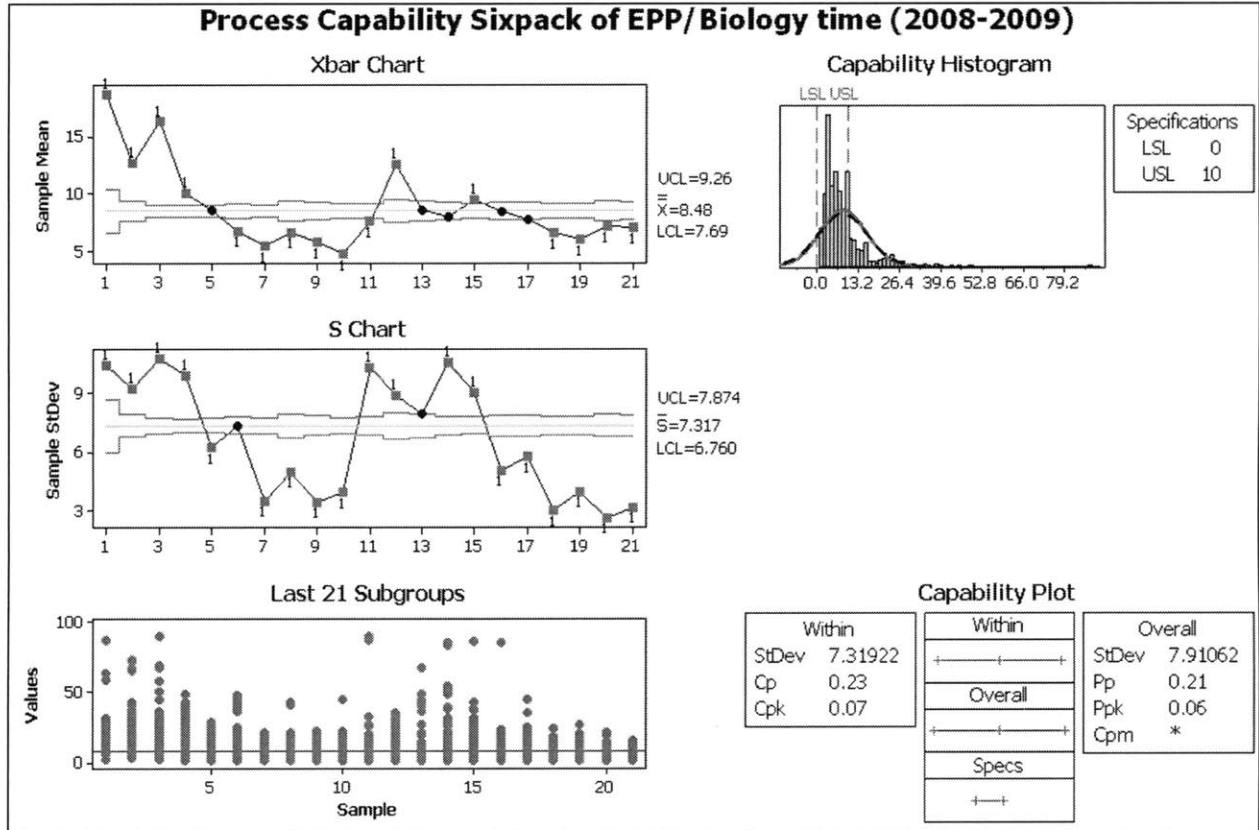
Figure 30. Results from the Capability Analysis for the process in EPP/Biology Lab

Starting with EPP/Biology turnaround times, when combining the data distribution presented in Figure 27, the Capability Histogram in Figure 30 is obtained. As noticed, a significant portion of the data is under the 10-day limit but some outliers are skewing the final average turnaround result to the right (towards higher turnaround times).

Looking at the Xbar and S Charts in the left of the figure, it is evident that, even when upper and lower control limits (UCL, LCL respectively) are calculated based on the data and seem to have ideal values, because of process instability, only few data points are within 3 standard deviations around the mean and only 2 months in the 2 years are within the boundaries of standard deviation. The explanation for these

results is given again by the influence of the outliers: a large number of low times (represented mostly by internal requests considered in the turnaround time) pulls down the mean, while unexpected high turnaround times provide noise to the outcomes. As a result, the mean value is close to the goal of 10 days (8.48 days as shown in the chart) and control limits are tight to that value, while most of the dataset is out of the range. Values marked with red in these charts are the result of the *test for special cause variation*, which mark points that present a distinct trend and should be analyzed. Given that most of the data points are marked, it can be concluded that noise in the process is causing the unpredictability and instability perceived by the chemists.

The only analysis left now is the process capability and performance. As observed, Cpk and Ppk have values of 0.07 and 0.06 respectively. Given that these values are significantly lower than 0.5 (theoretical used threshold) it is demonstrated that special and common causes for variation can be separated until the process is stabilized.

As a general reference, Figure 31 shows the results of the same capability study discussed before, but the time for the total turnaround times. As observed, findings are comparable to the ones presented for EPP/Biology time, reinforcing once again that the Biology lab generates the total profiling time variation.
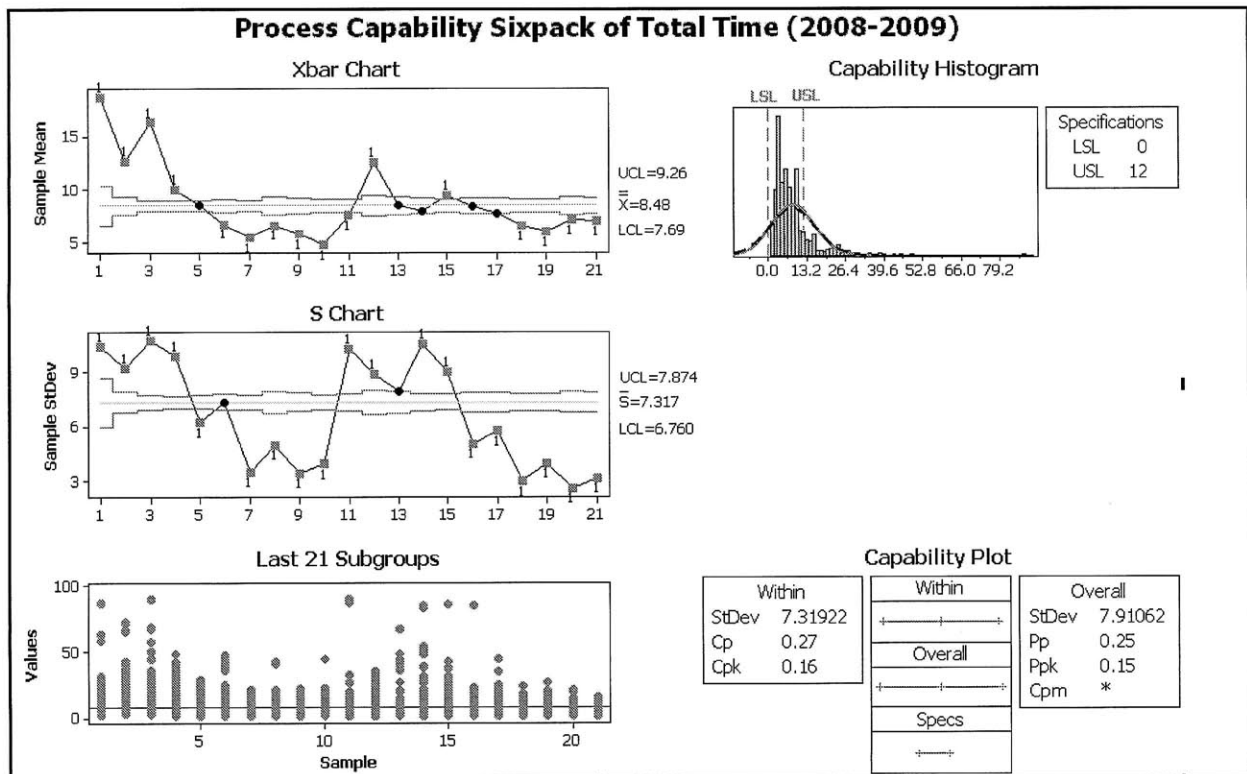
Figure 31. Results from the Capability Analysis for the total process

Finally, one last comment has to be made about the results obtained when outliers are considered. When setting a boundary for the data outliers, ideal case is that the dataset presents a clear gap, so that outliers can be are clearly observed and the threshold can be established with confidence. In contrast, EPP historic turnaround times cover all possible durations; therefore 90 days seemed an arbitrary decision. If all data points are evaluated, capability analysis calculates UCL and LCL of 18.6 and 0 for EPP/Biology times with a mean of 5 days. This result strengths the need for reducing process noise before concluding the effect of changes in the final process result, which can be achieved by standard procedures and the utilization of a measurement system.

In conclusion, turnaround times in the EPP/Biology lab deceased from an average of 8 days in 2007 to an average of 4 days in 2009. To define a significant difference between these values, internal requests should not be considered, thus increasing the mentioned turnaround time. Turnaround times for the entire loop are affected and defined by the length of the profiling time in the biology lab, and are characterized by a high variation. Variables that are affecting turnaround time's dynamics cannot be strictly defined until process noise is separated from special cause variation.

5.2. AC50 of Reference Compound

Shifting to the quality side of the established KPIs, the first metric is plate acceptance rate, which represents the assay plate integrity. This measurement is realized by tracking the reference compounds (RCs) in the plate and relies on the extrapolation of RC's integrity to the rest of the compounds in the plate.

As commented in prior chapters of this project, the extensive number of available variables for each compound inhibition curve has lead to the usage of different metrics by each lab associate, without comparing data consistency across assays. This lack of cross-assay review creates the data variation perceived by chemists and decreases the reliability in the lab's profiling data.

The first sub-class under the quality metrics shown in Figure 18 is RC AC50, which compares expected RC potency with the result in each plate.

5.2.1. Current state description

During the last few years, EPP has been implementing several software changes with the objective of improving data collection and analysis. With these changes, many procedures and measured variables have been changing also, limiting data comparability. To provide an analysis with recent data and in

comparable conditions, data used in the quality analysis correspond to the period between January 2009 and October 2009.

When analyzing historic data from RC, some surprising observations are identified:

- *8 RC are used in 83% of the assays:* From the data in analysis, 83% of the assays used 8 RCs out of the 27 listed. Figure 32 has a descriptive Pareto chart of the result.

Several advantages can be mentioned from the use of the same RC across assays, aside from cost benefits of buying higher quantities of a reagent. Having the same RC can help to better understand how the specific enzyme and/or substrate affect the observed compound potency, helping also to understand specifics about the target in evaluation. Also, when improving the process, clear changes in most of the assays can be achieved when focusing only on these 8 RCs.
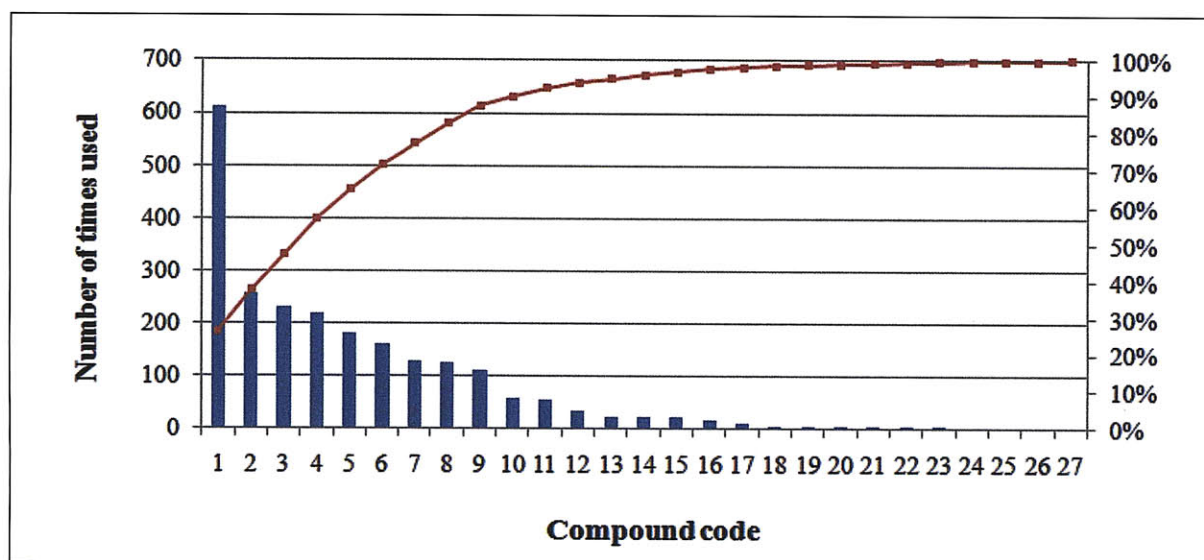
Figure 32. Pareto chart for RC usage

- *20% of the RCs have repeated tests:* this observation reveals the yield of the current profiling process and is related to the previous observation of assay repetition for compounds in D3 and D4. A root cause evaluation to understand the reasons for this 20% is performed, finding that 95% of this current 20% repetition rate is due to the presence of highly potent compounds, which need a second or third dilution for AC50 determination.

