ADVANCED DATA ANALYSIS AS AN ENABLER TO NEAR REAL-TIME CONTAMINATION CONTROL STRATEGY EVALUATION

Walid **El Azab**

QP Pro Services & QPM Consulting
Co-founder & Senior Consultant

QP PRO SERVICES

TRUSTED FOR **CONTINUITY.**

QP PRO SERVICES

Advanced Data Analysis as an enabler to near real-time Contamination Control Strategy Evaluation

El Azab Walid & El Azab Shady

Introduction:

The Contamination Control Strategy (CCS) should be documented and implemented for all pharmaceutical manufacturers producing sterile products (1-3). Manufacturers producing nonsterile products, where the control and reduction of microbial, particulate and endotoxin/pyrogen contamination are considered important, may elect to implement the CCS (1,3). The purpose of the CCS is to define all critical control points (design, procedural, technical, and organizational), monitor the measures in place, and assess the effectiveness of those controls to manage the risks to medicinal product quality and safety *(1)*.

The CCS should establish a robust assurance of contamination prevention and control. The periodic evaluation of the CCS must be associated with continuous improvements to the Pharmaceutical Quality System (PQS), contamination controls/measures, and manufacturing processes – all aiming to minimize the risks to product quality and safety. The effectiveness of the controls and measures in place should be part of ongoing and periodic management reviews, as required by the European Union (EU) guideline Annex 1 for Good Manufacturing Practice for Medicinal Products for Human and Veterinary (1).

Several manufacturers are in the process of documenting the CCS, while others have already done so *(4)*. One of the key questions to be answered today is how to best leverage the huge amount of data generated across our manufacturing processes and environments to minimize the risks to product quality and safety and fulfill a holistic assurance of contamination prevention. In this article, you will learn the three important steps in creating a near real-time CCS dashboard using advanced data analysis tools like Microsoft Power BI to unveil contamination trends and identify potential root causes of product contamination. Once this CCS dashboard is set up, it requires limited maintenance. It runs on a continuous basis providing a near real-time holistic view of contamination control performance by identifying and monitoring all critical controls influencing contamination and assessing the effectiveness and performance of all the contamination prevention measures in place. The use of Artificial Intelligence and Machine Learning Algorithms with tools like Microsoft Power BI could also help manufacturers predict CCS performance before production launch.

Using the Data, Information, Knowledge, Wisdom (DKIW) pyramid to set up the CCS:

The pharmaceutical industry is highly regulated to assure product quality and safety. To comply with the regulation, to prove adherence to those high standards, and to demonstrate product quality and safety, a high volume of data is generated and captured throughout the different stages of the manufacturing process of a medicinal product.

This data captures multiple dimensions of the pharmaceutical quality system and manufacturing process, such as facility, premises, equipment, utilities, personnel, raw material, environment, third parties, and product processing. Each element contains sub-elements that are controls or measures such as cleaning and disinfection, decontamination, sterilization, environmental monitoring, water monitoring, steam quality monitoring, aseptic process simulation, personnel monitoring, supplier performance analysis, in-process control, and final release product testing. These sub-elements are monitored through specifications such as particles, microorganisms, and endotoxin/pyrogen that generate the 'data.' The purpose of the data generated is to shed light on the process performance to distinguish trends that may lead to failure and enable us to decide on the best course of action to correct or prevent the failure.

This data comes from various digital (Excel, LIMS, Scada, applications, etc.) or, in certain cases, physical (sheets of paper) sources. With proper contextualization and interpretation of the data collected, it is possible to understand the drivers of process performance and decide on the best course of action to improve that process (**Figure 1**). As Carly Fiorina once said, "The goal is to turn data into information, and information into insight." The information is provided by connecting the data and the insight by interpreting the information. For the information and thus insight to be correct, the underlying data needs to be accurate, reliable, trustful, and complete *(5)* to guarantee the right conclusions and understanding of the contamination root causes and allow for reliable Corrective and Preventive Actions (CAPA) implementation that addresses the root causes of contamination.

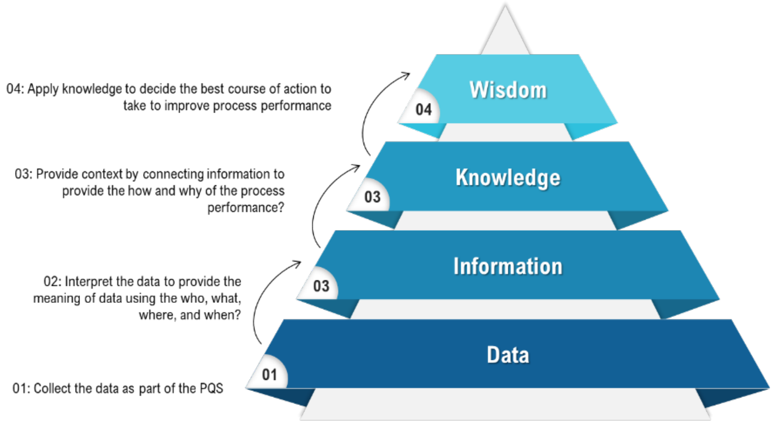

04: Apply knowledge to decide the best course of action to take to improve process performance
04
Wisdom
03: Provide context by connecting information to provide the how and why of the process performance?
03
Knowledge
02: Interpret the data to provide the meaning of data using the who, what, where, and when?
03
Information
01
Data
01: Collect the data as part of the PQS

Figure 1: Data (D), Information (I), Knowledge (K), Wisdom (W), or DKIW pyramid, the higher in the pyramid, the higher the human element is engaged. (6)

The steps to follow to build the first version of a near real-time CCS dashboard using actual data:

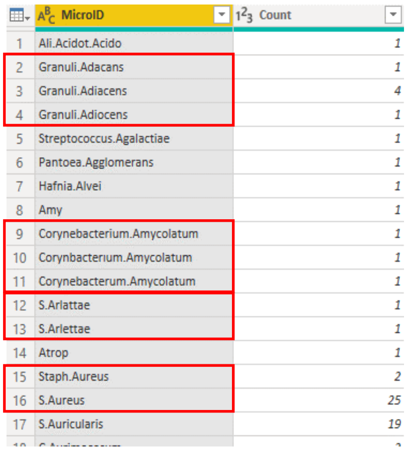
Building a near real-time CCS dashboard from scratch will expose you to many challenges, including how to go from the huge amount of data to clear insights that drive effective actions to prevent future contamination.

How many times have you stared at an Excel file with large amount of data and wondered how to structure the data to see a trend, a distribution, understand what is the root cause of a problem or get a better understanding of the process at hand? You tried the normal distribution, it didn't work. You plotted the data over time, it looked like a radio signal from Mars. You plotted a correlation graph; it just wasn't there. After hours of trying and staring at your screen, you hopped on to another task and hoped for tomorrow to be a luckier day.

In this section, you will learn the three important steps in creating near real-time CCS dashboards using advanced data analysis tools like Microsoft Power BI to unveil contamination trends and identify the potential root causes of drug contamination in your manufacturing process. Microsoft Power BI copies the data from one or multiple sources to achieve the objective; therefore, we should anticipate a minimal Computer System Validation compared to an excel sheet validation (7). Microsoft Power BI is already used by prominent (bio)pharmaceutical companies.

