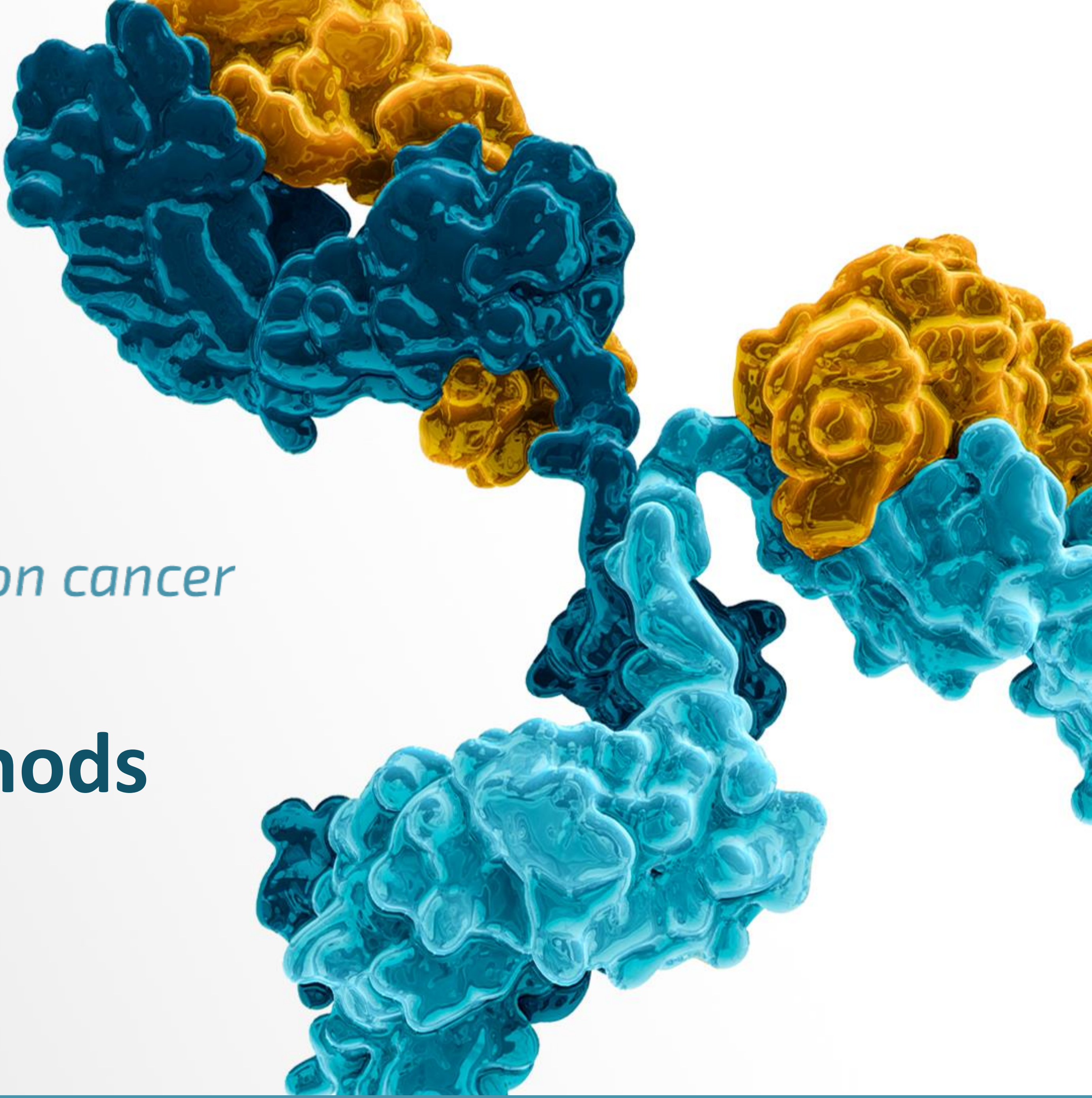
Merus closing in on cancer



Current Statistical Methods for Implementing QTLs

Sonia Amodio, Merus N.V.