- *4 RC are not properly identified as such:* When looking at the number of data points for each compound, 4 compounds appear to be used as RC but are not being tracked as such, nor have control limits to define plate integrity. If a compound is not defined as RC, even if the AC50 is manually checked for every assay, specific changes or trends in RC are not detected and plate validation is constrained.

Also, given that most of the assays share the same RC, any information on outliers or peculiarities with a specific compound batch can become beneficial for several other assays too.

- *Some of the UCL and LCL defined for RC can be improved:* Observation of particular trends of RC over time and its comparison towards the control limits established in the database shows that these limits are not constantly reviewed and/or updated. First, when considering each AC50 data point obtained for a particular assay plate, data shows that 48% of the RCs have no defined control limits, affecting about 12% of realized assays. This means that, for 12% of the independent assay plates, the RC used had no quality control limits in the system. It cannot be said that the values of the RC are not evaluated manually for these compounds or are incorrect, but is evident that these compounds are not being automatically checked.

Using the same dataset, 2% of the AC50s are below the LCL and 5% are higher than the UCL, representing a total of 19% of assays that fail the RC quality check. All the data for the compounds present in these assays was submitted to the chemists, meaning that chemists received data with unknown quality, which is one of the reasons for the perceived data quality variation by the chemists. Figure 33 graphically shows a summary of all above-mentioned results.
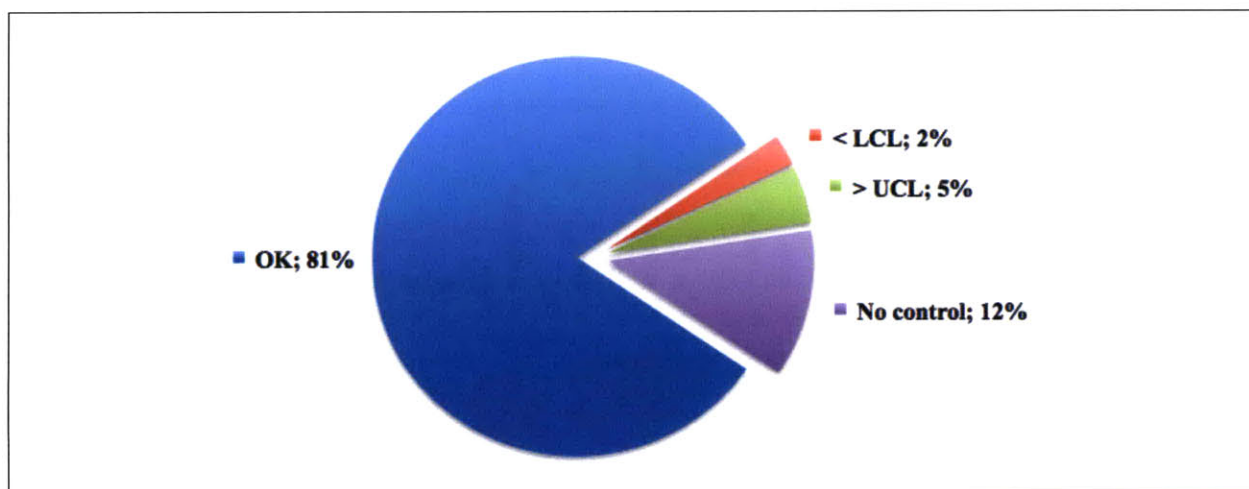
Figure 33. Distribution of RC AC50 with respect to the control limits in the database, from January to October 2009

A more in depth analysis shows that 97% of the assays without a control limit correspond to the new plate layout recently established, which means that assay reproducibility and repeatability was not realized before starting compound profiling. Even if new layout can change behavior of the RCs, leading to a different AC50 from the one in the database, at least historic data could be used as a first approximation to define plate acceptance. This situation will be discussed later again when identifying root causes.

- *Range UCL to LCL is too high for 65% of cases:* Given that the same RC is used in several assays but the conditions of the assay can lead towards different control limits, for the timeframe between January and October 2009, there are 144 different combinations of RC-assay. One conservative rule of thumb in this type of profiling is that a difference greater than 10-fold from one AC50 to the other is considered significant, meaning that, for repeated assays, if the final AC50 of a compound is in a 10-fold range (higher or lower), one could argue that the method is valid and any difference is explained by target and assay condition normal variability. This rule implies that any accepted value for a RC should also fall in this 10-fold window, which also implies that the ratio between UCL and LCL should be 10 or less. When analyzing the range established for the combination RC-assay it is observed that 65% of the RC-assay combination, have a ratio greater than 10. Figure 34 shows all the ratios stored in the database and the corresponding usage distribution.
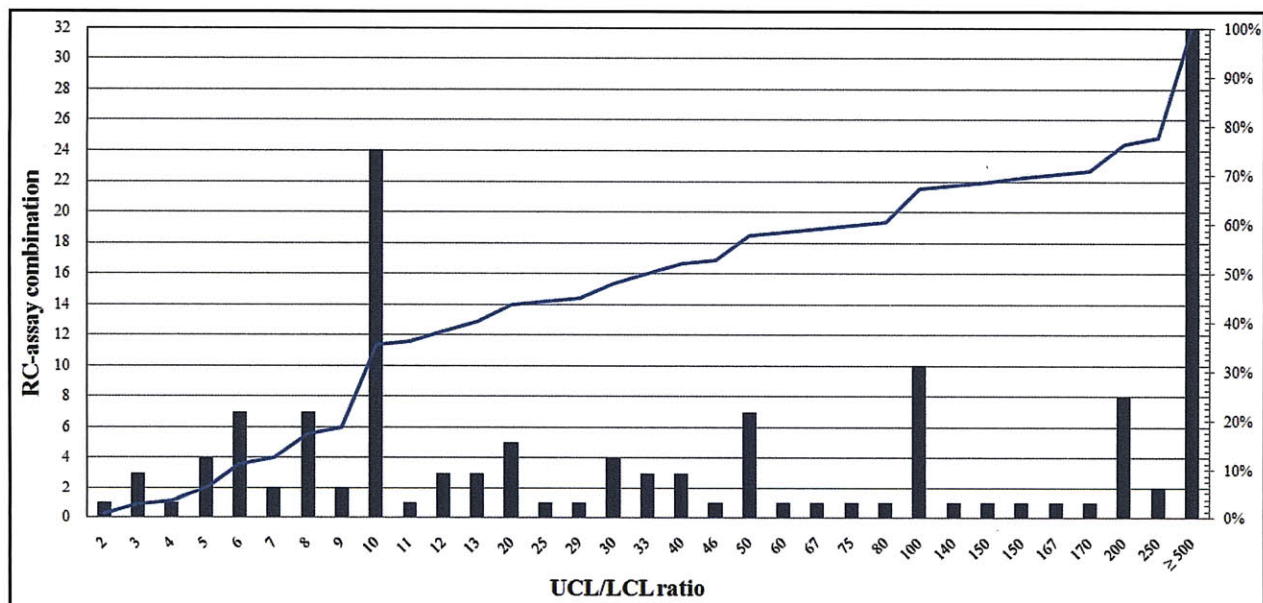
Figure 34. Distribution of UCL/LCL ratio for the total RC-assay combination used from January to October 2009

Even though most of the ratios are in fact 10, greater values should be clearly explained and agreement among all project teams should be present. Given that RC control limits are the first approximation for an evaluation of assay plate data quality, clear conditions and relation to assay conditions is key for future improvements.

5.2.2. List of possible root causes

After all above results, possible root causes for the encountered situations need to be listed. Appendix I presents the fishbone for root cause analysis, evaluating the limitations in the control limits and RC AC50 constant check. From it, the list can be reduced to:

- *Novelty of the assay:* When an assay is new and the RC is new to the platform, no historical data can be used to define the range of expected potency. Depending on the case, some assays are key to a project and its quick development is critical, so that the project team decides to start compound profiling without specifically knowing data for the RC. In some of these cases, ranges are not introduced or, if so, high ratios characterize the control limits. Over time, many of these initial values (or lack of values) remain for the entire length of the project campaign, and is the explanation of the observed 95% of the new plate layout without RC limits definition.

Statistical analysis of this factor will not be realized explicitly, given that there is not a direct variable saved in the database that relates the assay plate name with the phase of the compounds in evaluation. Also, it could be the case that, for a particular assay, the assay plate contains compounds from different projects at different phases so that straight comparison can lead to misinterpretations. As an approximation to observe the effect of knowledge curve in assay performance, experiment date will be used as the independent variable.

- *Non-standard procedures:* As discussed before in several sections, lack of standard procedures among lab associates also affects the review of data quality. In this case, CPs use different software indicators to validate the assay creating noticeable discrepancies when reviewing consistency among assays. The independent variable to study in this situation is operator.
- *RC out of range is observed when assay is finished:* Because actions in the current procedures are reactive, a RC is known to have an AC50 out of range when the assay has been already completed. Most of the times, the assay data is kept and actions are taken for future assays maintaining the wrong value in the database.

Along the same lines, RC aging can only be observed if data is manually stored and plotted. When having an out of range result, what the associate observes is just that specific data point. The result can be due to aging of the compound, a perturbation to the whole plate or a factor that affected only that specific well. Given the urgency of assays, equipment booking schedules and profiling priorities, most of the time a RC out of range is not a clear-cut decision to repeat an assay.

To evaluate these dynamics, the experiment date will be used as the independent variable, which will also

show the trends of the RC and permit clear observation of batch aging.

In the same category, another important analysis is the determination of batch-to-batch differences, when enzyme and/or substrate batches are changed. The database contains each of the batches used in every assay, so that a statistical evaluation of their differences can be realized, if the RC-assay combination had a change in the time of the study. Further analysis of the possibility for this study is discussed in the statistical analysis section.

- *Error in the robotic systems:* In this case, errors can be due to specific equipment failure or interaction between substrate and enzyme that interfere with the readouts. To distinguish between a robotic defect, pure variability of the readouts or an error carried out form the beginning of the assay plate generation the following variables are selected: readout timestamp, experiment time and assay plate name.

5.2.3. Statistical Analysis

A thorough statistical analysis for RC AC50 variation requires a separate evaluation of each of the 144 RC-assay combinations, considering in each case all the independent variables listed as possible root causes. Given its complexity and considering that 8 RC account for 80% of the assays, the statistical analysis is realized for these 8 compounds and the most significant assay configuration they represent. Results for particular discussions will be provided and the rest of non-published statistical outcomes will be mentioned and discussed.

Following the same procedure as the one for turnaround time evaluation, the effect of each of the mentioned independent variables on the RC AC50 is shown in Appendix J.

As observed, AC50 of RC varies significantly with the experiment date, which is actually a surprise. Even when non-controllable conditions may be affecting the resulted AC50, a significant variation of the RC's AC50 will potentially be related to a greater instability in the resulting AC50 of the tested compounds, given that RCs are usually the most potent compounds present in the assay plate. These types of trends are useful to observe in order to define which the days that have more variation are and to find root causes to eliminate result's fluctuation.

Even though the difference is significant, direct comparison with the control limits will show if the difference is enough to discard the data from the entire plate. In the next section, a process capability study will address this direct comparison.

The next variable evaluated is operator. As noticed, the effect of the operator can be significant and, as

expected, different for different assays depending on the operator's expertise and familiarity with the protocol. A lab associate usually owns a specific assay, so these differences are not an issue. If help between associates is required, it is important to measure the impact of that change in the final results, so that training and/or procedure analysis can be done before the temporary take-over.

Interesting results are derived when deviations from different substrate batches are analyzed. As noticed, some batches can insert an additional variation to the potency results, such that the final outcomes cannot be compared. This kind of evaluation is key before starting to profile so that batches can be characterized and disregarded before affecting the overall project data. Also, if needed to employ a substrate batch that is known to cause different data, measures can be applied to convert resulting data or compare in a relative scale results obtained with this batch with the other batches.

Readout timestamp also appears to have a significant impact in the AC50 obtained for RCs. As shown in Appendix J, AC50 values differ significantly depending on the time the readouts were performed, which could correspond to either a variation of the equipment or a particular situation with the assay conditions. As observed, for both revealed analyses, different timestamps have different statistical characteristics (i.e. mean, confidence intervals), which show that differences come mostly from equipment noise. This type of un-controllable noise adds variation to the process (common cause variation), and will be observed when evaluating the process capability.

Another compelling discovery is how the time a particular experiment is performed affects significantly some assays. As presented in the statistical analysis results, C1 and C2 behave comparably at any time of the day, while analyzed assays for C4 and C5 have a significant difference with the experiment time. Once again, it is important to mention that these results can be due not only to differences in the time itself but related to other specific variables not addressed in the analysis. Because of this, it is important to evaluate regularly the obtained RC's results so that specific causes of variation can be identified and corrected promptly.

Lastly, the assay plate name is evaluated. This variable is a consecutive number that provides not only an insight of the changes in results over time, but also from plate to plate, even if several plates were tested at the same time. As concluded in the statistical output from Appendix J, variation from plate to plate is related to the assay and could be linked also to specific perturbations due to robotic system defects or assay conditions that occurred only for that particular test. Once again, this reinforces the need for a protocol to continuously evaluate assay performance so that real root causes can be immediately defined.