The first step in any analysis or dashboard building is **cleaning up the data**. The cleaning up of the data in GMP activities should comply with data integrity guidelines (5). This step is cumbersome, annoying, and often overseen by eager analysts and managers trying to get to the answer too quickly. The biggest challenge facing the transition to the era of big data in pharmaceutical companies' – and other large companies across many industries – is not the amount of data being captured but rather the quality of the data generated. This data is encoded in different geographies and manufacturing plants with often different definitions for the same KPI. Building a worldwide dashboard in those conditions, makes data cleaning an essential step in ensuring a fair comparison across manufacturing plants and geographies – comparing apples to apples. Data cleaning remains crucial even when looking at a single drug manufacturing site. Why? Current data entry tools are still imperfect in that they may be used inconsistently from operator to operator. Even though big companies talk about using Artificial Intelligence and Machine Learning, most are still capturing data with a traditional Excel sheet completed by an operator. That operator writes ‘Jack’, ‘Jak’, and ‘Jac’ or ‘*Staphylococcus cohnii*’, ‘*Staphylococcus cohni’*, and ‘*Staphylococcus cohniii*’. Imagine how the next operator taking over the Excel sheet will write ‘Jack’ or ‘*Staphylococcus cohnii’* if the same operator already found three different ways of writing the same name. Silly, yes, but if not corrected, it could mislead you into drawing the wrong conclusion or no conclusion.

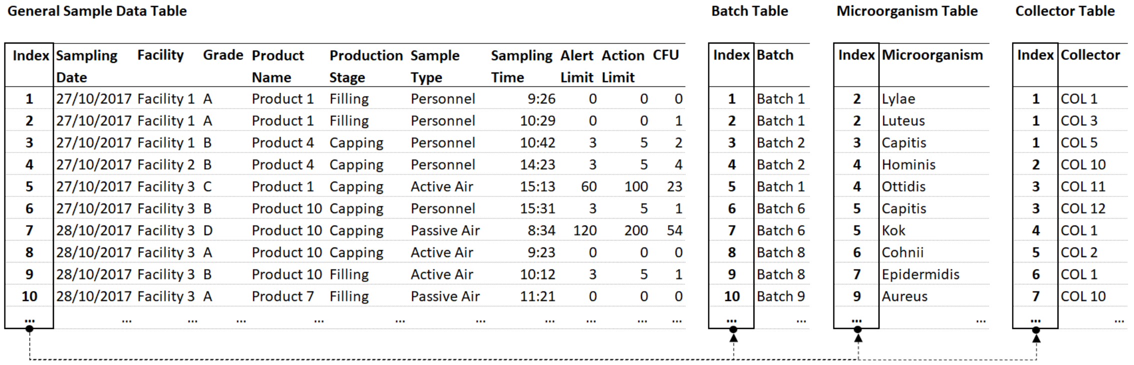
Be smart when cleaning up the data, or you could waste your time with limited return. Only clean up the data for the dimensions you believe are potential root causes or drivers of your problem or key metric you are analyzing. If you are looking at drug contamination, like in our case, and don’t believe the product to have any influence on drug contamination, then don’t waste your time cleaning up the ‘Product Name’ column. Don’t also bother cleaning up the ‘description of the sample location’ column as this is a text field with a long description of where the sampling was done – the chances of you finding a few samples with the same entry for this column is minimal. After identifying which columns to clean up, an easy way to clean up the data is to use the Power Query editor of Microsoft Power BI to isolate this one dimension you are cleaning up, count the number of matching entries or samples here for that one dimension, and rank them by ascending alphabetic order. In the case of the ‘Microorganism’ column, it will give you the results in **Figure 2**. By going through the table, you can easily identify the misspelled microorganisms and adjust them in the initial table with a *ReplaceText* function. You would also easily be able to identify the right ‘spelling’ – if you didn’t know it already – as this would be the one with the highest representation (cf. ‘Count’ column in **Figure 2**).



	MicroID	Count
1	Ali.Acidot.Acido	1
2	Granuli.Adacans	1
3	Granuli.Adiacens	4
4	Granuli.Adiocens	1
5	Streptococcus.Agalactiae	1
6	Pantoea.Agglomerans	1
7	Hafnia.Alvei	1
8	Amy	1
9	Corynebacterium.Amycolatum	1
10	Corynbacterium.Amycolatum	1
11	Corynebacterum.Amycolatum	1
12	S.Arlattae	1
13	S.Arlettae	1
14	Atrop	1
15	Staph.Aureus	2
16	S.Aureus	25
17	S.Auricularis	19

Figure 2: Cleaning up the data is a crucial step to unearth the ‘correct’ information and insights. In this example, the microorganism column (MicroID) is cleaned up for spelling mistakes.

The second step in any analysis or dashboard building is **structuring your data**. Three sample collectors were involved in collecting that sample in which several microorganisms were detected. The way the data is encoded on the Excel sheet is one line represents one sample. Nothing dictates that you to keep this structure for your analysis or dashboard. It is easier, yes, but it will prevent you from reaching the right conclusions. Why? Let’s take an example: ‘Jack’ has made 350 samples with other collectors and 50 samples on his own. If you were to assess ‘Jack’s influence on samples contamination, with the encoded data structure, you would only have 50 datapoints, while if you were to restructure the data like in **Figure 3**, where the sample collector dimension was removed from the general sample data table and put in a separate “Collector” table, you would have 400 datapoints to assess ‘Jack’s influence on samples contamination. This should yield a more representative sample and allow for a fairer assessment of Jack’s contribution.



General Sample Data Table

Index	Sampling Date	Facility	Grade	Product Name	Production Stage	Sample Type	Sampling Time	Alert Limit	Action Limit	CFU
1	27/10/2017	Facility 1	A	Product 1	Filling	Personnel	9:26	0	0	0
2	27/10/2017	Facility 1	A	Product 1	Filling	Personnel	10:29	0	0	1
3	27/10/2017	Facility 1	B	Product 4	Capping	Personnel	10:42	3	5	2
4	27/10/2017	Facility 2	B	Product 4	Capping	Personnel	14:23	3	5	4
5	27/10/2017	Facility 3	C	Product 1	Capping	Active Air	15:13	60	100	23
6	27/10/2017	Facility 3	B	Product 10	Capping	Personnel	15:31	3	5	1
7	28/10/2017	Facility 3	D	Product 10	Capping	Passive Air	8:34	120	200	54
8	28/10/2017	Facility 3	A	Product 10	Capping	Active Air	9:23	0	0	0
9	28/10/2017	Facility 3	B	Product 10	Filling	Active Air	10:12	3	5	1
10	28/10/2017	Facility 3	A	Product 7	Filling	Passive Air	11:21	0	0	0
...	...	...	...	...	...	...	...	...	...	...

Batch Table

Index	Batch
1	Batch 1
2	Batch 1
3	Batch 2
4	Batch 2
5	Batch 1
6	Batch 6
7	Batch 6
8	Batch 8
9	Batch 8
10	Batch 9
...	...

