User Manual for Patient-Based Quality Control Excel Spreadsheet Tool

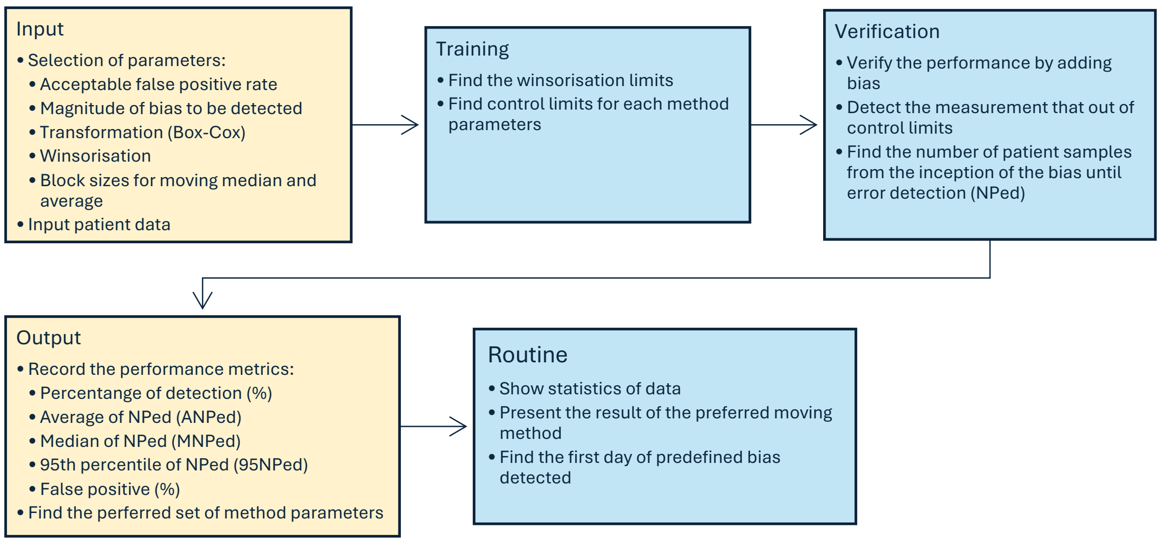
This is the user manual for the Patient-Based Quality Control (PBQC) Excel tool. This tool provides a complete process for setting up (training of algorithm), optimization, verification and routine running of PBQC algorithms. Data visualizations are modelled to illustrate data transformation (Box-Cox), data winsorization and routine data monitoring in control chart. Of note, winsorisation is the process of replacing a specified number of extreme values with a smaller data value.

There are five major steps in the use of the Excel tool, which are:

1. Inputting parameters of model,
2. Model training,
3. Model verification,
4. Outputting the preferred parameters of model and
5. Routine monitoring of PBQC.

The statistical algorithms available in the Excel tool are moving median and moving average.

The framework below is an overview of the PBQC Excel tool.



The compressed file contains:

1. PBQC Excel tool,
2. Sample data of sodium,
3. Sample data of aspartate transaminase (AST),
4. User manual.

measurement(AST)
measurement(Na)
PBQC_v2.2
User_manual_PBQC_v2.2

This sample data is provided for the convenience of user to explore and familiarise with the Excel tool. This data is in a format that can be 'copy' and 'paste' into the Excel tool. The headers are 'id', 'date' and 'measurement' as depicted below for the sodium example.

id	date	measurement
1	2020-01-01	138
2	2020-01-01	141
3	2020-01-01	145
4	2020-01-01	139
5	2020-01-01	137
6	2020-01-01	141
7	2020-01-01	140
8	2020-01-01	143
9	2020-01-01	140
10	2020-01-01	140

The Excel tool has five sheets, which are 'Input', 'Training', 'Verification', 'Output', and 'Routine'.

The figure below shows the 'Input' sheet. Only the 'Acceptable False Positive Rate (%)'. 'Desirable Bias for detection(%)', 'Box-Cox transformation', 'Winsorisation %' and 'Block size' require user input based on their preferences.