In conclusion, as in the case for turnaround times, most of the variables studied affect one or more of the

RC-assay combination analyzed. Is evident that the noise of the process is so high and the interaction between variables so weakly understood, that specific root causes are hard to define though a statistical analysis.

5.2.4. Process capability study

The last evaluation for AC50 of RC is the capability of the process to replicate its value in separate assays. As an example, Figure 35 presents the study for C1 in assay 62-1. For this particular RC, most of the values fall into the upper and low control limits established in the system, as observed in the capability histogram in the right of the figure. It also underlines again the importance of a graphical data representation when aging of the RC batch is to be forecasted. Looking at the Xbar chart in the left of the figure, constant increase in AC50 over time is clear, and with the tools in place, batch changing can become proactive. In addition, is evident how after a batch change, AC50 of RC drops to a value within the expected range.
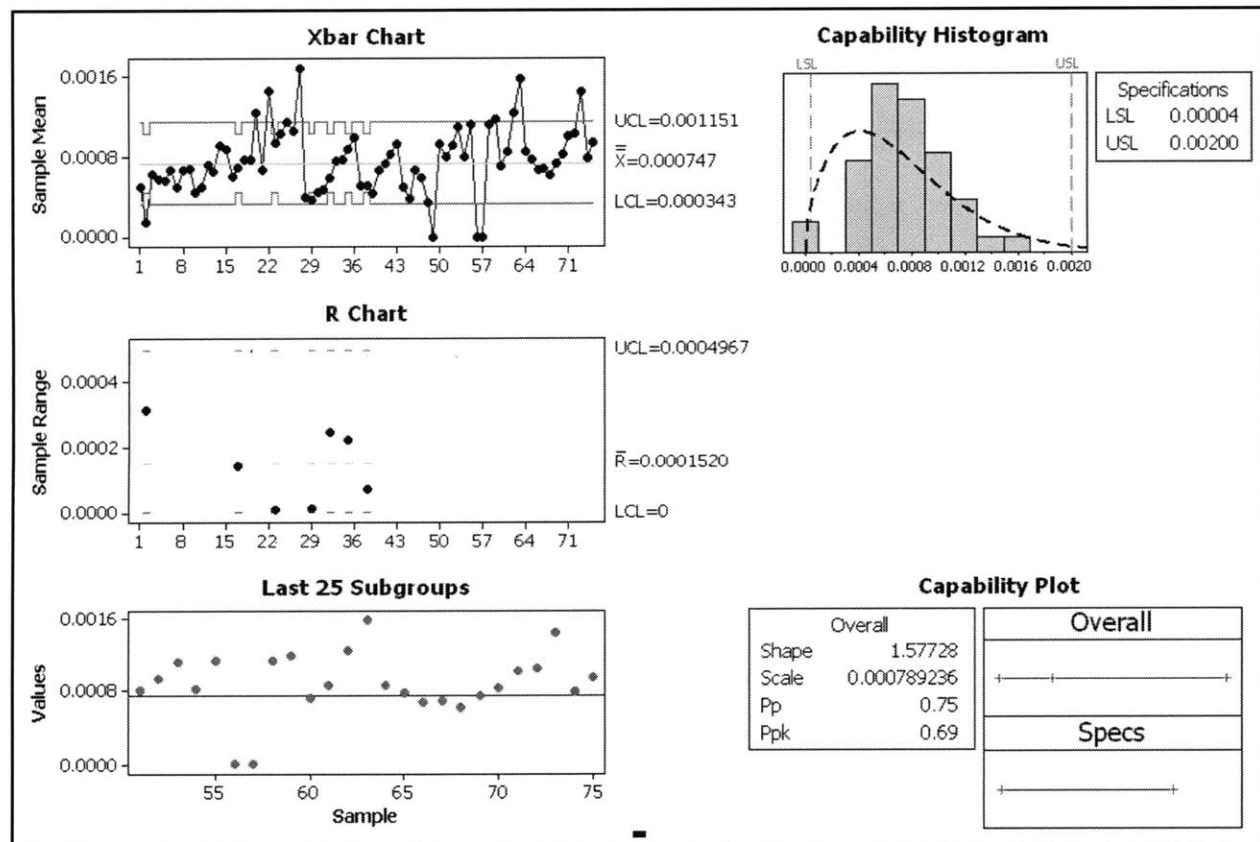
Figure 35. Process capability analysis for C1 in assay

5.3. RZ' of Reference Compound

5.3.1. Current state definition

As described before, parameters that do not contain any information about data variation are less appropriate for assay quality evaluation. As a result, robust Z' was defined as one of the most important variables to determine assay plate quality control and assay performance[45]. This indicator combines the signal amplitude and the variability of the assay, directly relating statistically significant thresholds of activity for compounds.

Because high RZ' values do not guarantee that an assay is valid or that the data is reproducible, this KPI is used in this case to monitor trends in time, evaluating assay consistency. Given that RZ' values are assay specific, these values are analyzed in the same way as the parameters for AC50 of RC. Considering data for the two most used RC, current state is shown in Figure 36.
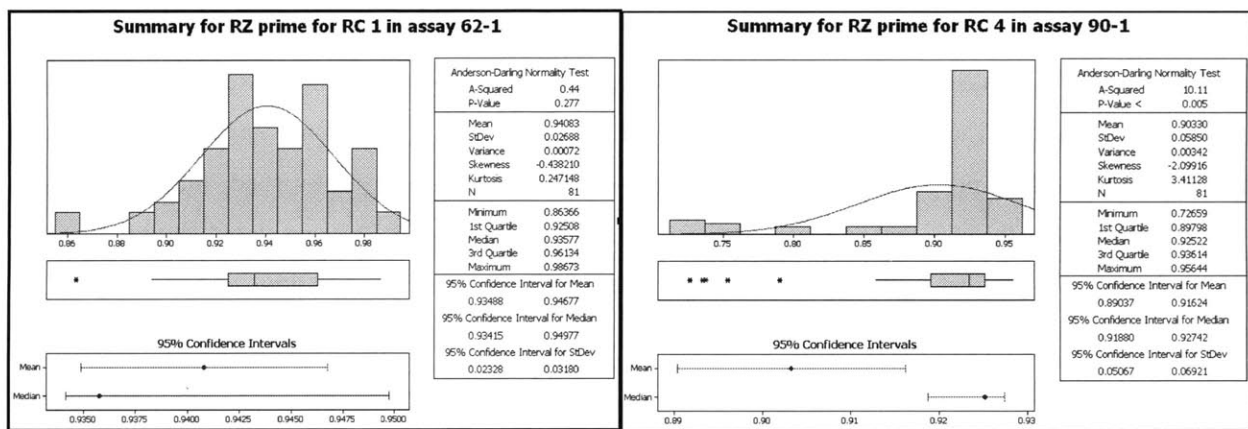
Figure 36. Basic statistics for current state description for selected RCs

It is important to observe that monitoring of RZ' is vital to characterize each particular assay and to define from there also how variable is it over time and how normal assay variability will affect the final potency of compounds tested. Also, an evaluation of this indicator can show, when combined with the rest of selected KPIs, variation produced by intrinsic assay noise from other direct special cause variations.

With the aim of expressing the variation between assays Table 8 presents basic statistics for 12 assays, which correspond to about 50% of realized tests for the timeframe in evaluation. These values reinforce the prior statement that characterization of the profiling test variation can be clearly defined by RZ' and that RZ' tracking has to be done on an assay-to-assay basis.

---

[45] (Gubler February, 2007)

Table 8. Basic statistics for RC1 in different assays performed from January to October 2009

Assay Code	Mean	Range	SdDev
19-1	0.8935	0.1363	0.0332
62-1	0.6627	0.0091	0.0052
88-1	0.9491	0.1086	0.0244
89-1	0.9077	0.2291	0.0508
90-1	0.9023	0.2299	0.0537
91-1	0.8632	0.4358	0.0898
41-1	0.8554	0.4605	0.0671
62-2	0.6627	0.0091	0.0052
39-1	0.7682	0.4034	0.1149
38-1	0.7144	0.4226	0.0996
55-1	0.8894	0.1843	0.0468
54-1	0.9085	0.1301	0.0335

5.3.2. List of possible root causes

Because all the possible factors that may cause RZ' to vary out of the expected range were listed also when evaluating root causes for RC AC50, the fishbone analysis in this case leads to the same output as the previous section's. The list of root causes is reduced to the same ones mentioned earlier as novelty of the assay, non-standard procedures, RZ' out of range is observed when assay is finished, and error in the robotic systems. For the statistical analysis the same independent variables are considered: experiment date, operator, substrate batch, readout timestamp, plate group name, experiment time, and assay plate name.

5.3.3. Statistical Analysis

It is not surprising to obtain identical results as the ones presented for AC50, when evaluating the significance of independent variables on RC's RZ'. If potency of RC is significantly affected by the variables described, RZ' has to be also affected in a similar manner.

Even though RZ' and AC50 of RC are equally influenced by assay variables, their study and tracking is still beneficial when validating assay plate data. With the current data, it is concluded that the noise of the process is limiting real definition of significant variables. If special and common cause variations are separated in the future, individual analysis of RC AC50 and RZ' can be helpful when determining specific reasons for results variation.

5.3.4. Process capability study

As usual, process capability diagrams allow a simple and straightforward evaluation of the trends and historic behavior. Figure 37 presents the results for two selected RCs. As observed, this study confirms that each assay will perform on a different RZ' window and specific control systems should be developed. For the RC-assay selection shown in the diagram, RC 1 can be characterized with a more stable and generally higher RZ' than RC 5. In the same way, RC 5 in assay 70-1, can be described by a more variable RZ', which can be part of the intrinsic noise of the assay. Track of RZ' can give insight to assay performance and provide additional information when in doubt about resulting data.
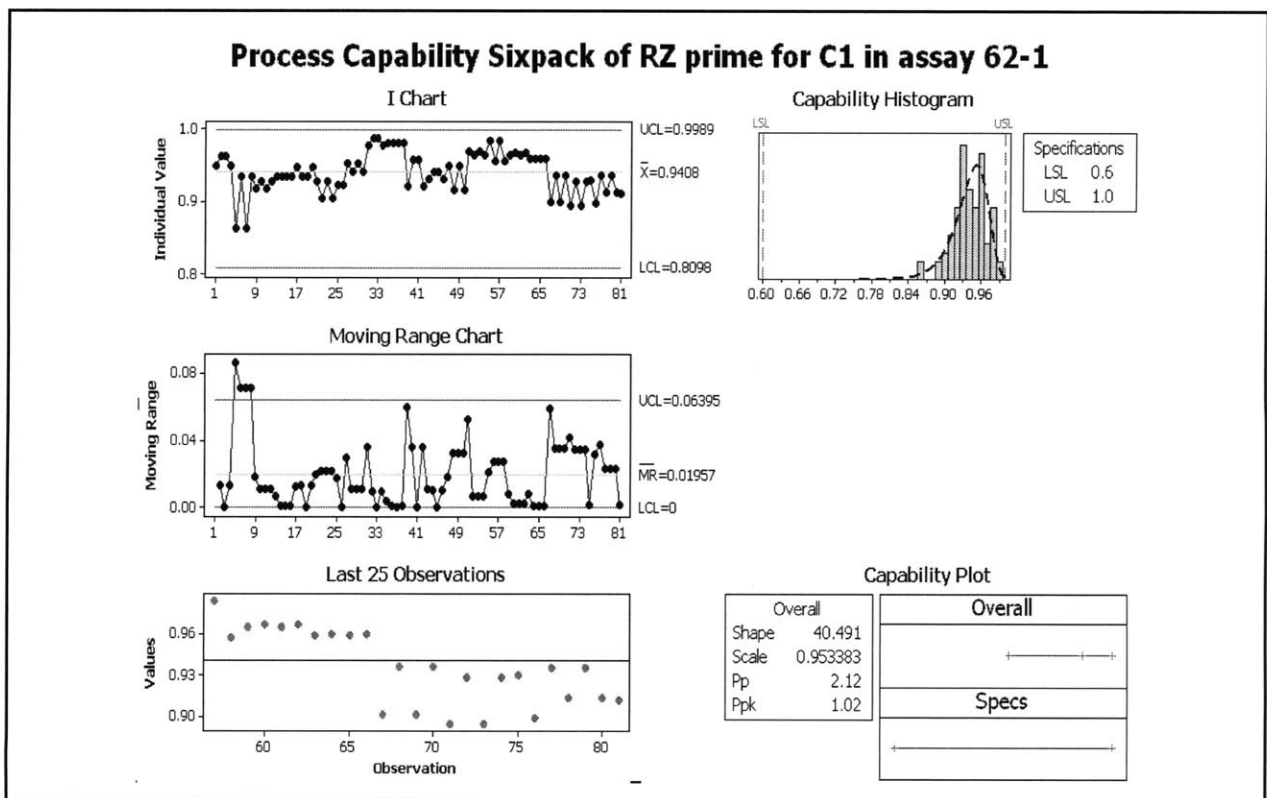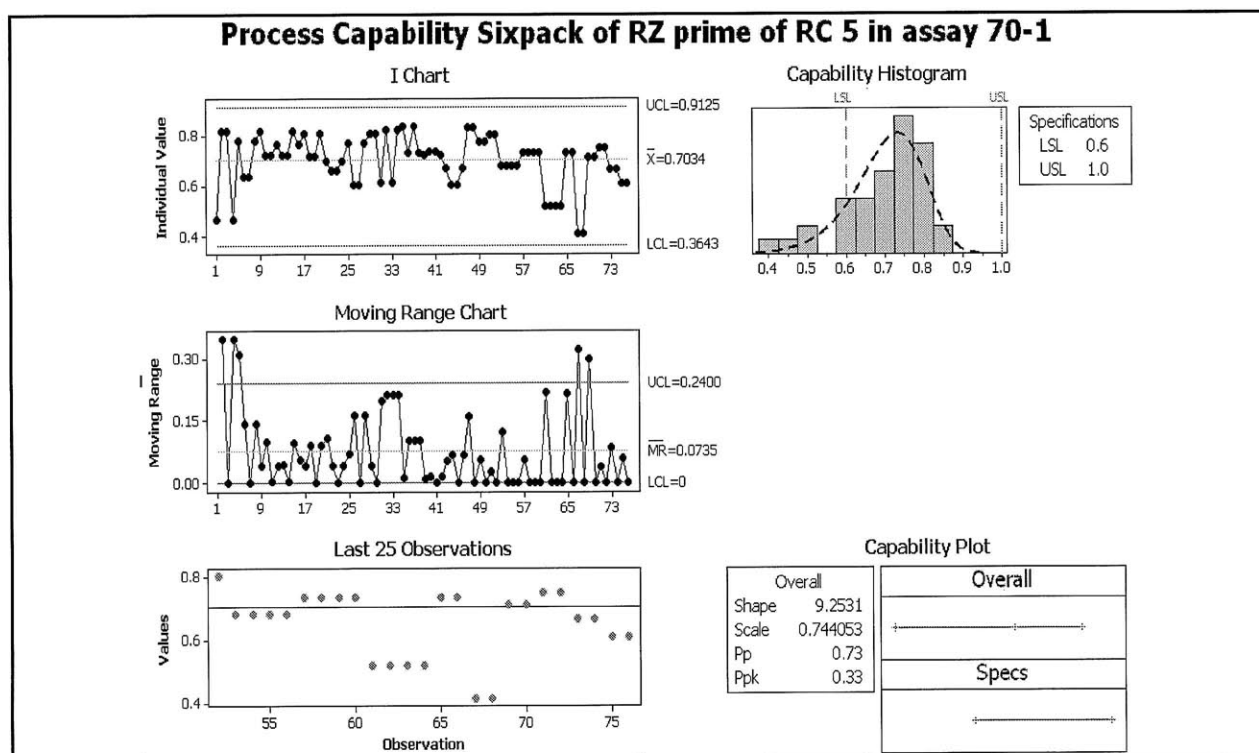
Figure 37. Process capability study for selected RC-assay combination