Microorganism Table

Index	Microorganism
2	Lylae
2	Luteus
3	Capitis
4	Hominis
4	Ottidis
5	Capitis
5	Kok
6	Cohnii
7	Epidermidis
9	Aureus
...	...

Collector Table

Index	Collector
1	COL 1
1	COL 3
1	COL 5
2	COL 10
3	COL 11
3	COL 12
4	COL 1
5	COL 2
6	COL 1
7	COL 10
...	...

Figure 3: Selecting the best data structure is important to identify the drivers of contamination and evaluate their influence

Once the data is cleaned up and properly structured, the last step in any analysis or dashboard building is **displaying the data in an insightful manner** to help the user derive the right conclusion. The granularity of the data shown will depend on the audience targeted – an executive may not like to deep dive in the details and will prefer a high-level overview showing the trends on the core KPIs and some benchmarking highlighting potential improvements. A manager will want a high-level view on core KPIs with the ability to deep dive and understand the root causes for low performance to identify levers for improvements. An operator will likely want self-performance KPIs and recommendations on how to improve. The best way to build your dashboard visuals is to draw them on paper and discuss them with – a sample of – your audience to deliver the most suited views for your audience.

Selecting the 'right' graph to help your audience easily capture the insights is an art. Take, for example, the 'Mekko Chart' in **Figure 4**. It allows the user to easily see the importance of each 'Period', 'Facility', 'Grade', *etc.* by just looking at the width on the X-Axis, which represents the relative number of Environmental Monitoring (EM) samples. This information is crucial in reading the graph and driving the right conclusion. Imagine you had only two samples in 2019 with one fail (CFU above Action Trigger). Showing your results in a conventional 100% Stacked column chart would lead you to conclude that 2019 was the worst year with 50% failed samples when there were only two samples. Looking at **Figure 4**, one can conclude that 2018 was a great year with the highest number of samples and yet the lowest percentage of failed samples.

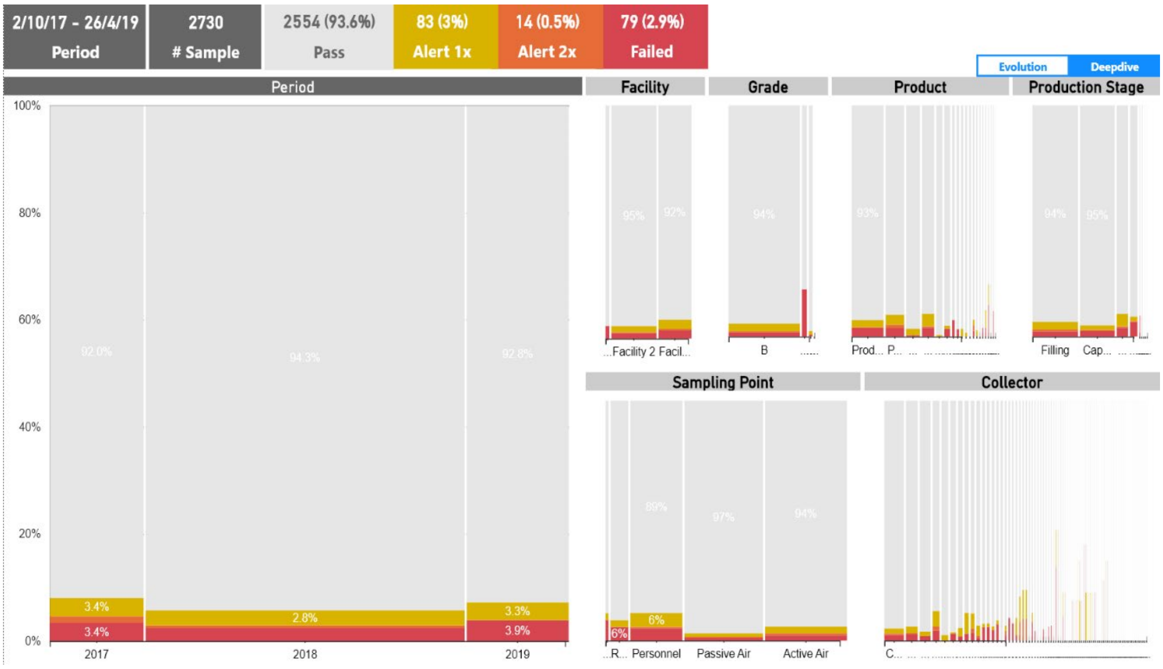
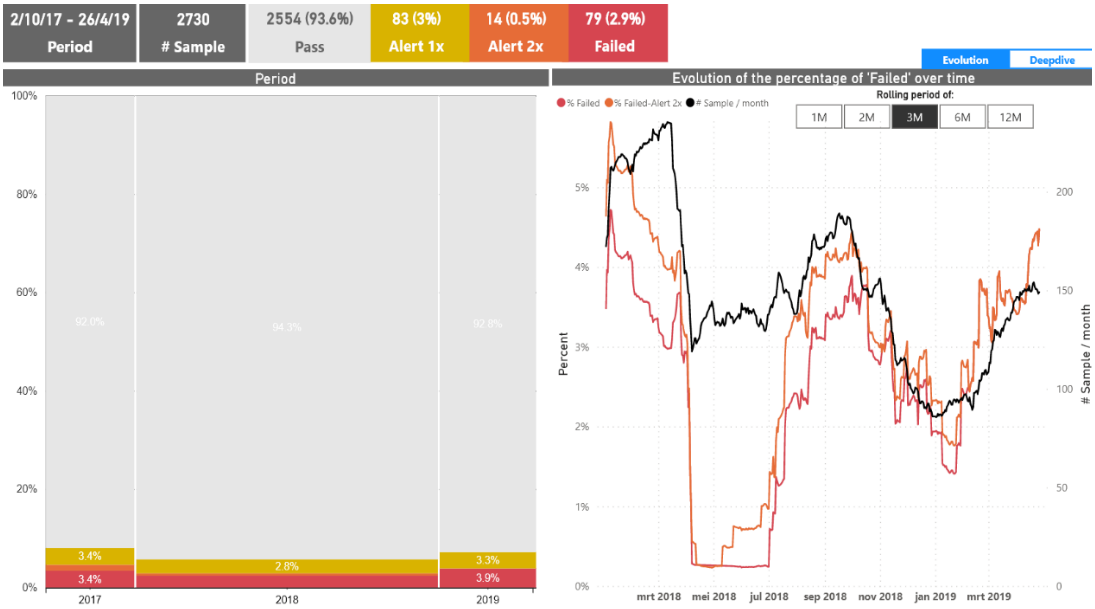


Figure 4: A view of the CCS dashboard summarizing the number of samples taken across three facilities and the percentage of Failed samples and samples with Alert levels *(top part of the graph)*. The charts used in this view are 'Mekko' charts where the size of the X-Axis for each value (2017, 2018, 2019) represents the number of samples for that value while the percentage of failed and alert samples is represented on the Y-Axis. 'Mekko' charts allow for a more accurate read than conventional 100% Stacked column charts.