- There are two options for 'Box-Cox transformation', which are 'Yes' and 'No'. When selecting 'Yes', the 'Box-Cox lambda' will be automatically calculated and included in the other Excel sheets.
- Similarly, the user can provide the preferred percentage of data for winsorization (%). A high percentage value will convert a lower number of extreme values and vice versa. Winsorization is bilateral. For example, a winsorization of 95% means 2.5% of the highest and 2.5% of the lowest values will be converted to the winsorization limit.
- The users can define the block sizes (n1, n2, and n3) for calculating the moving median (Med) and moving average (Avg).

Predefined setting	
Acceptable False Positive Rate(%)	0
Desirable bias for detection (%)	5
Box-cox transformation	No
Box-cox lambda	#N/A
Winsorisation %	98
Winsorisation lower limit	#NUM!
Winsorisation upper Limit	#NUM!

Method parameters	
Method1	Med
Method2	Avg
Block size1, n1	5
Block size2, n2	10
Block size3, n3	20

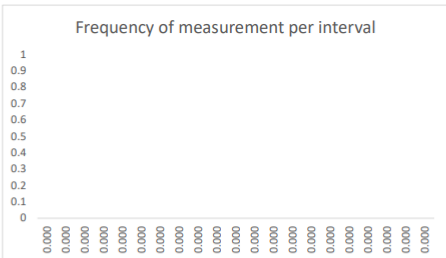
Column	Method	block size, n	lower control limit	upper control limit
MMed1	Med	5	#NUM!	#NUM!
MMed2	Med	10	#NUM!	#NUM!
MMed3	Med	20	#NUM!	#NUM!
MAvg1	Avg	5	#NUM!	#NUM!
MAvg2	Avg	10	#NUM!	#NUM!
MAvg3	Avg	20	#NUM!	#NUM!

The figure below shows the 'Training' sheet. The header 'ID', 'date' and 'measurement' requires user input. The sample data can be paste into these three columns to train the algorithm. The excel tool only works on numerical values for 'measurement'. Please remove any data containing symbols or alphabets in before entering into the Excel.

*For the purpose of initial familiarization using the sodium sample data, the user may copy and paste the first 300 days of data in the 'Training' sheet to train the algorithm, and the next 300 days of data can be copied and pasted in the 'Verification' sheet.

ID	date	measurement

Result	
min date=	0/1/1900
max date=	0/1/1900
min date (number)=	0
max date (number)=	0
number of days=	0



After pasting training and verification data, the first three sheets would show as below. The winsorisation and control limits are automatically determined from training data. Here, the winsorisation (%) was set as 99.5%, indicating that 0.5% of the most extreme results will be winsorized.

Predefined setting	
Acceptable False Positive Rate(%)	0
Desirable bias for detection (%)	5
Box-cox transformation	No
Box-cox lambda	#N/A
Winsorisation %	99.5
Winsorisation lower limit	132.000
Winsorisation upper Limit	147.000

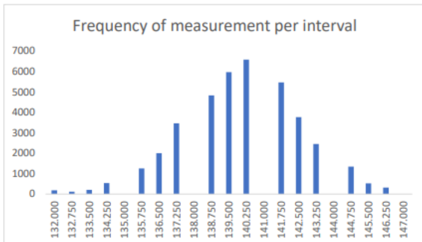
Method parameters	
Method1	Med
Method2	Avg
Block size1, n1	10
Block size2, n2	20
Block size3, n3	30

Column	Method	block size, n	lower control limit	upper control limit
MMed1	Med	10	136.500	144.000
MMed2	Med	20	138.000	143.000
MMed3	Med	30	138.000	142.500
MAvg1	Avg	10	137.200	143.400
MAvg2	Avg	20	138.300	142.550
MAvg3	Avg	30	138.800	142.233