5.4. Remaining KPIs

The AC/NC ratio for plate acceptance, AC50 of compound ratio, and number of unmasked data points for compound data validation are the remaining indicators to evaluate. Because none of these variables are currently calculated by the software or tracked by the lab associates, the analysis cannot be performed. Even though, based on literature and lab's personnel experience, it is recommended to add these variables in future metrics to study compound data quality.

5.5. Observations from data analysis

From the analysis of all KPIs, it can be concluded that the current process presents a high and random variation that limits strict root cause determination. The statistical analysis showed that most, if not all, of the independent variables have a significant impact in the result outcome, which demonstrates that the intrinsic noise of the process (common cause variation) is being influenced by a series of unknown special cause variations. The limitation of not being able to discern between special and common cause variation is that trends cannot be improved, leading to unpredictable data patterns.

In order to develop specific process improvements thee following steps are recommended for

EPP/Biology lab to follow:

1. Establish a routine of monitoring recommended KPIs for each assay, with the goal of having more data points to increase the quality of the statistical analysis.

2. Prioritize assays given their impact on particular projects and/or the amount of compounds profiled over time. Select one or two projects to focus on in future improvement.

3. Perform a root cause analysis for results out of the expected limits, increasing the feedback to the process. This step will increase learning, not only of cause and effect dynamics but also of specific areas to implement in process controls.

4. Make changes to the process in order to adjust for the recognized causes and monitor the impact of the performed changes over the outcome. Continue with different variables until expected process behavior is obtained.

These four steps are iterative, so that, after improving certain assay, loop can be started again with a different assay, incorporating also learning from previous improvements.

3. CONCLUSIONS AND RECOMMENDATIONS

6.1. Overview of Results

Through out the previous sections, process and data analysis have described the current state for EPP/Biology's processes, have evaluated the weaknesses and strengths of the utilized tracking methods, and have shown the variables that drive process behavior. Conclusions from all these studies can be integrated to provide a list of areas for improvement and their impacts on measured results.

First, it can be concluded that the initial hypothesis of the project has been confirmed. As in section 3.1, the hypothesis stated that focus on providing chemistry with a potency value, without further analysis of the procedures or even of the quality of the data, has limited sustainable process improvement. As observed from the analyzes, constant technology upgrades and equipment changes, even when providing some decrease on profiling times, have promoted most of the NVA activities in CM and CP's thus limiting time available for continuous improvement. In the same way, these constant upgrades in technology have created a lack of standard procedures, which have impacted the stability of profiling times, and predictability and repeatability of potency values.

Second, it is important to reinforce that, even when focus on turnaround times can boos process speed, real root causes have to be addressed in order to achieve real improvement. All the concerns from the chemists and the biology lab associates are related to the lack of a common system that facilitates profiling, tracking, and data revision.

Lastly, from the statistical analysis of response variables with respect to possible independent factors, it became evident how the constant equipment upgrades and the lack of common methodologies have incorporated noise to the process, to the point where common and special cause variation cannot be differentiated. In order to have a correct process control, activities need first to have known patterns, so that special causes can be observed and resolved.

In order to create a sustainable improvement, it is important to maintain the current inertia towards process upgrade while adding control and tracking methodologies. Current momentum is good, and is important to keep it alive, but it is also necessary to add some control to observe the impacts these changes have on the final outcomes. The recommended metrics were listed and evaluated in section 4.2 and are summarized in Table 9. This table also shows the direct benefits of implementing these variables.

Table 9. Summary of suggested KPIs and their benefits

KPI	Current state	Area of impact	Benefits
Mean total turnaround time	Available, not tracked	Effectiveness / efficiency	Evaluate impact of results and changes in the overall loop time perceived by the chemists
Mean EPP/Biology lab turnaround time	Available, tracked	Effectiveness / efficiency	Define improvements in the highest portion of assay profiling duration
Range total turnaround time	Available, not tracked	Effectiveness / efficiency	Observe outliers and evaluate root causes to improve the overall compound cycle
Range EPP/Biology lab turnaround time	Available, not tracked	Effectiveness / efficiency	Observe EPP/Biology lab's outliers and evaluate root causes to improve internal processes
AC50 of RC	Available, tracked (not tracked over time)	Efficiency / data quality	Define strict control limits that help explain RC out of range leading to a better understanding of the assay's biology too
RZ' of RC	Available, not tracked	Data quality	Monitor normal assay noise window to assess abnormal behavior and plate validation
AC/NC ratio	Can be calculated, not tracked	Data quality	Track assay's potency range for plate validation and identify special cause variation
AC50 of cmpd ratio	Can be calculated, not tracked	Efficiency / data quality	Examine cmpd's repeatability to identify abnormal values or intrinsic assay performance
Number of masked data points	Recently incorporated, not tracked	Data quality	Classify the quality of cmpd's AC50 curve given the correspondence of data points with curve fitting

6.2. Recommended areas for improvement

In a more detailed manner Table 10 recommends specific changes at different stages in the future in order to achieve expected benefits. The table summarizes the areas to focus on, along with the actions recommended and perceived advantages. Figure 38 integrates all three focus areas of the project: quality, effectiveness, and efficiency.

Table 10. Summary of recommended actions, timeframe and direct benefits

WHEN	ACTION	IMPACT
Now	Incorporate chemist's knowledge about cmpd potency	- Eliminate the 20% assay repetition of potent compounds, which equals almost 30% of CP's NVA time and an entire run for CM
	Integrate equipment with available information	- Remove about 30 mins in CM's time (29% of NVA for CM) and 15 mins of CP's time (11% of NVA time for CP)
	Change booking system procedures so that CP book time slots and CM prepares based on that schedule	- Change process to full push flow - Reduce NVA time for CP by 10 mins
BENEFITS OF IMMEDIATE ACTIONS		1.Decrease almost 40% in NVA time of CM (from 104 to 64 mins) 2. Decrease 41% in NVA time of CP (from 134 to 79 mins)

(cont.) Table 10. Summary of recommended actions, timeframe and direct benefits

WHEN	ACTION	IMPACT
Next 3-6 months	Incorporate assay plate and cmpd data validation criteria	- Start tracking defined criteria to find special cause variation and optimize internal processes - Understand the assay performance and knowledge sharing among associates
	Develop graphic controls	- Facilitate outlier and trend observations, eliminating assay repetition - Relate any process change to a data point creating a learning environment about impact of changes – knowledge sharing among similar assays
	Define procedures to asses outliers and find special cause variation	- Decrease assay repetition solicited by chemists (5% of realized assays)
BENEFITS OF MIDTERM ACTIONS		1. **Eliminate 5% of unnecessary cmpd testing = 11 mins for CM and more than 13 mins for CP (complete cycles)** 2. **Change from reactive to proactive when aging of RC**
Next year	Incorporate standard procedures in assay development	- Standardize definition of RC control limit - Eliminate differences among associates so that best practices can be shared and assay profiling time is more stable
	Develop a scheduling tool	- Allow CP to select cmpds from TRT and book a robot. Results in freeing a computer and eliminating 15 mins of CP's NVA time (11%) - Provide available times for robots and calculate needed times depending on specific cmpds-assays, optimizing equipment usage - Link robot's information to update availability (not used or down) - Alarm new changes in robotic systems - Provide information to CM eliminating 5 different Excel files, need for developing, and supporting macros internally. Eliminate 25 mins of CM (24% of NVA time)
	Integrate an assay development tool with the process	- Make final procedures available to all associates - Require complete information such as dilution, enzymes in a panel, RC used, and control limits, providing a complete report dataset for repetition - Update of assay versions is constant and associates are aware of changes
	Integrate data generation with data submission	- Deliver email to chemists with a link to the database in the form needed - Eliminate 10 to 20 mins of CP's NVA time (7-14%) and having the same information in several places which currently leads to confusion about updated results
BENEFITS OF FUTURE ACTIONS		1. **Inform of timeframes for data submission** 2. **Close gap between mean and median values in about 2 days** 3. **Decrease turnaround times greater than 60 days in more than 50%** 4. **Decrease 39% in NVA time of CM(from 64 to 39 mins from previous improvement)** 5. **Decrease almost 32% in NVA time of CP(from 79 to 54 mins from previous improvement)** 6. **Decrease 30% in CP and CM total duration per assay plate produced**

An important observation from the analysis in prior chapters is that process variability comes from the current scheduling system and changing the planning procedures will allow to increase profiling time predictability, even if assays have also variability in demand. The current process relies on steady-state assumptions and flexible schedules to work around the variability in the number of compounds requested

for profiling. This adaptability has created observed outliers and affected process reliability. It is proposed to consider a rough estimation of 12 hours as the total process time (TPT) for the Biology lab, which , assuming that the production rate is about 14 compounds per day (a plate) per person, leads to an optimization point 14 compounds. This means that, the best results are obtained when profiling processes start whenever there are 14 compounds solicited for a particular panel. The 14 compound number is known as *process queue* and will permit the lab to operate in a smoother activity phase.
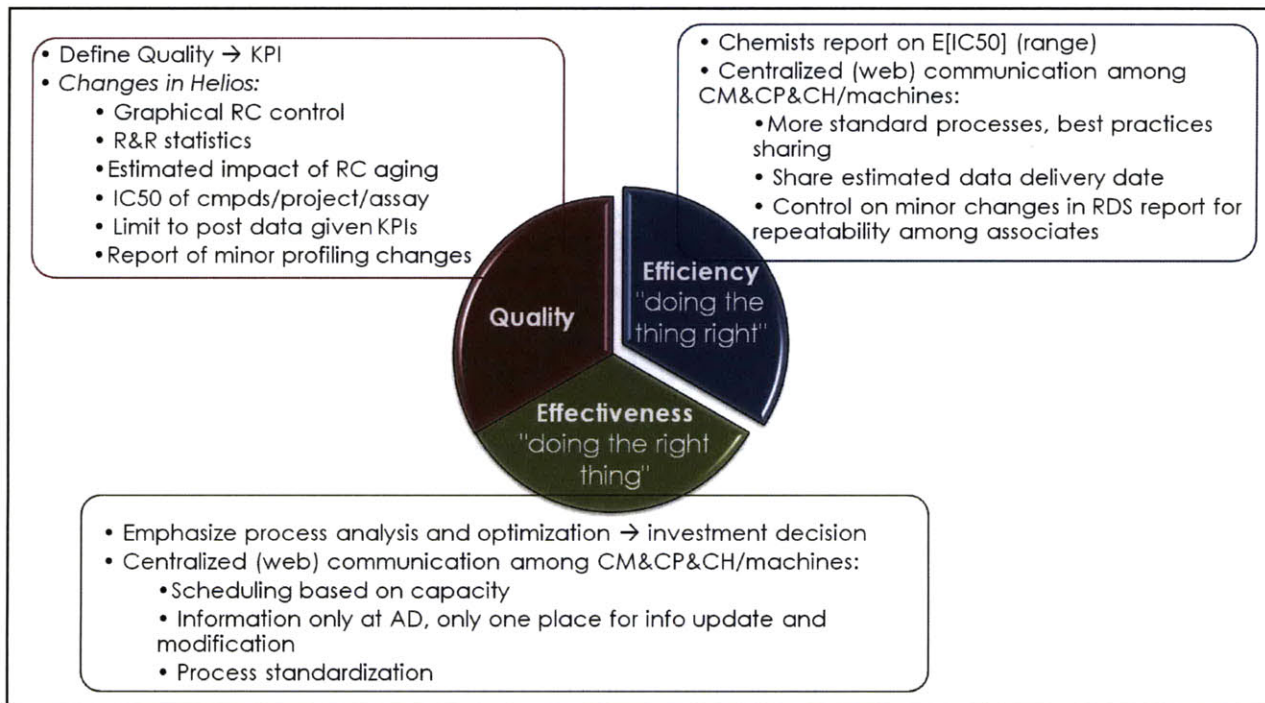
Figure 38. Relation between recommendations the three areas of the project