It is difficult to see trends when looking at timeseries graph. One needs to smoothen the curve to see the trend. How? By looking at the results using a rolling period of 'X' months. You only have a few samples per day; calculating the percentage of failed samples per day and plotting that on a line chart will give you an 'Alien' like signal. Rather, if, for every day, you look back at three months of sample data and calculate the percentage of failed samples, this would give you a smooth curve showing you where you are headed: to higher or lower fails? **Figure 5** shows the evolution of the percentage of failed samples over time with a 3-month rolling period. One can see that there was a decline in the percentage of failed in the first half of 2018, followed by an increase in Q3 2018, a decline in Q4 of 2019, and finally, an increase at the start of 2019. Clearly, the site was not heading in the right direction at the time and needed to identify what was driving this increase in failed samples.

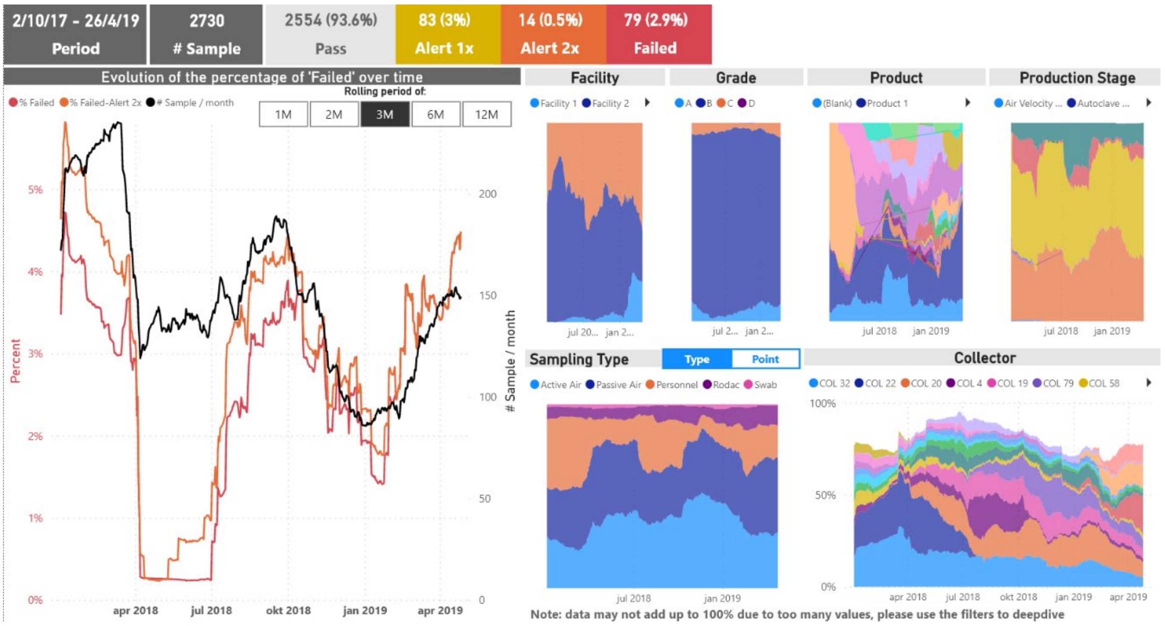
A few things to keep in mind when using a line chart with a 'X-month rolling period: (1) The longer the rolling period, the more data you will need to see a trend. If you select a 12-month rolling period, the first 12 months of your data won't be visible on your line chart as they will be averaged on the first point of your graph. (2) The longer the rolling period, the harder to identify potential one-off events as these will be smoothened, and the harder to assess the impact of your latest improvement measures as these will have a lower weight in a longer period. (3) If you believe seasonality may affect your performance and want to neutralize for it, find and use the shortest period length that neutralizes seasonality. You might not want to *de facto* go for the 12-month rolling period for the drawbacks mentioned earlier.



2/10/17 - 26/4/19	2730	2554 (93.6%)	83 (3%)	14 (0.5%)	79 (2.9%)
Period	# Sample	Pass	Alert 1x	Alert 2x	Failed

Figure 5: On the right side of this view, a line-graph with the evolution of the percentage of failed samples over a 3-month rolling period. Each data point on this graph is the average of the last 3 months' data. This approach smoothens the curve and allows for a better read of the improvement/deterioration trends in sample contamination.

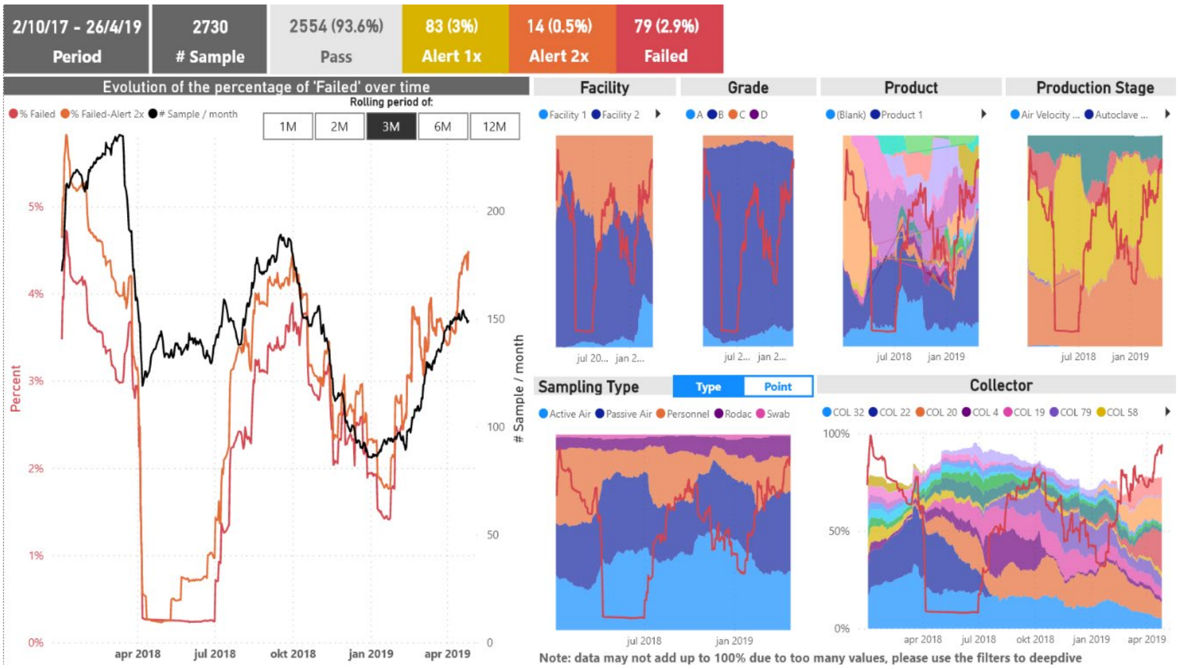
One should not hesitate to build 'bold' views if one believes these will drive more insights to your audience. Such a view could be **Figure 6a** which shows, on the left side, the evolution of the percentage of failed samples on a 3-month rolling period and, on the right side, the evolution on a 3-month rolling period of each identified potential driver of microorganism contamination. This view may seem 'noisy' at first – *gosh, so many colors* - but looking with a more probing eye, one can find potential correlations between a higher percentage of failed samples and 'Personnel' sampling, use of 'Facility 1' or sampling by 'Collector 22'. The share of 'Personnel' sampling increased in early July 2018 at the same time as the percentage of failed samples, and both decreased in early October. The share of samples done in 'Facility 1' increased in January 2019 at the same time as the percentage of failed samples. The share of samples done by 'Collector 22' dropped in March 2018 at the same time as the percentage of failed samples. Those correlations are visible to the trained eye but help yourself and superpose the curve representing the percentage of failed samples (cf. red curves superposing each driver graph on **Figure 6b**) onto each driver's graph to detect those correlations easily. Showing-hiding this red line is done with a single click on Microsoft Power BI.