On behalf of the PSI SIG on CSM/QTL

Acknowledgments & Disclaimer

This presentation is delivered on behalf of the PSI/EFSPI/ASA-BIOP Special Interest Group on Centralized Statistical Monitoring/Quality Tolerance Limits

Lead: Tim Rolfe (GSK) and Susan Talbot (Amgen)

Members: Rakhi Kilaru (PPD), Sonia Amodio (Merus N.V.), Yasha Li (Bei Gene), Yuling Zeng (Bei Gene), Sharon Love (UCL), Arseny Khaplanov (Paraxel), Zhongkai Wang (Gilead), Chris Wells (Roche)

- The views and opinions expressed in this presentation are solely those of the members of the PSI SIG on CMS/QTL and not of their respective employers
- No confidential business information is presented nor any specific ongoing programs within companies of members of PSI/SIG is discussed
- The primary purpose of this talk is educational – it is intended to raise awareness on statistical methodologies for implementing Quality Tolerance Limits monitoring
- No conflicts of interest to disclose

Historical perspective on QTLs

E6(R2) 2016

5.0.4. Risk Control
… Predefined quality tolerance limits should be established, taking into consideration the medical and statistical characteristics of the variables as well as the statistical design of the trial, to identify systematic issues that can impact subject safety or reliability of trial results. Detection of deviations from the predefined quality tolerance limits should trigger an evaluation to determine if action is needed.

E8(R1) 2021

Integrating Quality by Design (QbD) into study design and conduct

E6(R3) 2025

3.10.1.3 Risk Control
… Where relevant, the sponsor should set pre-specified acceptable ranges (e.g., quality tolerance limits at the trial level) to support the control of risks to critical to quality factors. These pre-specified ranges reflect limits that when exceeded have the potential to impact participant safety or the reliability of trial results. Where deviation beyond these ranges is detected, an evaluation should be performed to determine if there is a possible systemic issue and if action is needed.

Grounded in the foundational principle of QbD - Applying the foundation of E8 to the conduct of clinical trials

QTL definition (TransCelerate 2017)

A level, threshold or value that defines an acceptable range which is associated with a parameter that is critical to quality (CtQ)

Parameter Trial variable to be assessed at trial level, CtQ factor	+	**Threshold** Designed to detect potential issues that could be mitigated	=	**QTL**

"A QTL parameter is appropriate for a mitigable risk if it is detectable, interpretable and measurable" (Keller et al., 2024)

CtQ factors define **WHAT** is worth of measuring

QTLs implementation defines **HOW** to monitor these CtQ factors (setting **ACCEPTABLE DEVIATIONS)**

Examples:

% or # of participants randomized who do not meet inclusion/exclusion criteria	% or # of participants with withdrawal of IC *AND/OR* are lost to follow-up
% or # of participants with premature discontinuation of IP	Etc.

Methods for implementing QTL monitoring

Home > Therapeutic Innovation & Regulatory Science > Article

An Overview of Current Statistical Methods for Implementing Quality Tolerance Limits

Review | Open access | Published: 26 December 2023

Volume 58, pages 273–284, (2024) Cite this article

Download PDF

You have full access to this open access article

Rakhi Kilaru, Sonia Amodio, Yasha Li, Christine Wells, Sharon Love, Yuling Zeng, Jingjing Ye, Monika Jelizarow, Abhinav Balakumar, Maciej Fronc, Anne Sofie Osterdal, Tim Rolfe & Susan Talbot

Statistical Process Control (SPC) Charts:	Bayesian Methods:
• Observed Minus Expected (O-E) Difference Charts • Observed/Expected (O/E) Ratio Charts • Cumulative Proportion Charts	• Beta-Binomial Model • Bayesian Hierarchical Model (BHM)

Merus

Introductory example and notation

- Suppose our QTL is the % of patients who prematurely discontinue treatment.
- The contribution of each patient is identical (patients are independent of each other) and the probability of the event of interest is constant (patients enrolled in the clinical trial are sharing similar baseline characteristics following the same protocol).
- Occurrence of the event of interest for each patient can be regarded as a Bernoulli event