Lastly, it has to be recognized that EPP/Biology lab has a great advantage of possessing all the knowledge in house. Also, is characterized by associates with defined responsibilities and with the expertise, not only for a great assay profiling but also for equipment support and software tool development. All this together with some process monitoring and control, can lead to sustainable and impactful benefits for the overall R&D length.

6.3. Internal challenges

Even when proposed changes are well defined and results are compelling, application can face with people's personal perceptions and resistance towards change. The aim of the recommendations presented is to provide a framework for improvement utilizing current momentum, but the limitation is that current

momentum is also perceived for some as chaotic and unorganized. Thus, suggestions in this project can be observed as more changes that will confuse further current procedures.

As concluded by many researchers, one of the most important drivers for productivity increase in science-driven companies is motivation[46]. When individual contribution is high and team cohesion is strong, process upgrades are, not only more helpful, but sustainable. In this way, process changes are perceived as a tool for better realizing day-to-day operations and their implementation becomes natural.

Another constraint for improvement is the individual owning of assays. Given that each assay is performed by a particular associate, knowledge sharing can be perceived as lack of freedom or trust. For the changes to be successful, an environment where everyone can measure the benefits from integration is vital, so metrics have to be aligned to show the advantages of boosting team focus.

The last limitation can be observed when standardizing procedures. This standardization often times is perceived as lack of trust, and may be seen as a limitation for innovation. When people are focused on uniqueness of work, having common frameworks can decrease people's motivation and joy for new projects. In order to overcome this mentality, a balanced measurement system, as the one recommended in this project, is key for engaging associates while maintaining the innovation in the teams.

6.4. Future studies

This project is an initiation towards analyzing lab activities from a process perspective. As a result, many areas were identified as potential topics for future projects:

1. *Evaluation of the historic trends of selected KPIs*: Recommended KPIs are a first approximation to process monitoring and control, based on the current data available. In the future, it will be necessary to revisit the impact and applicability of the recommended KPIs to ensure that these are still the best ones to define the lab's performance.

2. *Study of the shift of incentives and/or priorities*: Even when changing current lab KPIs, each drug discovery project relies on a group of people with different backgrounds, decision making procedures and responsibilities. For a real decrease in the total duration of drug discovery, indicators for every stakeholder should be aligned to reflect platform performance. Some of indicators that can help aligning incentives are: number of sPOC/time, number of NMEs in new areas (to boost innovation) and speed of decision making processes (measured even for stopped projects in order to know how useful the data is in

---

[46] (Andreas Sewing, 2008)

making decisions and stopping projects on-time). Studies to evaluate the impact of having more aligned KPIs can give more insight about the platform dynamics and make evident future areas for improvement.

3. *Simulation of benefits*: When noise of the process is reduced and measured outcomes are in control, additional benefits can be obtained from process simulation. This tool will allow observation of critical steps in the flow of activities and help differentiate key areas for improvement. Lastly, this application can show the impact that any variation in the methods can have on the rest of the drug discovery loop.

7. BIBLIOGRAPHY

Andreas Sewing, Toby Winchester, Pauline Carnell, Davod Hampton and Wilma Keighley. "Helping Science to succeed: improving processes in R&D." *Drug Discovery Today* (Elsevier) 13, no. 5/6 (March 2008): 227-233.

Army Business Transformation Knowledge Center. *Transformation Focus-Continuous Process Improvement (CPI).* August 25, 2009. http://www.army.mil/ArmyBTCK/focus/cpi/tools3.htm (accessed January 21, 2010).

Bedford, Jordan M. *Optimization of Chemical Reagent Storage and Distribution at Novartis Institutes for BioMedical Research.* Published MBA/MS Thesis. Leaders for Manufacturing Program, Massachusetts Institute of Technology,Cambridge, MA, 2007.

Boutellier, Fredrik Ullman and Roman. "A case study of lean drug discovery: from project driven research to innovation studios and process factories." (Elseiver) 13, no. 11/12 (June 2008): 543-550.

Boutellier, Fredrik Ullman and Roman. "Drug discovery: are productivity metrics inhibiting motivation and creativity?" (Science Direct) 13, no. 21-22 (November 2008): 997-1001.

Business Dictionary. *Dictionary.* http://www.businessdictionary.com/definition/customer.html (accessed February 7, 2010).

Center, Eli Lilly and Company and the National Institutes of Health Chemical Genomics. *Assay Operations for SAR Support .* 2008.

Chad W. Johnson, et al. "Efficient and Effective Compound Management to Support Lead Optimization." (Journal of Biomolecular Screening) 14, no. 5 (May 2009): 523-530.

Drug Discovery and Development. *Glossary .* http://www.dddmag.com/Glossary.aspx (accessed April 1, 2010).

Eillott, Michael H. "Pharma Asia." *Integrating the Drug Discovery Laboratory.* Reed Business Information Asia. June 1, 2006. http://www.pharmaasia.com/article-5679-integratingthedrugdiscoverylaboratory-Asia.html (accessed November 16, 2009).

Eric Reidenbach. *Market Value Solutions.* 2005. http://www.marketvaluesolutions.com/SixSigmaandtheVoiceoftheCustomer.pdf (accessed September 29, 2009).

Fee, Robert. "The Cost of Clinical Trials." *Drug Discovery and Development* 10, no. 3 (March 2007): 32.

Food and Drug Administration. "Challenges and Opportunities Report." *Science and Research.* August 12, 2009. http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportuniti esReports/ucm077262.htm (accessed March 20, 2010).

Goodman, Elisabeth. "How Lean Can Bring Real Benefits to Innovation in Pharmaceutical Research?"
Six Sigma and Process Excellence. January 8, 2010.
http://www.sixsigmaiq.com/article.cfm?externalid=1720 (accessed March 10, 2010).

Goudarzi, Sara. "Perspective On: A Pharmaceutical Research Lab." *Lab Manager Magazine*, January 6,
2010.

Grant, Warren J. Ewens and Gregory R. *Statistical Methods in Bioinformatics: An introduction.* New
York, New York: Springer, 2001.

Gubler, Hanspeter. "Methods for statistical analysis, quality assurance and management of HTS data."
High-Throughput Screening in Drug Discovery 8 (2006): 151–205.

Health Performance Partners. *8 Wastes with Healthcare Examples.*
http://leanhealthcareperformance.com/page.php?page=8%20Wastes%20with%20Healthcare%20
Examples (accessed January 20, 2010).

Heebner, Peter Gwynne and Gary. "Technologies in Drug Discovery Drug Development: The Next
Generation." *Science*, August 3, 2002.

IBM. *WebSphere Business Modeler.* IBM.
http://public.boulder.ibm.com/infocenter/dmndhelp/v6rxmx/index.jsp?topic=/com.ibm.ntools.hel
p.modeler.bmeasures.doc/doc/concepts/measures/kpis.html (accessed February 10, 2010).

iSix Sigma. http://www.isixsigma.com/index.php?option=com_k2&view=item&id=1275:&Itemid=49
(accessed February 3, 2010).

Nathalie Malo, James Hanley, et al. "Statistical Practice in high-throughput screening data analysis."
(Nature Biotechnology) 24, no. 2 (February 2006): 167-175.

Next Generation Pharmaceutical. "The lean solution." *Next Generation Pharmaceutical*, April 5, 2010:
130-132.

Novartis . "Novartis Media Releases." *Strategy and Financial Highlights Information from ICIS.* January
28, 2009. http://www.icis.com/v2/companies/9146085/novartis-chemicals/financial.html
(accessed September 29, 2009).

Novartis AG. "Novartis Annual Report." Basel, Switzerland, 2009.

—. *Novartis Media Releases.* November 19, 2008.
http://cws.huginonline.com/N/134323/PR/200811/1271207_5.html (accessed November 5, 2009).

Novartis Company Website. http://www.novartis.com (accessed February 10, 2010).

Novartis Technical Operations . "Novartis IQP Presentation." *DavidH_eBPR_Presentation.*

Novartis Technical Operations. "IQP Presentation." *Data_Analysis.*

—. "IQP Presentation." *RCA_Tools_basicsfishbone5whyisisnot.*

—. "Novartis IQP Presentation." *Process Stability.*

—. "Novartis IQP Presentation." *Operational_Risk_Assessment-FMEA.*

—. "Novartis IQP Presentation." *IoannaK_ IQPchampions_qualimetrics.*

—. "Novartis IQP Presentation." *Getting_started_with_Qualimetrics_processmapping.*

—. "Novartis IQP Presentation." *Discrete_Data_ControlCharts.*

—. "Novartis IQP Presentation." *Qualimetrics_IQP.*

—. "Novartis IQP Presentation." *Qualimetrics_IQP.*

—. "Novartis IQP Presentation." *Capability Analysis - Cp Indices.*

Pollok, Brian. "Assay development: an increasingly creative endeavour." *Nature Reviews Drug Discovery* 4 (December 2005): 956-957.

Rice, John A. *Mathematical Statistics and Data Analysis.* Third. Belmont, CA: Thomson Brooks/Cole, 2007.

Rosenfield, Sara Beckman and Donald. *Operations Strategy.* New York: Mc Graw Hill, 2008.

Steven Paul, Daniel Mytelka, et al. "How to improve R&D productivity: the pharmaceutical industry's grand challenge." *Nature Reviews Drug Discovery* 9 (March 2010): 204-214.

Trevor Hastie, Robert Tibshirani and Jerome Friedman. *The Elements of Statistical Learning.* Second. New York, New York: Springer, 2009.

United States Department of Labor. *Bureau of Labor Statistics.* December 17, 2009. http://www.wordwebonline.com (accessed January 27, 2010).

Wood, Simon. "Lean Principles of Process Improvement." *IPTonline.* August 31, 2006. http://www.iptonline.com/articles/public/page787980loresnonprint.pdf (accessed March 20, 2010).

World Health Organization (WHO). *The world medicines situation.* WHO, 2004.