2/10/17 - 26/4/19	2730	2554 (93.6%)	83 (3%)	14 (0.5%)	79 (2.9%)
Period	# Sample	Pass	Alert 1x	Alert 2x	Failed

Evolution of the percentage of 'Failed' over time
● % Failed ● % Failed-Alert 2x ● # Sample / month
Rolling period of:
1M
2M
3M
6M
12M
Percent
5%
4%
3%
2%
1%
0%
200
150
100
50
0
# Sample / month
apr 2018
jul 2018
okt 2018
jan 2019
apr 2019
Facility
● Facility 1 ● Facility 2
jul 20... jan 2...
Grade
● A ● B ● C ● D
jul 2... jan 2...
Product
● (Blank) ● Product 1
jul 2018 jan 2019
Production Stage
● Air Velocity ... ● Autoclave ...
jul 2018 jan 2019
Sampling Type
Type
Point
● Active Air ● Passive Air ● Personnel ● Rodac ● Swab
jul 2018 jan 2019
Collector
● COL 32 ● COL 22 ● COL 20 ● COL 4 ● COL 19 ● COL 79 ● COL 58
100%
50%
0%
apr 2018 jul 2018 okt 2018 jan 2019 apr 2019
Note: data may not add up to 100% due to too many values, please use the filters to deepdive

Figure 6a: View of the CCS Dashboard, which may seem intimidating at first but will help uncover potential correlations and drivers of higher failed samples. The left part shows the percentage of failed samples over a 3-month rolling period, while the right part shows the value over a 3-month rolling period of the potential drivers of contamination. Comparing the two parts in this view may help uncover correlations and drivers of contamination to be further investigated.



2/10/17 - 26/4/19	2730	2554 (93.6%)	83 (3%)	14 (0.5%)	79 (2.9%)
Period	# Sample	Pass	Alert 1x	Alert 2x	Failed

Evolution of the percentage of 'Failed' over time
● % Failed ● % Failed-Alert 2x ● # Sample / month
Rolling period of:
1M
2M
3M
6M
12M
Percent
5%
4%
3%
2%
1%
0%
200
150
100
50
0
# Sample / month
apr 2018
jul 2018
okt 2018
jan 2019
apr 2019
Facility
● Facility 1 ● Facility 2
jul 20... jan 2...
Grade
● A ● B ● C ● D
jul 2... jan 2...
Product
● (Blank) ● Product 1
jul 2018 jan 2019
Production Stage
● Air Velocity ... ● Autoclave ...
jul 2018 jan 2019
Sampling Type
Type
Point
● Active Air ● Passive Air ● Personnel ● Rodac ● Swab
jul 2018 jan 2019
Collector
● COL 32 ● COL 22 ● COL 20 ● COL 4 ● COL 19 ● COL 79 ● COL 58
100%
50%
0%
apr 2018 jul 2018 okt 2018 jan 2019 apr 2019
Note: data may not add up to 100% due to too many values, please use the filters to deepdive

Figure 6b: same as **Figure 6a** with the left graph being superposed on right graphs to more easily uncover potential correlation and negative influence on EM samples contamination

Powerful tools like Microsoft Power BI will allow you to easily filter graphs and deep dive into a specific dimension. **Figure 7** displays on the left side a ranking of detected microorganisms with the number of samples in which they were detected. '*Staphylococcus hominis*' is the most prominent microorganism followed by '*Staphylococcus epidermidis'*. The graph on the right side shows the number of samples in which those microorganisms were detected over time. By just clicking on the microorganism on the left graph, Microsoft Power BI filters the graph on the right only to show the contamination by this selected microorganism over time. A small but important parenthesis: had you kept the data structure the same as the one encoded in the Excel sheet source, i.e., one single table with all the data, you would not be able to show the samples where multiple microorganisms were detected at the same time. Decoupling the detected microorganisms from the samples, like in the data structure shown in **Figure 3,** allows you here to show all the samples where the microorganism selected was detected, even those where several other microorganisms were detected along. Hence, the importance of properly structuring your data.

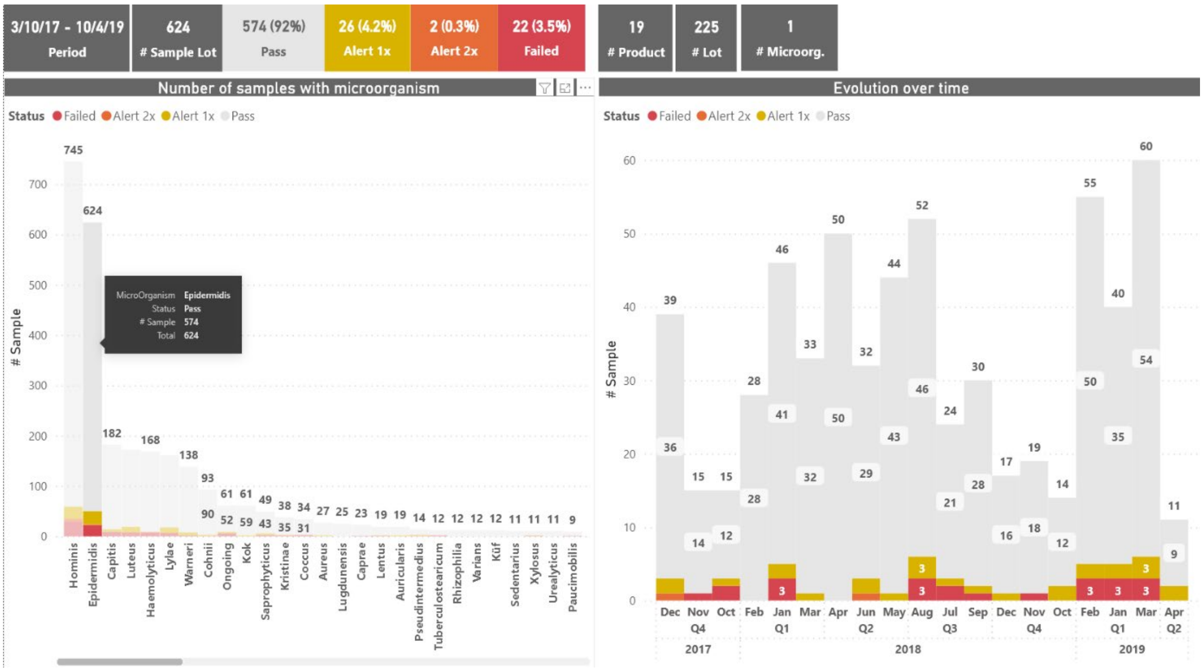


Figure 7: Deep dive on 'Staphylococcus epidermidis' detection in samples over time with just a click. The left graph displays the microorganisms detected and the number of samples affected, while the right graph displays the detection of these microorganisms over time in EM samples.