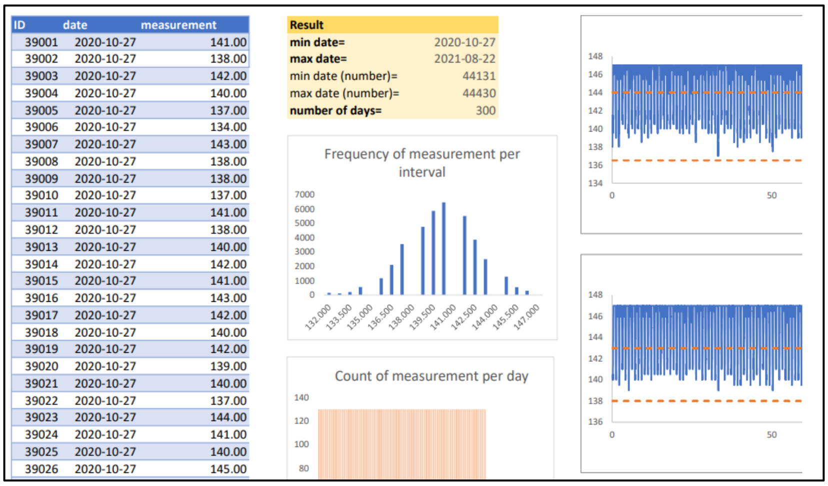
Basic information about the data input into the Excel are summarized along with a histogram of the data distribution. The histogram can be read to inform the laboratory user if Box-Cox transformation is required if a skewed distribution is present.

ID	date	measurement
1	2020-01-01	138.00
2	2020-01-01	141.00
3	2020-01-01	145.00
4	2020-01-01	139.00
5	2020-01-01	137.00
6	2020-01-01	141.00
7	2020-01-01	140.00
8	2020-01-01	143.00
9	2020-01-01	140.00
10	2020-01-01	140.00
11	2020-01-01	144.00
12	2020-01-01	141.00
13	2020-01-01	141.00
14	2020-01-01	140.00
15	2020-01-01	139.00
16	2020-01-01	141.00
17	2020-01-01	137.00
18	2020-01-01	135.00
19	2020-01-01	137.00
20	2020-01-01	136.00
21	2020-01-01	141.00
22	2020-01-01	138.00
23	2020-01-01	137.00

Result	
min date=	2020-01-01
max date=	2020-10-26
min date (number)=	43831
max date (number)=	44130
number of days=	300



In this sheet, a visual representation of the impact of the introduced bias on the daily data is shown on the right hand graphs with the predefined control limits (in red dashed lines).



For each algorithm (moving median, moving average), the results that are out of the control limits are marked as ‘positive’ (error detected). The ‘Perc detected’ (%) is the percentage of ‘positive’ (error) detected when the bias is introduced into the daily data of ‘verification’ sheet. On the other hand, the ‘NPed’ is the number of patient samples from the introduction of the (simulated) bias until error detection (NPed). This metric provides an indication of the ability of the algorithm to detect the predefined bias. A smaller NPed indicates a faster bias (error) detection and lower number of patient samples affected by the bias.

The fourth sheet, ‘output’ sheet records the performance metrics, including percentage of error detection, average of NPed (ANPed), median of NPed (MNPed), 95th percentile of NPed (95NPed) and false positive rate (%) for each algorithm. The performance metric that ranked the highest (best) among the algorithms are highlighted in green. The algorithm with the highest number of ‘best’ parameters are automatically selected (‘preferred’) and used for the subsequent ‘routine’ running of the PBQC algorithm.

Method	block size	Perc detected(%)	ANPed	MNPed	95NPed	False positive (%)	Preferred
Med	10	100	5	5	6	0	Preferred
Med	20	100	8	8	11	0	
Med	30	100	11	12	15	0	
Avg	10	100	5	5	7	0.015	
Avg	20	100	7	7	9	0.0125	
Avg	30	100	8	8	12	0.01	

Best metrics	
max per detected (%)	100
min ANPed	5
min MNPed	5
min 95NPed	6
min fasle positive (%)	0

Method parameters for real-time data	Value
Method	Med
Block size	10
lower control limit	136.500
upper control limit	144.000
Acceptable False Positive Rate(%)	0
Desirable bias for detection (%)	5
Box-cox transformation	No
Boc-cox lambda	#N/A
Winsorisation %	99.5
Winsorisation lower limit	132.000
Winsorisation upper Limit	147.000