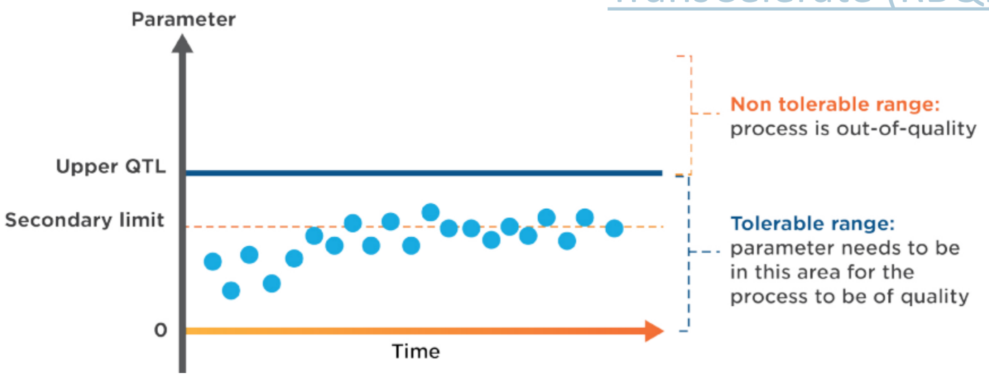
$$X \sim Bernoulli(p)$$

- Let $n = 1, \ldots, N$ be the number of patients randomized at a given time point in the trial

- Since patients are independent, the theoretical distribution of the occurrence of the events in the trial is a Binomial

$$Y_n \sim Binomial(n, p)$$

TransCelerate (RBQM)



Goal of QTL monitoring:

Trigger an action early enough to prevent QTL breach at the end of the trial (and minimizing the False alarm rate)

Merus

SPC Control charts

Frequentist approach, but secondary limits not fixed but dynamic

Observed minus Expected (O-E)

Cumulative sum of differences between Observed events and Expected events

$$Y_n - np_E$$

Secondary limits corresponds to upper 95th quartile of the binomial distribution (asymptotic limits are possible too)

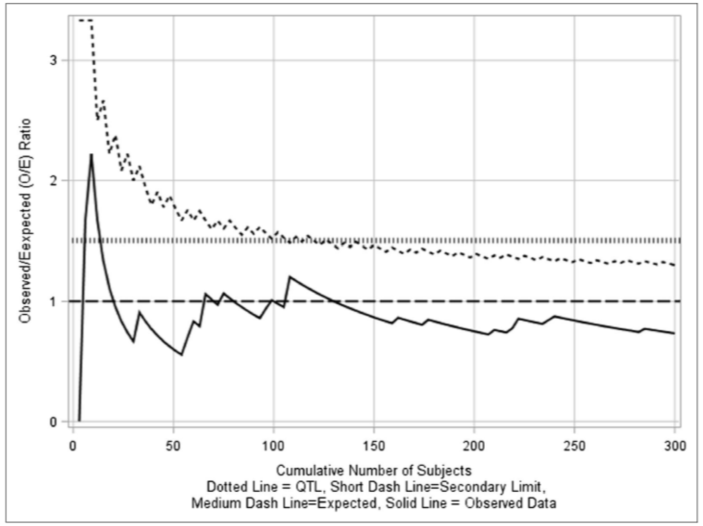


Figure 4 O/E chart on significant protocol deviations.

Cumulative (rolling) proportion

Proportion of events is expected to be constant throughout the trial

Secondary limits can be determined using 95th quartile of binomial distribution, exact limits or asymptotic limits

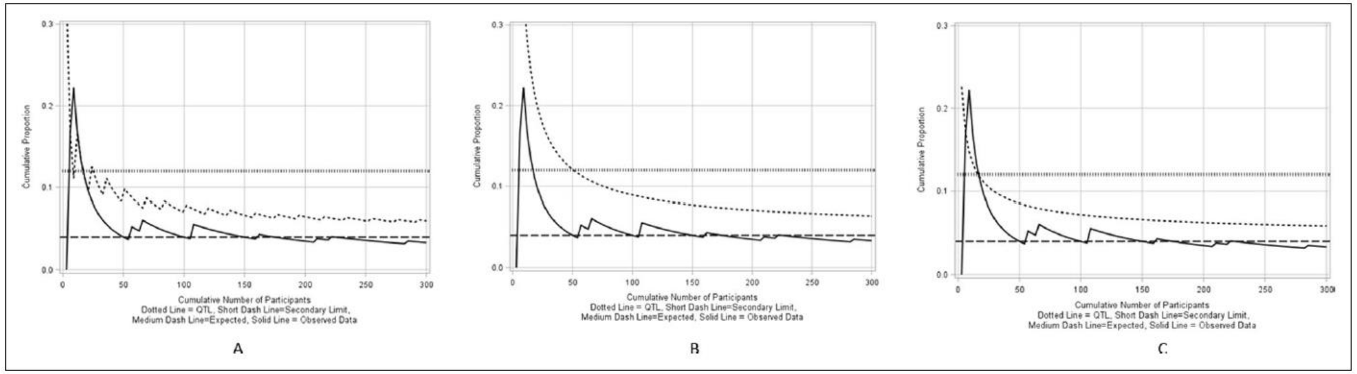


Figure 5 **A** Cumulative proportion chart using a binomial quantile method for an in-control process, **B** Cumulative proportion chart using an exact method for an in-control process, **C** Cumulative proportion chart using asymptotic method for an in-control process.