Enough looking at samples, you may say; what about manufacturing batches? How many batches of the same product are needed to deliver at least one batch with no failed EM sample? **Figure 8a** displays over time the number of batches needed to deliver at least one batch with no failed sample with a certainty of 80%, 95%, and 98%. How is it calculated? Consider the percentage of failed batches over a 3-month rolling period – *same principle as failed samples as explained earlier* – and identify this value to the probability of a failed batch *'q.'* Now, consider the probability of having at least one batch with no failed samples using the binomial distribution formula:

$$Probability\ (\,Batch\ with\ no\ failed\ sample \geq 1) \geq 95\%$$

This is equivalent to:

$$1 - \ Probability\ (\,Batch\ with\ no\ failed\ sample = 0) \geq 95\%$$

This can be rewritten as:

$$1 - C_n^0\, p^0(1-p)^n \geq 95\%$$

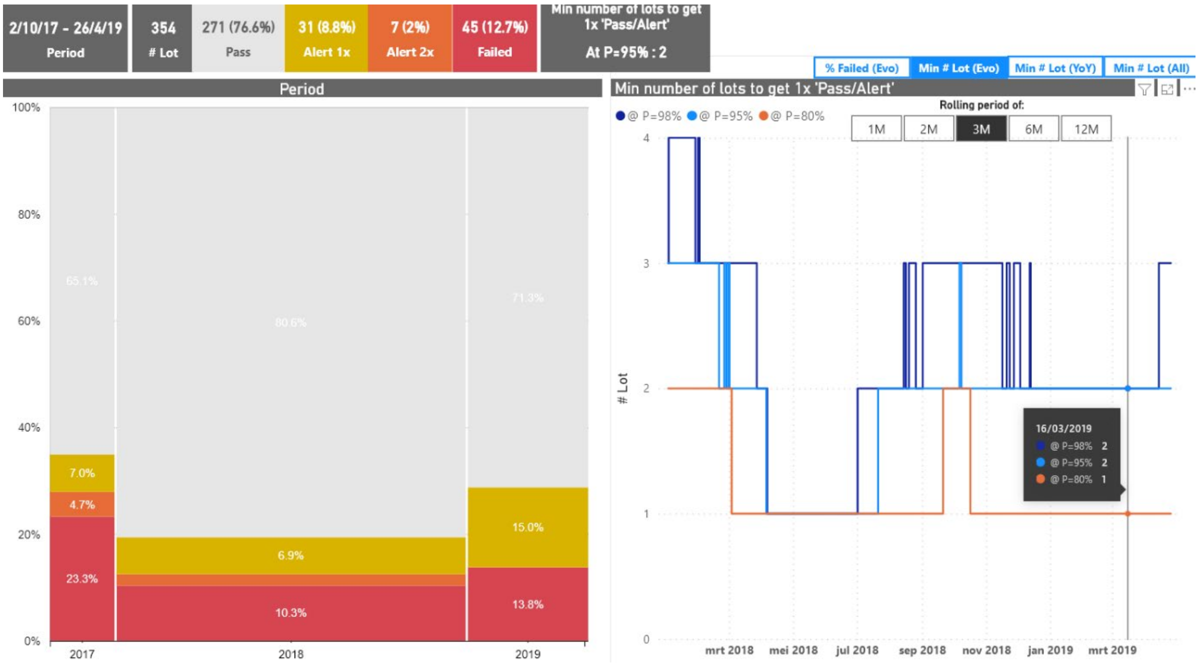
where $C_n^0 = \frac{n!}{(n-0)!\,0!} = 1$ is the binomial coefficient, n the number of batches and p the probability of a successful batch.

The equation can be rewritten using $q = 1 - p$ the probability of a failed batch:

$$1 - C_n^0\,(1-q)^0(1-q)^n \geq 95\%$$

Solving for n, leads us to the curves in **Figure 8a**.

Figure 8b shows the evolution of the needed batches to have at least one batch with no failed samples depending on the certainty needed (80%, 95% or 98%). The higher the importance of delivering the batch on time with no fails, the higher you will set this certainty / probability. 2018 was the best year as it required less batches at the same certainty to have at least one batch without failed samples. This matches with the left side graph where the percentage of failed batches is lowest in 2018 with yet the highest number of batches across the three years.



2/10/17 - 26/4/19	354	271 (76.6%)	31 (8.8%)	7 (2%)	45 (12.7%)	Min number of lots to get 1x 'Pass/Alert'
Period	# Lot	Pass	Alert 1x	Alert 2x	Failed	At P=95% : 2

Figure 8a: On the left side, a 'Mekko' chart showing the evolution of failed batches (Lots) over time. On the right side, a stepped line graph based on 3-month rolling data showing the number of batches needed over time to deliver at least one batch with no failed sample at three selected certainties needed: 80%, 95%, 98%.