8. APPENDIX

Appendix A: Compound Manager Detailed Process Map

PROCESS MAP-EPP/BIO/COMPOUND MANAGER (1)

START

1. TAKE COMPOUNDS TO EPP/BIO LAB
- Cmpds arrive to fridge from NCA
- 1.1. Go downstairs to the fridge
- 1.2. Take rack for Bio Lab
- 1.3. Bring rack to Bio Lab

2. RACK PREPARATION
- Are cmpds in flask or rack? — Rack → Are cmpds first columns? — YES → NO → Sort tubes manually
- Flask → Place and sort tubes on rack

3. RACK SCAN
- Cmpds in rack
- 3.1. Label rack temporarily
- 3.2. Scan temporary label from Entry rack (saved automatically in shared drive)
- 3.3 Scan of cmpds's 2D barcodes
- 3.4 File generated is saved in CM's folder
- (Backup file of cmpds position in the rack)

4. TRT
- File with scanned cmpds
- 4.1. Open TRT and create worklist
- 4.2. Browse for file with scanned barcodes
- 4.3. Copy and paste 2D barcodes in TRT, then organize by submitted list
- 4.4. Check that # of cmpds in TRT = # cmpds in scanned rack file
- 4.5. Export TRT generated table to Excel and Copy/paste in macro for *correct tube sorting* for robot

5. PREP FOR TUBE SORTING
- File with *correct cmpd order*
- 5.1. Open tube sorting software and browse file with cmpd's correct order
- 5.2. Rename the document (needed for automatic sorting)
- 5.3. Use manual barcode reader to scan source racks
- 5.4 Take needed destination racks from the drawer and scan their position before placing them
- 5.5. Repeat scanning of destination racks until all panels are in the robot

6. AUTOMATIC TUBE SORTING
- Position of cmpds in source rack and position of destination racks

7. SORTING PROBLEM SOLVING
- Were all cmpds sorted? — NO → Find out specific reasons why cmpds are not sorted — Repeated until solution is found and cmpd is placed on correct storage rack and position
- YES → Place and sort tubes on rack

8. TRANSFER CMPDS FOR SORTING TO STORAGE RACK
- Sorting racks with cmpds for each assay panel
- 8.1. Save file generated by tube sorter as .txt
- 8.2. Pick up storage rack from a fridge in EPP/Bio lab
- 8.3. Change the tubes from sorting to storage racks changing from horizontal to vertical sequence, also observing if cmpds are in solution
- 8.4. Place sorting rack again in drawer and source and destination racks back in the fridge
- 8.6. Take out temporary label from entry rack
- (Cmpds organized in racks for each assay panel)

9. TRT
- File with sorted cmpds
- 9.1. Copy/Paste cmpd's codes in TRT
- 9.2. Select all cmpds as "arrived"
- 9.3. Change status of all cmpds to "accepted"
- (Change in cmpd's TRT status to accepted)

10. UPDATE INTERNAL CMPD LIST
- Excel file with cmpd tracking
- 10.1. Browse in shared drive for the excel file
- 10.2. Copy/Paste info of cmpds accepted
- 10.3. Change txt file name to show that day's date (signal that that day's cmpds are sorted and status changes in TRT)
- (Document in shared folder with date's input so that CP know that cmpd status is up to date)

11. DATA PREPARATION FOR CMPD DILUTION
- File with cmpd position in each rack
- 11.1 Browse shared drive for CP's request sheet (DRD request sheet)
- 11.2. Copy/Paste file in CM's personal folder adding CM's initials to file name
- 11.3. Browse Master Plate sequence file and Copy/Paste cmps info
- 11.4. Fill spaces for CM in DRD sheet
- 11.5. Print another label for the rack, for the long side (short side has label)
- 11.6. Scan rack again to verify data introduced by the CP in DRD sheet
- 11.7. Copy/Paste cmpd's codes in 2nd tab of DRC sheet
- 11.8. Use macro to order cmpds correctly (corresponding to place in rack)
- 11.9. Print temporary label for work rack
- 11.10. Scan source rack and work rack's label to enter codes in DRC sheet
- 11.11. Use macro to relate cmpd's position in source rack with future position in work rack
- 11.12. Save this last work rack list in another document (.csv) and change name using barcode reader so that the name is the barcode of the working rack
- 11.13. Save work rack document in shared drive
- 11.14. Print a hard copy of DRC final sheet
- (Hardcopy of DRC request sheet with information about cmpds in working rack, date, dilution concentration)

CM-2

CM-2

12. COMPOUND PREPARATION FOR DILUTION

Cmpd's tubes and master plates

12.1. Centrifuge
12.2. Use Multiprobe tool to transfer cmpds from tube to master plate position
12.3. Use CyBio® for DMSO addition to each well for dilution
12.4. Centrifuge

Master plate with first concentration of cmpds in each well and empty wells with DMSO

1.

2.

13. AUTOMATED DRC

Master plate with Initial cmpd concentration

Residual quantities of cmpds in tube

14. PLACE CMPD REMAINDER IN FRIDGE

14.1. Take tubes with cmpd remainder and place them in fridge

DRC curves for each cmpd and Backup of cmpds in case retest is needed

Is plate ready? —NO→ Wait

YES

15. TAKE PLATE WITH CMPDS DRC TO THE FRIGDE

Plate with cmpd's DRC for each panel

15.1. Place plate with DRC in correspondent storage place depending on panel

DRC plates for each panel

Are all CP's DRC finished? —NO→ Back to Step12.

YES

Does CP scheduled profiling for the same day ? —NO→ FINISH

YES

16. PLATE COPY

DRC in the fridge

16.1. Take DRC plate out of the fridge
16.2. Enter data in Hummingwell® device
16.3. Copies will be generated automatically by the device

Copies of DRC for specific panel

17. FINISH DATA ENTRY

17.1. Enter data manually in DRC hardcopy(date, number of copies, ect)
17.2. Enter data in DRC excel file
17.3. Save modified file in outbox folder in shared drive and another copy in CM's personal folder
17.4. Place copies in the robotic system where CP will perform assay

DRC document in outbox telling CP that copies are ready . Hardcopy of DRC with all data too

FINISH

Appendix B: Compound Profiler Detailed Process Map

PROCESS MAP-EPP/BIO/COMPOUND PROFILER (1)

PROCESS MAP-EPP/BIO/COMPOUND PROFILER (2)

CP-2

Cmpd and plate barcodes

10. WORKLIST CREATION

10.1. Register plate with barcode
10.2. Copy/paste cmpd barcodes into TRT to generate worklist
10.3. Select all cmpds and change status to "inprogress"

Status of cmpds changes to "in-progress" and a work list for data analysis

Cmp and plate barcodes, assay characteristics (concentrations, plate layout etc)

11. PLATE REGISTRATION IN AEOS

11.1 Copy/paste cmpd barcodes in ICP
11.2. Change all default starting concentration if different with the one used in the assay
11.3. Check manually plate layout in screen, change if necessary
11.4. Search and select the panel
11.5. Manually name and create plate
11.6. With manual barcode reader, scan barcode of master plate

Status of cmpds changes to "in-progress" and a work list for data analysis

Plate's barcode, folder where raw data is located

12. PREPARATION FOR DATA ANALYSIS

12.1. Go back to ICP
12.2. Select the reader used of assay and then select get plate file
12.3. Open explorer to get data from device manually typing plate's barcode
12.4. Select send data, ICP generates one file per plate profiled
12.5. Open Helios
12.6. Browse assay panel and select created plate group typing plate group name
12.7. Find individual plates that will create the group and generate the plate group

Plate group to analyze all plates at a time

Expected IC 50 values for reference cmpds

13. DATA ANALYSIS

13.1. Visually evaluate reference cmpd's curves and I50 values, compare with expected
13.2. Visually analyze tested cmpds' curves
13.3. Depending on individual procedure, evaluate cmpds' curve validity

Analysis of resulted curves for cmpds tested

Are cmpds too potent and curves can't be analyzed? — YES → **14a. CMPD DILUTION** → **Back to step1a.**

NO

14. DATA SUBMISSION

14.1. Select all curve data and click on "set to publish"
14.2. Un select reference cmpds and any other data that won't be published
14.3. Store final results in databases
14.4. Export the file to excel
14.5. Open Pharon, copy and paste exported data to excel and save file to post data in Avalon
14.6. Save Excel file in CP personal folder
14.7. Copy raw data in excel to generate another report for chemists
14.8. Send email around to all team with last excel report

Cmpd data available for chemists

Does the assay needs to be repeated? — YES → *Is chemist generating another request or cmpd batch?* — YES → **Back to step1.**

NO → (FINISH)

NO → **Back to step1a.**

Appendix C. Codes used in Process Mapping

FIGURE	SHAPE KEY
	DEFINITION
	Manual operation or copy/paste of data
	Decision points
	Standard/established operation
	Preparation
	Manual operation not critical to the process
	Manual operation that can impact the process outcome
	Sum/junction of steps
	Manual operation that requires move from work station
	External link
	Manual Input of info/data
	Waiting time
	A document is generated or used
	Starting or ending points or denotes inputs or outputs in activities when accompanied by a dotted arrow
	Sort
	Process that has to occur in order to continue activities
	Feedback to a previous activity
	Shows inputs or outputs that are currently needed for a particular activity

Appendix D. Fishbone diagram for Turnaround time variation (mean and range)

PEOPLE

EPP projects will be done first than rare requests from other DAs

Projects with less priority have higher profiling times

Each Lab Head may has different preferences and that changes CP schedule

Preference for some projects

MATERIALS

CP do things depending on Lab Head

Need for "innovation"

Everybody has their own procedure

Assay development activities may have priority depending on Lab Head

No specific schedule for profiling

CP schedule assays depending on their workload

EQUIPMENT

Booking is done in half days, not a calculation of the real duration of the assays

Sometimes people don't use them and never update data

Robots can be booked for days

No communication channel when robot is down

Too many constant changes in software

Equipment is down

WHY TRUNAROUND TIMES ARE UNSTABLE AND UNPREDICTABLE?

Every CP fills in different way

No better way to communicate

Filling out requests from CP to CM take time

Copy and paste or double entry needs leads to errors

Check in all excel files to find cause of errors

Errors in DRC requests take time to solve

Out of the established process need "tweaks" for system

Direct requests from chemists make CP and CM loose time

METHODS **ENVIRONMENT**

Appendix E. Statistical Analysis for EPP/Biology turnaround times

Moods Median Test for Type of screening:

Mood Median Test: Time in EPP/Biology lab versus Type of screening

```
Mood median test for Y2-TIME IN EPP/BIO
Chi-Square = 4007.46    DF = 47    P = 0.000

X2-Type of                               Individual 95.0% CIs
screening   N<=    N>    Median   Q3-Q1  +---------+---------+---------+------
As1         143    196    5.0      9.0    *
As2           1      2   10.0     23.0    (-*------)
As4           9      6    1.0      4.0   *-)
Cy16          9    102    8.0      1.0       *
Cy4          38     49    6.0      8.0    *-)
Cy7         155    147    4.0      8.0    *
Cy8           0    165    8.0      1.0      *
Cy9         164     10    1.0      3.0   *
EP1           0      2    6.5      3.0     (*
EP10         33    210    5.0      4.0    *
EP11         17     15    4.0      2.0    (*
EP12          6      0    2.5      2.3    (*
EP13          5      0    2.0      1.0    *
EP14          0      1    7.0   Not Used
EP15        119     25    2.0      2.0    *
EP16         10      3    4.0      2.0    (*
EP17         11      1    2.5      1.0    *
EP18          3      0    4.0      1.0    (*
EP19         13     48    6.0      4.0     *)
EP20         17      0    3.0      0.5    *
EP21         49     40    4.0      2.0     *
EP22          9     15    5.0      1.8     *
EP23         36     39    5.0      2.0     *
EP24        149     80    4.0      3.0     *
EP25         76     55    4.0      4.0     *
EP26         59     54    4.0     10.0     *)
EP27         10      9    4.0      4.0   (-*
EP3           2     21    6.0      1.0     *
EP4           6      4    3.0      5.3   *-)
EP5           8     13    6.0      3.0    (*
EP6          23     18    4.0      2.0     *
EP7          34     41    5.0      2.0     *
EP8          88     83    4.0      2.0    *
EP9          18      5    1.0      3.0   *
Me1           0      1   38.0   Not Used
Me10         72      0    1.0      0.0   *
Me11         77      8    4.0      1.0     *
Me2          59     20    2.0      4.0   (*
Me6          90      0    2.0      0.0   *
Me7         446    117    3.0      3.0   *
Me9         834    185    2.0      3.0   *
Pr1         299    841    7.0      7.0    (*
SE1           5      2    4.0      2.0   (*)
Se13       1943    181    2.0      1.0   *
Se14       6363   4973    4.0      5.0    *
Se16          2      1    2.0     86.0   *--------------------------------)
Se17         37      2    2.0      1.0   (*
Se18        622    142    2.0      2.0   *
Se4           0     33   20.0     22.0          *-----)
Se5         324   1312   10.0     12.0     *)
                                         +---------+---------+---------+------
                                         0        25        50        75

Overall median = 4.0
```

Moods Median Test for Protocol type:

Mood Median Test: Time in EPP/Biology versus Protocol Code

```
Mood median test for Y2-TIME IN EPP/BIO
Chi-Square = 878.82    DF = 1    P = 0.000

X4-PROTOCOL                              Individual 95.0% CIs
CODE          N<=     N>   Median  Q3-Q1  -------+---------+---------+---------
P            2143    384    2.00    2.00  *
S           10350   8893    4.00    6.00                                      *

                                         -------+---------+---------+---------
                                             2.40      3.00      3.60
|
Overall median = 4.00


A 95.0% CI for median(P) - median(S): (-2.00,-2.00)
```

Moods Median Test for Profile type:

Mood Median Test: Y2-TIME IN EPP/BIO versus X3-PROFILE CODE

```
Mood median test for Y2-TIME IN EPP/BIO
Chi-Square = 2897.53    DF = 30    P = 0.000

X3-PROFILE                             Individual 95.0% CIs
CODE       N<=     N>   Median  Q3-Q1  -------+---------+---------+---------
EP1         22      0    1.0     0.0   *
EP10        55     73    5.0     9.0              (---*-------)
EP11        33    210    5.0     4.0                    *--)
EP12         5      0    2.0     1.0       *--)
EP13       137     31    2.0     2.0       *
EP14       830    180    2.0     3.0       *
EP15        10      3    4.0     2.0         (---*)
EP16        82     90    7.0    11.0          (-------------*---------)
EP17       741    575    4.0     5.0              *
EP18        13     48    6.0     4.0                  (-*--)
EP19       188    162    4.0     2.0             *---)
EP2          0      2    6.5     3.0                (----*----)
EP20        66     40    4.0     3.0             *
EP21         9     15    5.0     1.8             (---*
EP22       866    164    2.0     2.0        *
EP23       185    119    4.0     3.0             *
EP24       135    109    4.0     6.8            *---)
EP25        10      9    4.0     4.0   (---------*---)
EP26        54     38    3.0     3.0          *------)
EP27       402     24    2.0     3.0        *
EP28       719    690    4.0     7.0            *---)
EP4          2     21    6.0     1.0               (-*
EP5        353    462    6.0     7.0               (--*--)
EP6        271   1265   12.0    13.0                                     (--*
EP7       4614   2341    3.0     4.0          *
EP8       2379   1750    4.0     5.0         (--*
EP9          8     13    6.0     3.0         (---------*
Pr5         67    224    7.0     7.0                    (-*
Pr6         93    231    6.5     6.0                    (-*)
Pr7        139    386    7.0     7.0                    (--*
SE1          5      2    4.0     2.0          (*-------)
                                       -------+---------+---------+---------
```

Moods Median Test for type of request (internal vs. external):

Mood Median Test: Y2-TIME IN EPP/BIO versus X1-REQ GROUP

```
Mood median test for Y2-TIME IN EPP/BIO
Chi-Square = 592.11    DF = 2    P = 0.000

X1-REQ                              Individual 95.0% CIs
GROUP    N<=    N>   Median  Q3-Q1   +---------+---------+---------+------
EPP/BIO  4124  1703   2.00    3.00   *
EPP/MCH  5360  4679   4.00    5.00                       *
OTHER    3009  2895   4.00    6.00                          *---------)
                                     +---------+---------+---------+------
                                    2.0       3.0       4.0       5.0

Overall median = 4.00
```

Moods Median Test for Phase of Project:

Mood Median Test: Y2-TIME IN EPP/BIO versus X7-PHASE

```
Mood median test for Y2-TIME IN EPP/BIO
Chi-Square = 801.11    DF = 9    P = 0.000


X7-PHASE                     N<=     N>   Median  Q3-Q1
D0                           922    323    2.00    4.00
D1                          1001    310    3.00    2.00
D2a                         2916   1747    3.00    5.00
D2b                         5500   4128    4.00    5.00
D2b->D3 (LEAD Candidate)     297    388    5.00    6.00
D3                          1501   2052    5.00    8.00
D3->CSP (CSP Candidate)      334    274    4.00    5.00
D4                            11     25    9.00   11.00
DevSupport                    3      9    6.50    2.75
ReferenceCpd                  8     21    6.00    4.50

                             Individual 95.0% CIs
X7-PHASE                     ---+---------+---------+---------+---
D0                           *--)
D1                           (--*
D2a                            *
D2b                              *
D2b->D3 (LEAD Candidate)          *--)
D3                                *
D3->CSP (CSP Candidate)       (--*
D4                                     (---------*---------)
DevSupport                        (-------*)
ReferenceCpd                       (--*---)
                             ---+---------+---------+---------+---
                             3.0       6.0       9.0       12.0

Overall median = 4.00
```

Moods Median Test for location (country):

Mood Median Test: Y2-TIME IN EPP/BIO versus COUNTRY

```
Mood median test for Y2-TIME IN EPP/BIO
Chi-Square = 226.75    DF = 5    P = 0.000

                                        Individual 95.0% CIs
COUNTRY    N<=     N>   Median  Q3-Q1    ---------+----------+----------+---------
AT           0     30   12.00   7.00           (----------------------------*
CH       11256   8435    4.00   5.00      *
CN           2      2    5.50   6.00      (-----*----------------------)
GB          38    141    7.00   3.00             (---*
SG          13     53    7.00   2.25             (---*
US        1184    616    3.00   5.00     *
                                         ---------+----------+----------+---------
                                              5.0        7.5       10.0

Overall median = 4.00
```

Mood Median Test: Y3-TOTAL TIME versus COUNTRY

```
Mood median test for Y3-TOTAL TIME
Chi-Square = 1031.44    DF = 5    P = 0.000

                                        Individual 95.0% CIs
COUNTRY    N<=     N>   Median  Q3-Q1    +---------+---------+---------+------
AT           0     30   16.0    7.0            (----------------------*
CH       10699   8992    6.0    6.0      *
CN           0      4    9.0    6.3        (-------*----------------)
GB          16    163   11.0    5.0               (---*
SG           9     57    9.0    4.0         (--*------)
US         339   1461    9.0    5.0         (--*
                                         +---------+---------+---------+------
                                         6.0       9.0      12.0      15.0

Overall median = 6.0
```

Appendix F. Minitab output for Process Capability study

EPP/Biology Lab

Xbar Chart — Sample Mean: UCL=6.74, X̄=6.00, LCL=5.26

Capability Histogram — Specifications: LSL 1, USL 10

S Chart — Sample StDev: UCL=7.384, S̄=6.861, LCL=6.339

Last 21 Subgroups — Values

Capability Plot:

Overall	
Shape	1
Scale	6.00051
Pp	0.23
Ppk	0.16
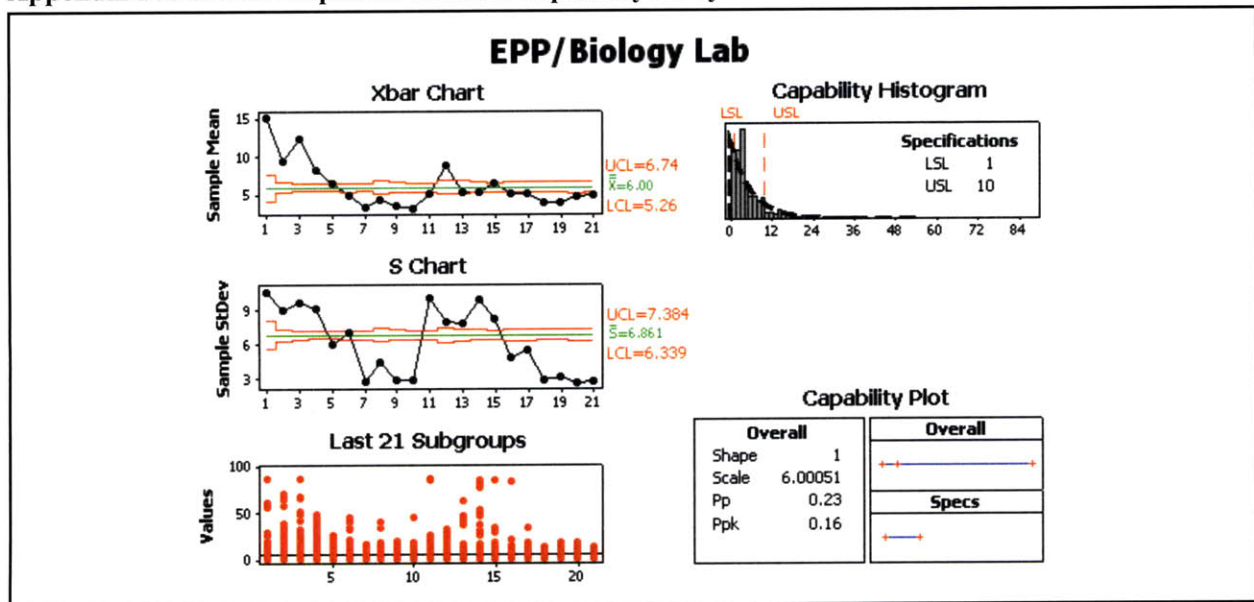
Appendix G. Fishbone diagram for Quality of data

MATERIALS

Batch is known to be aged after test

RC is aging

PEOPLE

When AC50 of RC is not clear

High UCL/LCL permits to work with no flags

Projects in late phases require more accuracy than cmpds in early project phases

Lab technicians use different tools to know if assay is valid

No standard criteria, depends on chemist and team

Projects/assays differ in acceptance criteria

EQUIPMENT

Robot dispensed incorrect amounts

Readouts are affected by assay conditions

WHY RC DATA QUALITY CAN FAIL?

No automatic system to observe trends

Manual tracking is required

Some assays need a quick test and data can't be thoroughly checked

Urgent requests

No RC available for the conditions needed

Assay is urgent, no time to test RC as needed before profiling

Assay is new and RC behavior is not well defined

Data not available

If AC50 is out of range, batch of RC is changed for next assay but the current stays in the database

RC batch is changed after observed data

METHODS

ENVIRONMENT

Appendix H. Mood's Median Test for AC50 of selected

With respect to the experiment date

Results for: C1 in assay 62-1

Mood Median Test: Qualified AC50 versus Experiment date

```
Mood median test for Qualified AC50
Chi-Square = 20.38    DF = 7    P = 0.005


Experiment date  N<=  N>  Median   Q3-Q1
3-Mar-2009        4    0  0.0004   0.0005
10-Mar-2009      13    7  0.0007   0.0002
11-Mar-2009       0    2  0.0009   0.0000
16-Mar-2009       0    4  0.0011   0.0005
18-Mar-2009      13    7  0.0006   0.0003
25-Mar-2009       5    8  0.0008   0.0006
1-Apr-2009        5    6  0.0008   0.0006
8-Apr-2009        0    6  0.0010   0.0003


                    Individual 95.0% CIs
Experiment date  +---------+---------+---------+------
3-Mar-2009       (-------*----)
10-Mar-2009                 (-*)
11-Mar-2009                      *
16-Mar-2009                           (*-----------)
18-Mar-2009             (-*--)
25-Mar-2009          (----*-----)
1-Apr-2009               (-*---------)
8-Apr-2009                  (---*-----)
                 +---------+---------+---------+------
              0.00000   0.00050   0.00100   0.00150
```

Results for: C2 in assay 88-1

Mood Median Test: Qualified AC50 versus Experiment date

```
Mood median test for Qualified AC50
Chi-Square = 39.14    DF = 17    P = 0.002


                    Individual 95.0% CIs
Experiment date  -------+---------+---------+---------
4-Dec-2008       (---------------*----------------)
17-Dec-2008
18-Dec-2008      (------*-----------------)
16-Jan-2009         (*-----)
23-Jan-2009             (*--)
3-Feb-2009          (*)
10-Feb-2009        (-*)
16-Feb-2009         (*)
24-Feb-2009       *)
27-Feb-2009         (*)
4-Mar-2009        (*)
17-Mar-2009          (---*-)
24-Mar-2009       (-*)
1-Apr-2009         (*)
6-Apr-2009        (-*)
17-Apr-2009       (----*-)
21-Apr-2009      (--*--)
22-Apr-2009      (*)
24-Apr-2009        (-*-)
30-Apr-2009      -------+---------+---------+---------
                     3.0      6.0      9.0
```

With respect to the operator

Results for: C1 in assay 62-1

Mood Median Test: Qualified AC50 versus Operator

```
Mood median test for Qualified AC50
Chi-Square = 6.49    DF = 1    P = 0.011

                              Individual 95.0% CIs
Operator  N<=  N>  Median   Q3-Q1  ------+---------+---------+---------+
A          40  34  0.0007   0.0004 (-*--)
B           0   6  0.0010   0.0003       (--------*----------------)
                                   ------+---------+---------+---------+
                                    0.00080  0.00100  0.00120  0.00140

Overall median = 0.0008
```

Results for: C5 in assay 70-1

Mood Median Test: Qualified AC50 versus Operator

```
Mood median test for Qualified AC50
Chi-Square = 4.21    DF = 1    P = 0.040

                              Individual 95.0% CIs
Operator  N<=  N>  Median   Q3-Q1  +---------+---------+---------+------
C          40  36  1.53     1.59   (--*)
D           0   4  3.42     2.21          (-----*--------------)
                                   +---------+---------+---------+------
                                   1.2      2.4      3.6      4.8

Overall median = 1.55
```

With respect to substrate batch
Results for: C2 in assay 88-1

Mood Median Test: Qualified AC50 versus Substrate Batch

```
Mood median test for Qualified AC50
Chi-Square = 0.75    DF = 1    P = 0.387


Substrate Batch    N<=  N>  Median  Q3-Q1
BS-1               6    9   4.13    3.57
BS-2               32   29  3.06    1.14


                   Individual 95.0% CIs
Substrate Batch    --------+---------+---------+---------
BS-1               (--------------------*-----------)
BS-2                               |(-*-)
                   --------+---------+---------+---------
                        3.0       4.0       5.0
```
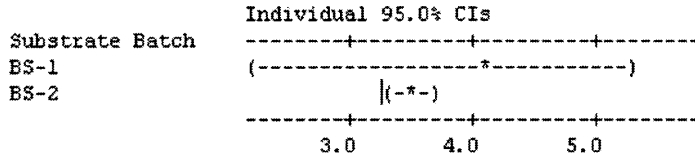
Overall median = 3.10

Results for: C4 in assay 90-1

Mood Median Test: Qualified AC50 versus Substrate Batch

```
Mood median test for Qualified AC50
Chi-Square = 21.27    DF = 1    P = 0.000

                                      Individual 95.0% CIs
Substrate Batch   N<=  N>  Median  Q3-Q1   ----+---------+---------+---------+--
BS-3              21   2   0.061   0.048  (--*)
BS-4              20   36  0.248   0.121                       (---*---)
                                          ----+---------+---------+---------+--
                                          0.070     0.140     0.210     0.280
```