Beta-binomial model

Historical data can be used for the prior distribution of the unknown parameter $p \sim Beta(n_0 p_0, n_0(1-p_0))$

Then the posterior distribution of $p|y_{n_c}, n_0 p_0, n_0(1-p_0)$ follows a $Beta(n_1 p_1, n_1(1-p_1))$

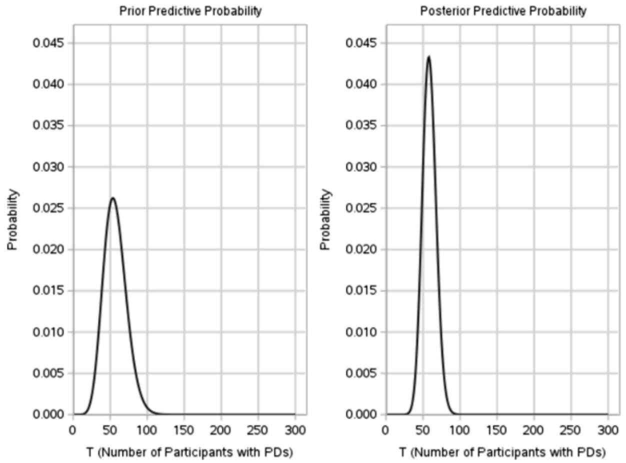
where $p_1 = \frac{n_0 p_0 + y_n}{n_1}$ and $n_1 = n_0 + n_c$

The posterior predictive distribution can be derived as a beta-binomial distribution

The median of the posterior predictive distribution is then used as a measure to assess whether the current observations exceeds the QTL threshold

The QTL monitoring uses the 95th quantile of historical data with the 80th quantile of the prior predictive distribution as secondary limit

Historical data used should be sufficiently homogeneous with current trial data



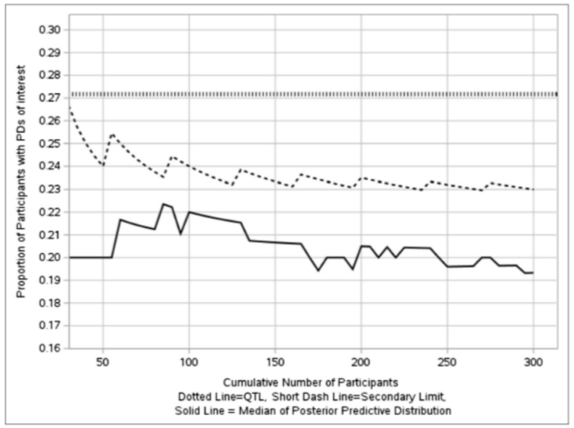


Figure 8 In-control process for proportion of participants with PDs of interest.

Merus

Bayesian Hierarchical Model (BHM)

- BHM for use in QTLs was proposed by the statisticians in the Roche Central Statistical resource team (lead Chris Wells) and inspired by a Bayesian meta-analysis of clinical trials data presented by Berry et al (2010)
- It allows to consider sites heterogeneity $R_i \sim Bin(p_i, k_i)$

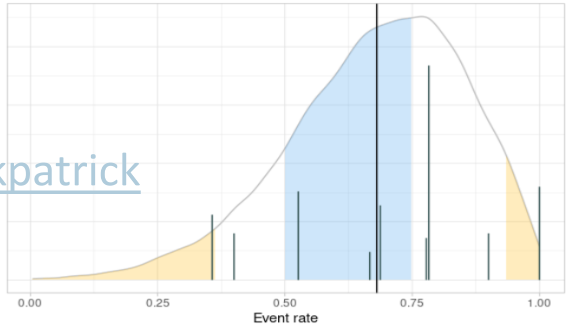
  Patients are nested in sites and probability of event may vary among sites
- Each site proportion is a sample from a population with distribution $\theta_i \sim Beta(\alpha, \beta)$.