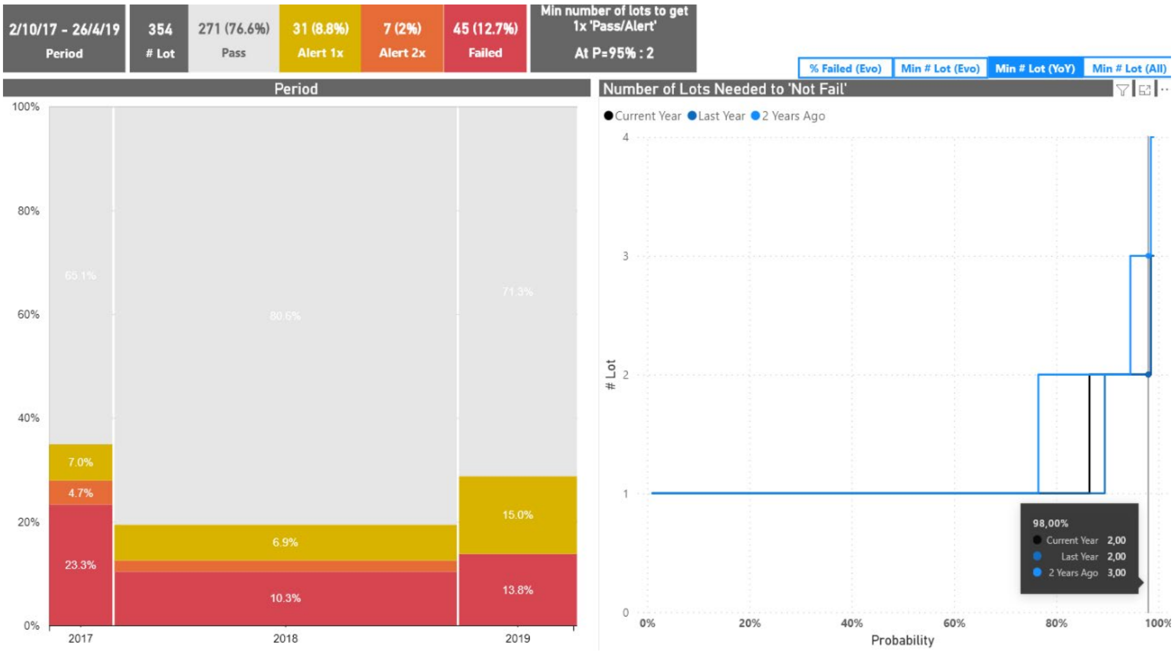
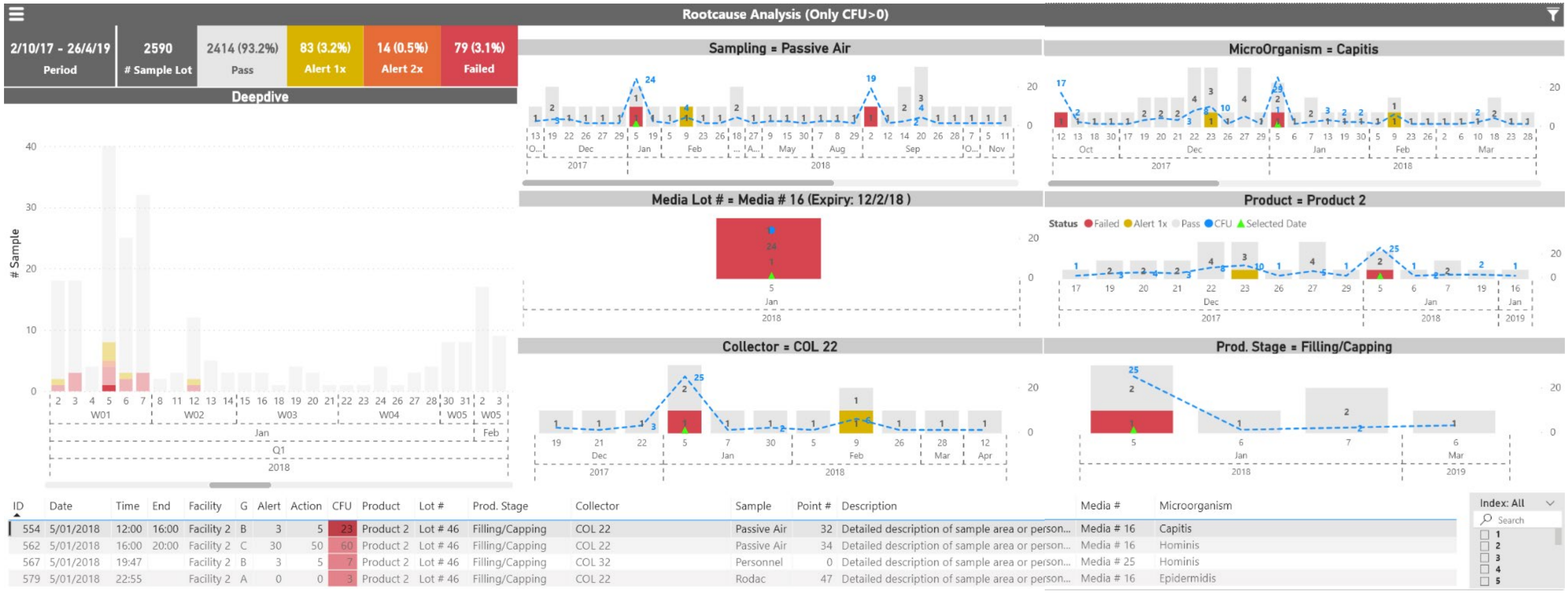


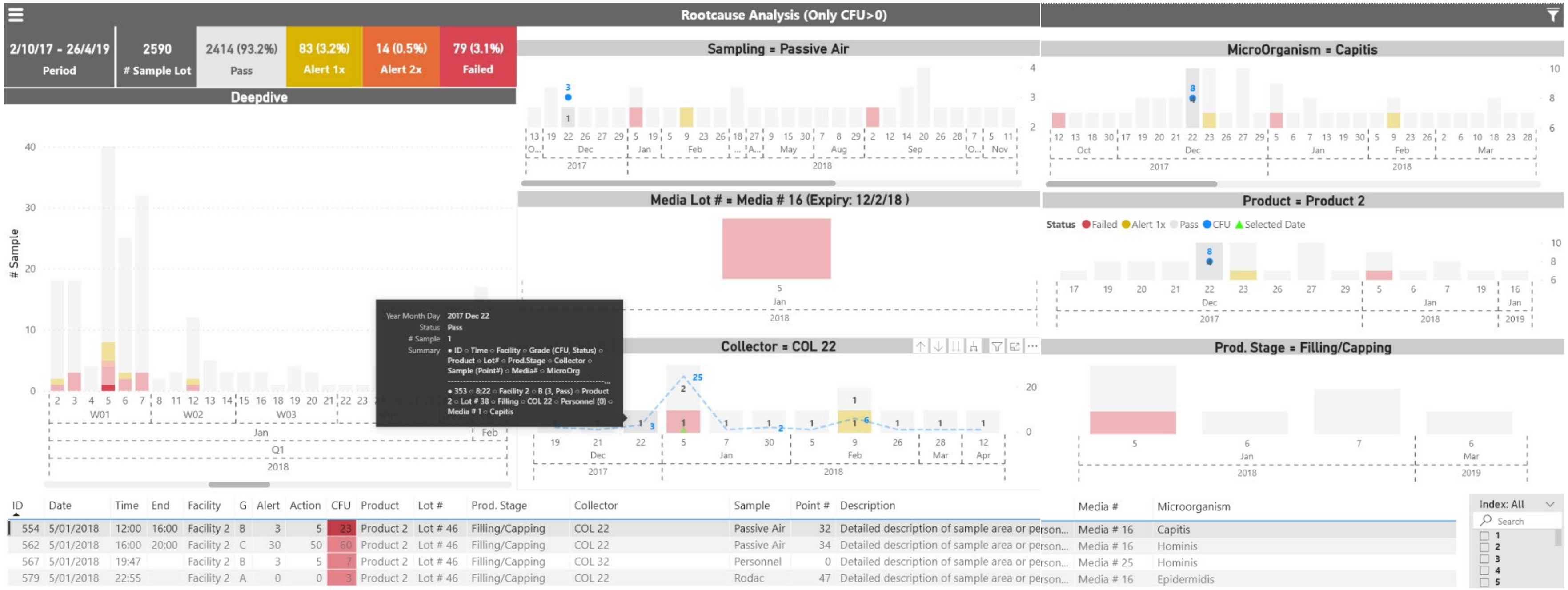
Figure 8b: On the left side, a 'Mekko' chart showing the evolution of failed batches (Lots) over time. On the right side, a stepped line graph showing the number of batches needed to deliver at least one batch with no failed sample depending on the certainty needed (the higher the certainty needed, the higher the number batches required)

Powerful data analysis solutions like Microsoft Power BI will allow you to create tools like the one displayed in **Figure 9a** that allow the user to detect potential causes of sample contamination by looking at several dimensions at the same time. How? The left graph shows the samples with a CFU>0 over time and highlights the failed samples and those with an alert. By clicking on failed samples for a specific date – the 5th of January 2018 –, the table below populates with the failed samples on that date. The user can then click on the specific sample to deep dive into – sample ID 554. The six graphs on the right light up. They display all the samples with the same microorganism as the one(s) detected in the selected sample keeping the dimension mentioned in the graph title the same as the one of the selected samples. For example, the 'Collector' graph displays all the samples done by the collector who collected the selected sample ('COL 22') and for which the same microorganism was detected ('*Staphylococcus capitis*'). The user can click on the sample 'COL 22' done on the 22nd of December 2017 to see if the collector could be the driving channel of the contamination (cf. **Figure 9b**). In that case, the user easily sees that there is another batch of the same product for which the sampling type ('Passive Air'), the product ('Product 2'), the collector ('COL 22') and the microorganism ('*Staphylococcus capitis*') matches with the 5th of January 2018 failed sample. Could there be a link between the collector and the contamination on the 5th of January 2018 that led to the failed sample and batch?