Overall median = 0.209

With respect to timestamp
Results for: C4 in assay 90-1

Mood Median Test: Qualified AC50 versus Readout timestamp

```
Mood median test for Qualified AC50
Chi-Square = 52.33    DF = 18    P = 0.000


Readout                             Individual 95.0% CIs
timestamp N<=  N>  Median  Q3-Q1    +---------+---------+---------+------
04:28.0   0    4   0.247   0.043             (*-)
04:30.0   0    4   0.306   0.052              (*--)
10:18.0   2    2   0.175   0.507    (-------*--------------------)
12:25.0   3    1   0.183   0.082          (*---)
17:05.0   0    4   0.334   0.030                (*)
20:38.0   2    0   0.023   0.009    *
24:31.0   4    0   0.013   0.028    (*)
27:36.0   0    4   0.274   0.062              (*-)
28:08.0   2    2   0.235   0.075            (-*-)
32:12.0   4    0   0.086   0.018      *)
42:23.0   3    0   0.025   0.039    (*)
44:42.0   1    3   0.225   0.038            (*-)
46:27.0   4    0   0.060   0.016    (*
46:45.0   7    5   0.167   0.219      (---*------)
48:06.0   0    4   0.451   0.126                 (-------*
48:10.0   4    0   0.066   0.010     *)
52:20.0   1    3   0.231   0.109           (---*--)
57:51.0   0    4   0.251   0.040            (*-)
58:26.0   4    0   0.152   0.061        (-*)
                                    +---------+---------+---------+------
                                    0.00      0.20      0.40      0.60
```

Overall median = 0.209

Mood Median Test: Qualified AC50 versus Readout timestamp

```
Mood median test for Qualified AC50
Chi-Square = 39.64    DF = 19    P = 0.004

Readout                                 Individual 95.0% CIs
timestamp  N<=  N>  Median    Q3-Q1   -------+---------+----------+---------
04:04.0     0   4    3.7      0.5             (-*)
07:30.0     4   0    2.2      0.3         *)
07:36.0     1   0    2.3    Not Used
07:44.0     0   2    4.2      1.1                (-*-)
11:12.0     2   1    2.9      2.1             (*-----)
11:58.0     3   1    2.8      0.9          (-*-)
17:49.0     0   4    4.4      1.1                 (*--)
20:56.0     4   0    2.8      0.5           (*)
22:09.0     3   1    2.7      0.9          (-*)
25:47.0     2   0    2.5      0.4          (*)
34:39.0     2   2    3.0      0.5           (*)
36:22.0     1   3    3.4      0.7            (*)
38:22.0     1   0    1.9    Not Used
38:40.0     1   2    6.1      9.8      (---------------*----------------)
39:22.0     1   3    4.0      1.8          (----*-)
41:02.0     2   2    3.0      6.9     (-------*------------------)
42:45.0     3   1    2.6      1.7        (--*--)
43:07.0     0   4    3.7      1.2              *----)
48:04.0     1   3    3.3      0.4          (*)
50:55.0     4   0    2.4      0.5          (*)
51:12.0     0   4    5.4      0.9              (--*)
59:35.0     3   1    3.0      0.5          (*)
                                        -------+---------+----------+---------
                                        3.0       6.0        9.0


Overall median = 3.1
```

With respect to experiment time

Results for: C1 in assay 62-1

Mood Median Test: Qualified AC50 versus Experiment time (RED TO HOUR)

```
Mood median test for Qualified AC50
Chi-Square = 8.11    DF = 4    P = 0.088

Experiment
time (RED                            Individual 95.0% CIs
TO HOUR)  N<=  N>  Median   Q3-Q1   --------+----------+----------+--------
11          2   0  0.0006  0.0001      (--*-)
12         25  29  0.0008  0.0004          (----*-----)
13          0   4  0.0009  0.0002               (----*-------)
16          9   5  0.0005  0.0003   (--*-------------)
17          4   2  0.0007  0.0003     (---------*---------)
                                    --------+----------+----------+--------
                                    0.00060   0.00080   0.00100
```

Results for: C2 in assay 88-1

Mood Median Test: Qualified AC50 versus Experiment time (RED TO HOUR)

```
Mood median test for Qualified AC50
Chi-Square = 11.76    DF = 6    P = 0.068

Experiment
time (RED                           Individual 95.0% CIs
TO HOUR)  N<=  N>  Median  Q3-Q1   -------+---------+----------+---------
11          1   3   3.52   1.72         (-----*----------------)
12         17   9   2.83   1.34     (--*---)
13         11   7   2.79   2.13   (----*------------)
14          4   3   3.06   1.81     (-----*------------------)
15          2   3   3.24   3.02   (--------*----------------------)
16          1   9   3.70   1.31          (----*-------)
17          2   4   3.33   0.88     (------*-------)
                                   -------+---------+----------+---------
                                   3.0       4.0        5.0
```

Results for: C4 in assay 90-1

Mood Median Test: Qualified AC50 versus Experiment time (RED TO HOUR)

```
Mood median test for Qualified AC50
Chi-Square = 18.00    DF = 7    P = 0.012

Experiment
time (RED                         Individual 95.0% CIs
TO HOUR)   N<=  N>  Median  Q3-Q1  ------+---------+---------+---------+
11          2   0   0.123   0.022     (*)
12          9   5   0.178   0.125       (---*------)
13         13  13   0.214   0.209   (------------*--)
14         13   4   0.067   0.125  (-*-)
15          1   1   0.251   0.103          (---*---)
16          2   8   0.281   0.088          (---*---)
17          1   7   0.315   0.183            (-------*-----------)
18          0   2   0.334   0.015                      (*
                                   ------+---------+---------+---------+
                                      0.12      0.24      0.36      0.48
```

Results for: C5 in assay 70-1

Mood Median Test: Qualified AC50 versus Experiment time (RED TO HOUR)

```
Mood median test for Qualified AC50
Chi-Square = 18.75    DF = 6    P = 0.005

Experiment
time (RED                         Individual 95.0% CIs
TO HOUR)   N<=  N>  Median  Q3-Q1  ------+---------+---------+---------+
11          1   7   2.22    0.66                   (------*--)
12          0   4   2.26    0.85                  (-------*---)
13          6   4   1.53    0.89        (---------*-)
14         17   9   1.27    1.31     (---------*---)
16          8  10   1.83    1.66           (-------*-------)
17          7   1   0.50    0.33   (-*---)
21          1   5   2.53    1.23                (------------*-------)
                                   ------+---------+---------+---------+
                                      0.80      1.60      2.40      3.20
```

With respect to assay plate name
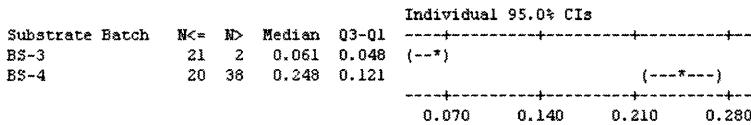**Results for: C3 in assay 19-1**

Mood Median Test: Qualified AC50 versus Assay plate name

Mood median test for Qualified AC50
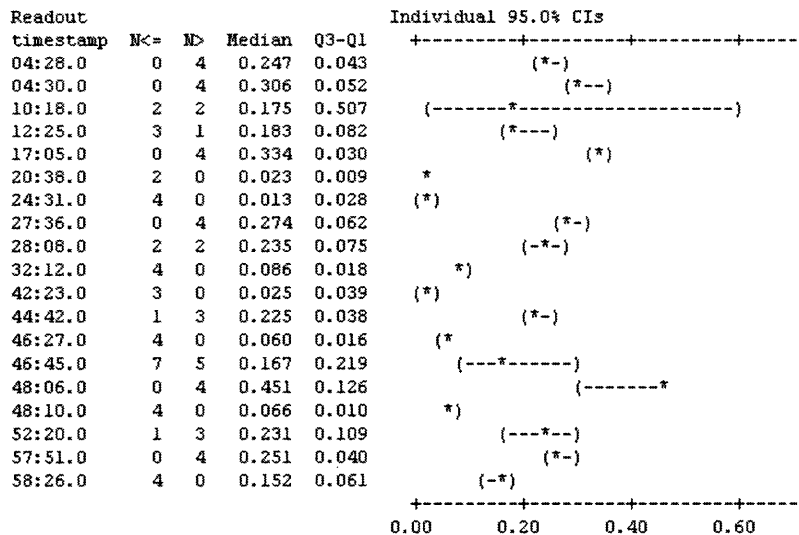Chi-Square = 68.00   DF = 35   P = 0.001

```
Assay                            Individual 95.0% CIs
plate name  N<=  N>  Median  Q3-Q1   +---------+---------+---------+------
PP00000049   2   0   0.058   0.006          (*
PP00000050   0   2   0.105   0.064             (---------*--------)
PP00000051   2   0   0.065   0.001          *)
PP00000052   1   1   0.076   0.010            (-*)
PP00000053   2   0   0.067   0.001           *
PP00000054   0   2   0.101   0.008                 (*)
PP10000121   2   0   0.042   0.003        *)
PP10000122   2   0   0.043   0.001        *
PP10000123   2   0   0.050   0.003         *)
PP10000124   2   0   0.050   0.002         *)
PP10000125   2   0   0.049   0.010       (-*)
PP10000126   2   0   0.049   0.009       (*)
PP10000127   2   0   0.049   0.012      (-*-)
PP10000128   2   0   0.046   0.009      (*)
PP10000154   2   0   0.041   0.004      (*
PP10000155   2   0   0.039   0.009    (*-)
PP10000230   4   0   0.057   0.023      (--*---)
PP10000231   2   0   0.059   0.014        (-*-)
PP10000232   0   2   0.084   0.001              *
PP10000233   0   4   0.083   0.008            (--*
PP10000276   0   2   0.108   0.010                 (*)
PP10000277   0   2   0.108   0.003                 (*
PP10000278   0   2   0.107   0.013                (-*-)
PP10000279   0   2   0.102   0.008               (*)
PP10000359   1   1   0.073   0.001             *
PP10000360   0   2   0.095   0.009              (*)
PP10000438   1   1   0.068   0.013        (*-)
PP10000563   0   2   0.126   0.043                    (-----*-----)
PP10000564   0   2   0.129   0.046                    (------*-----)
PP10000565   0   2   0.100   0.019               (--*-)
PP10000566   0   2   0.104   0.014               (-*-)
PP10000664   0   2   0.120   0.031                   (---*----)
PP10000665   1   1   0.087   0.035             (-----*----)
                                    +---------+---------+---------+------
                                  0.035     0.070     0.105     0.140
```

Overall median = 0.073

Results for: C5 in assay 70-1

Mood Median Test: Qualified AC50 versus Assay plate name

```
Mood median test for Qualified AC50
Chi-Square = 40.00    DF = 34    P = 0.221

Assay                            Individual 95.0% CIs
plate name  N<=  N>  Median  Q3-Q1  --------+---------+---------+--------
PP00000065   2   0    0.36   0.04   (*
PP00000066   2   0    0.34   0.03   *)
PP20000001   2   0    0.64   0.12      *)
PP20000002   2   0    0.28   0.11   (*
PP20000003   2   0    0.44   0.14    *)
PP20000004   1   1    1.08   1.18    (-----*-----)
PP20000009   2   0    0.30   0.13   (*)
PP20000010   2   0    0.29   0.13   (*)
PP20000011   2   0    0.36   0.27   (-*)
PP20000012   2   0    0.49   0.00     *
PP20000013   1   1    1.20   1.16     (-----*-----)
PP20000014   1   1    1.28   1.29     (------*-----)
PP20000015   1   1    1.33   0.56        (-*--)
PP20000016   1   1    1.20   0.79     (---*---)
PP20000018   1   1    1.53   0.03          *
PP20000020   0   2    2.55   0.08                   (*
PP20000021   0   2    1.82   0.38            (-*-)
PP20000022   1   1    1.78   0.98          (----*----)
PP20000023   0   2    1.98   0.59           (--*--)
PP20000024   0   2    2.24   0.56             (-*--)
PP20000025   0   2    2.32   0.29              (*-)
PP20000027   1   1    1.71   1.13        (-----*-----)
PP20000028   1   1    1.94   1.51         (------*-------)
PP20000030   0   2    2.87   0.89                 (----*---)
PP20000031   0   2    2.52   0.96              (----*----)
PP20000032   1   1    1.96   1.37         (------*-----)
PP20000033   2   0    1.35   0.20        (*-)
PP20000034   1   1    1.49   2.02   (---------*---------)
PP20000035   1   1    1.04   1.01   (----*----)
PP20000042   0   4    2.09   0.70             (--*------)
PP20000043   2   2    1.55   0.63        (---*--)
PP20000044   2   2    1.42   1.16   (----*---------)
PP20000045   1   1    1.92   1.02            (----*----)
PP20000046   0   2    1.83   0.31             (*-)
PP20000047   1   1    1.70   0.97         (----*----)
                                   --------+---------+---------+--------
                                      1.0       2.0       3.0

Overall median = 1.53
```