  The parameters α and β are hyperparameters that are also random, with distribution $h(\psi)$ and $g(\omega)$, respectively

- BHM uses the quantiles of the posterior distribution of the probability of the event to define the QTL for the metric of interest (M). This can happen in isolation or with reference to historical data obtained by similar previous studies.
- The classification rule for BHM is based on the median $\hat{m}$, that is compared with pre-defined limits $l_U(p)$.

  If $\hat{m} > l_U(p)$ then QTL status is defined as breach, otherwise the study is considered not needing further analysis

rbqmR package by John Kirkpatrick



Merus

Evaluation of current statistical methods for implementing Quality Tolerance Limits


Research Article

Evaluation of current statistical methods for implementing Quality Tolerance Limits

Rakhi Kilaru, Sonia Amodio, Yasha Li, Yuling Zeng, Arseny Khaplonov, Zhongkai Wang, Christine Wells, Sharon Love, Tim Rolfe & Susan Talbot ...show less

Received 26 Nov 2024, Accepted 09 Oct 2025, Accepted author version posted online: 29 Oct 2025

Statistics in Biopharmaceutical Research ›

Latest Articles

Submit an article

Journal homepage

All methods were compared using simulated data with N = 300 subjects and pre-defined probability of event.

Ten different probabilities (0.01, 0.03, 0.05, 0.06, 0.07, 0.1, 0.14, 0.15, 0.20, 0.30) were selected to simulate the data under IC and OOC scenarios

Table 1 Simulation Set-up (Objective 1)

Expected Values	Optimal q (BHM only)	True Values – In Control	True Values – Out of Control	Interim Reviews
0.01	0.80	0.01	0.03	20, 50, 100, 150, 200, 250, 300
0.03	0.87	0.03	0.06	20, 50, 100, 150, 200, 250, 300
0.05	0.83	0.05	0.1	20, 50, 100, 150, 200, 250, 300
0.07	0.79	0.07	0.14	20, 50, 100, 150, 200, 250, 300
0.1	0.77	0.1	0.2	20, 50, 100, 150, 200, 250, 300
0.15	0.795	0.15	0.3	20, 50, 100, 150, 200, 250, 300

For each scenario, 1000 independent simulations were generated.

Aim: provide insights into the performance of different QTLs monitoring methodologies under controlled conditions, using performance metrics

For BHM only, 5 sites were used in each simulation.

Objectives

Objective 3

- Assess starting point to initiate QTL monitoring (20,25,30,35,50)

Objective 2

- Assess methods on how well they performed when considering a confirmed breach (i.e., 2 consecutive breaches).

Objective 1

- Evaluate each method wrt FAR, AR, and ARL over a fixed sequence of interim looks (20,50,100,150,200,250, 300)

Metrics

False Alarm Rate (**FAR**): proportion of in-control processes that are incorrectly identified as out-of-control

Alarm Rate (**AR**): proportion of out-of-control processes that are correctly identified as out-of-control

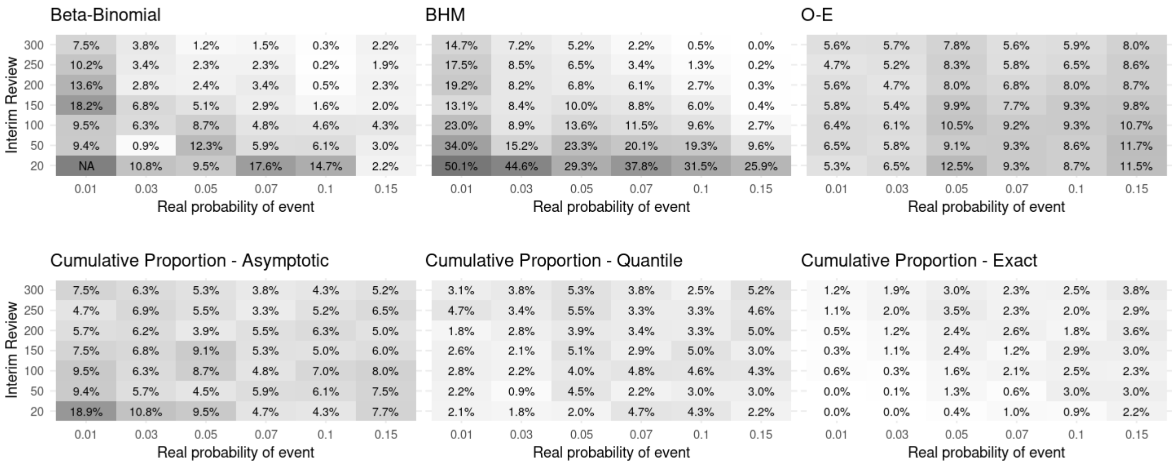
Average run Length (**ARL**): average number subjects at which a breach occurs for the first time among all interim looks:

- **In-control ARL**: average number of subjects before a false alarm occurs
- **Out-of-control ARL**: average number of subjects before a true alarm occurs

Merus

Objective 1: fixed sequence of interim looks

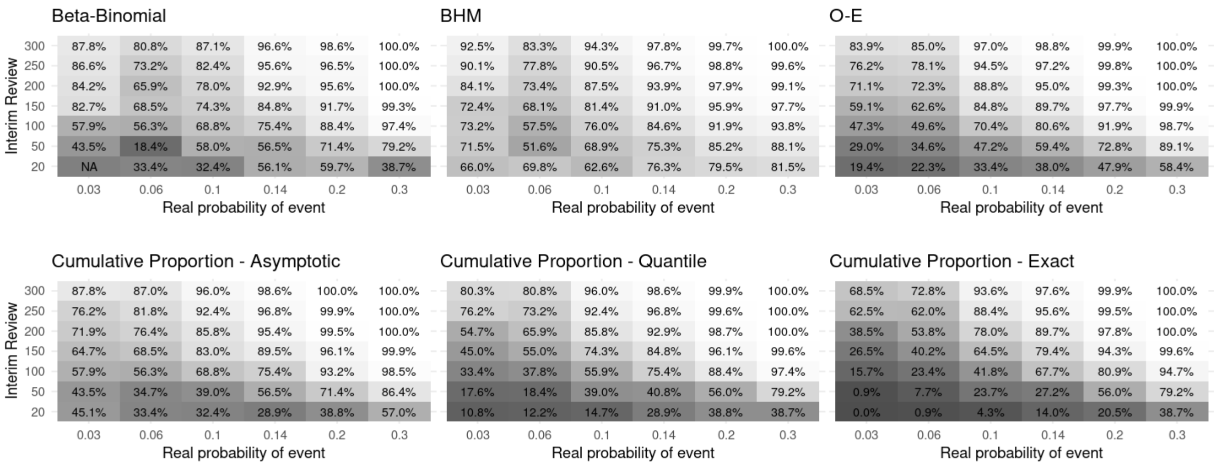
In-Control processes (FAR)



Objective 1

- Evaluate each method wrt FAR, AR, and ARL over a fixed sequence of interim looks (20,50,100,150,200,250, 300)

Out-Of-Control processes (AR)



- SPC methods perform well (especially cumulative proportion charts) when p_E and sample size (n) are sufficiently large
- BHM tends to work better than most method for early detection in OOC processes, but less well in IC processes.
- Beta-binomial does not yield reliable results with a small sample size ($n < 20$) and a low event rate ($p_{\text{exp}} = 0.01$), but performs well for higher p_E and is comparable with BHM for OOC processes

Merus

Objective 2: confirmed breaches

Table 2 Simulation Results (Objective 2)

Real probability of event (For out-of-control)	Expected probability of event	Method	ARL (in)	ARL (out)	FAR (unconfirmed)	FAR (confirmed)	AR (unconfirmed)	AR (confirmed)
0.03	0.01	Beta-Binomial	323	84	19.6%	19.1%	99.4%	98.9%
		BHM	233	85	48.2%	26.5%	97.5%	92.2%
		O-E	328	133	14.7%	7.4%	88.1%	85.0%
		Cumulative Proportion - Asymptotic	312	82	18.1%	9.8%	91.4%	82.3%
		Cumulative Proportion – Quantile	348	124	7.7%	4.6%	84.0%	74.7%
		Cumulative Proportion – Exact	364	175	2.1%	0.9%	71.8%	60.3%
0.06	0.03	Beta-Binomial	362	153	5.6%	4.4%	88.1%	83.7%
		BHM	298	121	2.7%	12.0%	91.6%	81.4%
		O-E	324	126	16.5%	8.3%	90.1%	87.9%
		Cumulative Proportion - Asymptotic	317	82	17.3%	9.1%	92.1%	86.2%
		Cumulative Proportion – Quantile	349	114	7.5%	3.8%	86.0%	85.3%
		Cumulative Proportion – Exact	360	151	3.6%	1.9%	77.0%	62.9%
0.1	0.05	Beta-Binomial	352	105	8.0%	7.6%	94.1%	91.9%
		BHM	278	86	31.3%	13.4%	97.5%	93.3%
		O-E	295	87	25.7%	12.7%	98.1%	97.4%
		Cumulative Proportion - Asymptotic	322	60	15.6%	9.0%	97.6%	94.2%
		Cumulative Proportion – Quantile	332	65	12.4%	6.7%	97.4%	93.7%
		Cumulative Proportion – Exact	351	82	6.6%	3.5%	95.5%	89.1%