ID	Date	Time	End	Facility	G	Alert	Action	CFU	Product	Lot #	Prod. Stage	Collector	Sample	Point #	Description	Media #	Microorganism
554	5/01/2018	12:00	16:00	Facility 2	B	3	5	23	Product 2	Lot # 46	Filling/Capping	COL 22	Passive Air	32	Detailed description of sample area or person...	Media # 16	Capitis
562	5/01/2018	16:00	20:00	Facility 2	C	30	50	60	Product 2	Lot # 46	Filling/Capping	COL 22	Passive Air	34	Detailed description of sample area or person...	Media # 16	Hominis
567	5/01/2018	19:47		Facility 2	B	3	5	7	Product 2	Lot # 46	Filling/Capping	COL 32	Personnel	0	Detailed description of sample area or person...	Media # 25	Hominis
579	5/01/2018	22:55		Facility 2	A	0	0	3	Product 2	Lot # 46	Filling/Capping	COL 22	Rodac	47	Detailed description of sample area or person...	Media # 16	Epidermidis

Figure 9a: Tool from the CCS dashboard to deep dive on contaminated sampled and identify what could be the driver of contamination (e.g., Sample collector, Sampling Area, Production Stage)


Rootcause Analysis (Only CFU>0)
Deepdive
# Sample
Sampling = Passive Air
MicroOrganism = Capitis
Media Lot # = Media # 16 (Expiry: 12/2/18)
Product = Product 2
Status ● Failed ● Alert 1x ● Pass ● CFU ▲ Selected Date
Collector = COL 22
Prod. Stage = Filling/Capping
W01
W02
W03
Jan
Feb
Q1
2017
2018
2019
Index: All
Search
□ 1
□ 2
□ 3
□ 4
□ 5

2/10/17 - 26/4/19	2590	2414 (93.2%)	83 (3.2%)	14 (0.5%)	79 (3.1%)
Period	# Sample Lot	Pass	Alert 1x	Alert 2x	Failed

Year Month Day 2017 Dec 22
Status Pass
Sample 1
Summary ● ID ○ Time ○ Facility ○ Grade (CFU, Status) ○ Product ○ Lot# ○ Prod.Stage ○ Collector ○ Sample (Point#) ○ Media# ○ MicroOrg
--...
● 353 ○ 8:22 ○ Facility 2 ○ B (3, Pass) ○ Product 2 ○ Lot # 38 ○ Filling ○ COL 22 ○ Personnel (0) ○ Media # 1 ○ Capitis

ID	Date	Time	End	Facility	G	Alert	Action	CFU	Product	Lot #	Prod. Stage	Collector	Sample	Point #	Description	Media #	Microorganism
554	5/01/2018	12:00	16:00	Facility 2	B	3	5	23	Product 2	Lot # 46	Filling/Capping	COL 22	Passive Air	32	Detailed description of sample area or person...	Media # 16	Capitis
562	5/01/2018	16:00	20:00	Facility 2	C	30	50	60	Product 2	Lot # 46	Filling/Capping	COL 22	Passive Air	34	Detailed description of sample area or person...	Media # 16	Hominis
567	5/01/2018	19:47		Facility 2	B	3	5	7	Product 2	Lot # 46	Filling/Capping	COL 32	Personnel	0	Detailed description of sample area or person...	Media # 25	Hominis
579	5/01/2018	22:55		Facility 2	A	0	0	3	Product 2	Lot # 46	Filling/Capping	COL 22	Rodac	47	Detailed description of sample area or person...	Media # 16	Epidermidis

Figure 9b: Tool from the CCS dashboard to deep dive on contaminated sampled and identify what could be the driver of contamination (e.g., Sample collector, Sampling Area, Production Stage)

The Future:

Imagine integrating AI, Machine Learning Algorithms, and real-time data reading systems to this CCS dashboard to automate identifying the importance of each control on contamination using historical data, read the current controls value, and predict the probability of failure, using historical data of a batch before even launching the production batch; you could adjust the measures and controls to maximize the probability of a successful batch at no cost and launch the production after you have maximized the chances of success. The savings that could be made! This future is close.

It all starts with improving the quality of the data captured by improving our data entry tools and aligning the definition across manufacturing sites and geographies. This will alleviate the investment in data cleaning and allow for faster ingestion of the data into AI and Machine Learning models.

Conclusion:

The use of modern tools such as Microsoft Power BI or a similar application enables data gathering from various sources and analyzing it to help identify potential contamination root causes while providing a near real-time holistic view on CCS trends and performance.

Building the CCS dashboard starts with cleaning up the data. Then, finding the ideal way of structuring the data is crucial and may need some deep thinking – don't hasten this step. The visualization, the final step, is a collaborative development with you and your audience. Select the visuals carefully to not overwhelm your audience, yet to surface the key insights and allow them to easily find the potential root causes of their problems and the measures that could address them. Practice will make you better and allow you to reach that fine balance.

Such a near real-time analysis provides similar or even more detailed information to the one contained in annual product review reports. In addition, it can also provide overall performance indicators of several CCS key performance indicators that will be useful to senior management or decision maker. Finally, the use of a near real-time CCS dashboard allows for 'continuous' periodic reviews of the CCS performance, resulting in quicker effectiveness checks of changes or improvements made in the pharmaceutical quality system to guarantee the ongoing assurance of contamination control.

Reference:

1. The Rules Governing Medicinal Products in the European Union Volume 4 EU Guidelines for Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use, Annex 1: Manufacture of Sterile Medicinal Products (Aug 2022)
2. El Azab W., Contamination Control Strategy: Implementation Roadmap, PDA Journal of Pharmaceutical Science and Technology March 2021,
3. ECA CCS guideline, How to Develop and Document a Contamination Control Strategy, accessed on 13 dec 2022: https://www.eca-foundation.org/news/ccs-task-force-issues-new-guideline.html
4. El Azab W, Hoenen I., Contamination Control Strategy: practices & a case study of a CCS implementation, La Vague, Jan 2022.
5. Source: PIC/S guidance Good Practices for Data Management and Integrity in Regulated GMP/GDP Environment (Jul 2021)
6. Jones B., Data Literacy Fundamentals, Understanding the Power & Value of Data, Data Literacy Press, 1st edition (2020)