Recommended preferred set of method Parameters is Moving Med with Block Size = 10 to detect 5% bias for 0% of Acceptable False Positve Rate

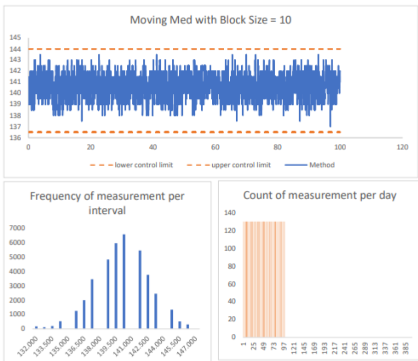
The last sheet is the 'routine' sheet, where laboratory users can input their routine laboratory data.

For the purpose of familiarization with the Excel tool, the user can input the last 100 days of the sodium sample data as 'routine data'. As noted above, the preferred set of algorithms parameters are automatically applied to the routine data.

The general descriptive statistics of data and the result of moving algorithms are presented. The number and percentage of positive detected is recorded. The first sequence number, day and date for the bias detected are determined.

ID	date	measurement
78001	2021-08-23	139.00
78002	2021-08-23	144.00
78003	2021-08-23	144.00
78004	2021-08-23	137.00
78005	2021-08-23	141.00
78006	2021-08-23	139.00
78007	2021-08-23	140.00
78008	2021-08-23	141.00
78009	2021-08-23	139.00
78010	2021-08-23	140.00
78011	2021-08-23	138.00
78012	2021-08-23	139.00
78013	2021-08-23	141.00
78014	2021-08-23	139.00
78015	2021-08-23	145.00
78016	2021-08-23	141.00
78017	2021-08-23	139.00
78018	2021-08-23	138.00
78019	2021-08-23	140.00
78020	2021-08-23	142.00
78021	2021-08-23	139.00
78022	2021-08-23	143.00
78023	2021-08-23	136.00
78024	2021-08-23	141.00
78025	2021-08-23	138.00
78026	2021-08-23	141.00
78027	2021-08-23	136.00
78028	2021-08-23	143.00
78029	2021-08-23	140.00
78030	2021-08-23	141.00
78031	2021-08-23	143.00
78032	2021-08-23	142.00
78033	2021-08-23	141.00
78034	2021-08-23	139.00
78035	2021-08-23	144.00
78036	2021-08-23	144.00
78037	2021-08-23	134.00
78038	2021-08-23	140.00
78039	2021-08-23	139.00

Result	Value
min date=	2021-08-23
max date=	2021-11-30
min date (number)=	44431
max date (number)=	44530
number of days=	100
average count per day =	130
number of positive detected =	0
percentage of positive detected (%)=	0
first sequence number of positive detected =	0
first day of positive detected =	0
first date of positive detected =	NONE



General troubleshooting for poor performing algorithm settings during verification step:

1. The percentage of error detection is too low.

a. Increase the block size. This has the effect of reducing data variability, which will reduce the control limits and improve error detection.
b. Select Box-Cox transformation if data is skewed.
c. Adjust winsorisation percentage lower. This has the effect of reducing the impact of extreme values.
d. Assign a larger bias to be detected. If the data variability is high, it will not be able to detect small biases.
e. Increase data size per day. If the amount of data per day is small, the algorithm will have less probability to detect the simulated bias. Note that this is an artifact of the simulation coded in this Excel and may not manifest in routine use since the algorithm will be applied continuously (and not on a daily basis as in the 'verification' sheet).

2. Box-Cox transformation does not improve performance.
   a. Check the data distribution. If the data without transformed looks normally distributed, the box-cox transformation is not required.
   b. Use the native result without Box-Cox transformation.

Precautions

1. The data source import to training, verification and routine sheet are limited to maximum 40,000 rows data within at most 400 days. The 40,000-row number restriction is applied to reduce the Excel computation time. If more data is input into Excel, a prompt will ask if user wishes to increase the table size, click 'yes' if that is the case but will increase computation time.
2. A warning that the run-time of the spreadsheet may be long if:
   a. The computer has low computation specification or
   b. The data size is large.
3. The number of daily entries is recommended to be at least three times larger than the maximum of block sizes.