Objective 2

- Assess methods on how well they performed when considering a confirmed breach (i.e., 2 consecutive breaches).

- ARL is generally high in IC processes and low for OOC processes across all methods.
- Alarm rate (AR) is generally high in both unconfirmed and confirmed breaches
- FAR for confirmed breaches is generally lower than for unconfirmed breaches
- Confirmation approach can reduce false alarm rate in monitoring.

Objective 3: interim looks starting point

Expected probability = 0.07

ARL (in)

	20	25	30	35	50
O-E	334	338	343	340	345
Cumulative Proportion - Quantile	334	338	343	340	345
Cumulative Proportion - Exact	354	356	353	356	357
Cumulative Proportion - Asymptotic	321	309	319	315	329
BHM	206	242	244	257	281
Beta Binomial	293	339	327	339	339

First Interim

ARL (out)

	20	25	30	35	50
O-E	84	88	90	88	96
Cumulative Proportion - Quantile	84	88	99	88	96
Cumulative Proportion - Exact	105	107	103	107	112
Cumulative Proportion - Asymptotic	74	68	73	70	84
BHM	39	46	50	55	69
Beta Binomial	65	90	81	90	93

First Interim

AR

	20	25	30	35	50
O-E	98.9%	98.9%	98.9%	98.9%	98.9%
Cumulative Proportion - Quantile	98.9%	98.9%	98.9%	98.9%	98.9%
Cumulative Proportion - Exact	97.9%	97.9%	98.0%	97.9%	97.9%
Cumulative Proportion - Asymptotic	99.0%	99.0%	99.1%	99.1%	99.0%
BHM	99.5%	99.3%	99.3%	99.3%	99.2%
Beta Binomial	98.3%	97.7%	97.8%	97.7%	97.7%

First Interim

FAR

	20	25	30	35	50
O-E	13.7%	12.5%	11.3%	12.0%	10.8%
Cumulative Proportion - Quantile	13.7%	12.5%	11.3%	12.0%	10.8%
Cumulative Proportion - Exact	7.3%	6.7%	7.5%	6.8%	6.4%
Cumulative Proportion - Asymptotic	17.6%	20.6%	18.0%	19.2%	15.6%
BHM	48.8%	39.4%	39.1%	35.6%	29.5%
Beta Binomial	23.7%	11.6%	14.6%	11.6%	11.6%

First Interim

Objective 3

- Assess starting point to initiate QTL monitoring (20,25,30,35,50)

- All methods seem to perform well for IC processes for varying p_E (BHM initially worse but becoming incrementally better as p_E increases)
- All methods perform reasonably well for OOC processes (especially with increase in sample size selected for the first interim look).
- BHM performs quite well for OOC processes

Merus

Conclusions

- Fixed limits are often set to align with what is desired rather than considering historical data and/or expert knowledge
- While there may be practical advantages of using fixed limits as early warning signals, the use of fixed limits as QTLs can be ambiguous
- Investigating a breach is a costly activity (i.e., it requires multiple perspectives and personnel performing the necessary due diligence to investigate root cause) and hence, setting the fixed limits too low can result in a false alarm resulting in wasting precious resources.
- **Statistically based monitoring methods have the potential for better operating characteristics than simple, fixed thresholds**

Take home messages

There is no best and no worst statistical method to perform QTL monitoring

- Cumulative proportion charts perform well for both IC and OOC scenarios

- Observed minus Expected and Beta Binomial model perform moderately well

- BHM works quite well in detecting breaches when process is OOC at earlier interim review, but not for IC processes especially with lower expected probabilities

The choice of the method to use to monitor QTLs should not be only based on overall performance

There is a **trade-off** between method's performance, statistical complexity, strength of assumptions, and feasibility that needs to be factored into the choice of the method to use.

Advantages of statistical methods thresholds over fixed limits

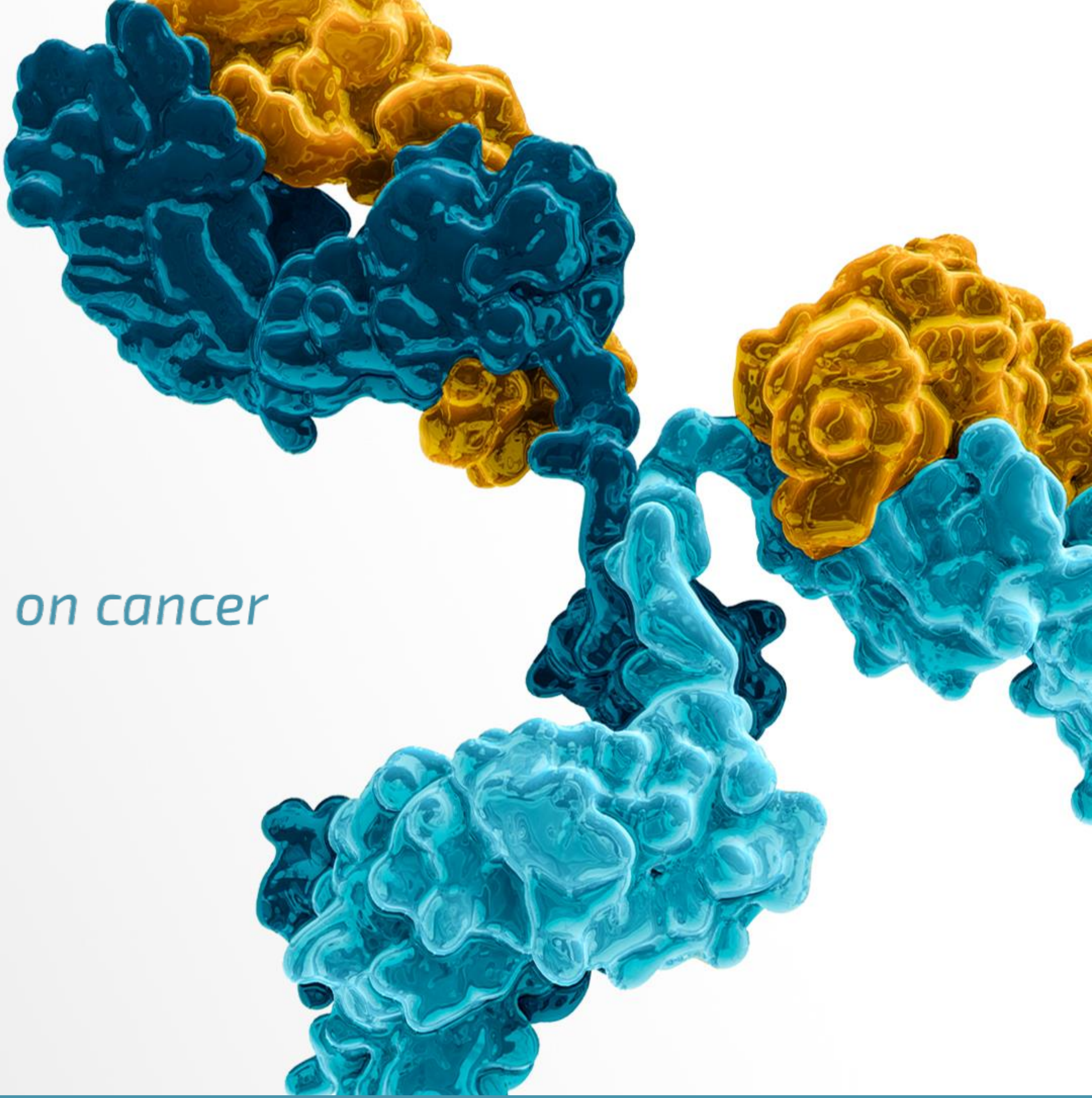
Statistically based monitoring methods seem to have better operating characteristics over fixed limits, which may be still advisable for small and early phase studies where close follow-up is not feasible

Selected references

- R. Kilaru, S. Amodio, Y. Li, et al., “An Overview of Current Statistical Methods for Implementing Quality Tolerance Limits,” Therapeutic Innovation & Regulatory Science 58 (2023): 273–284. Available at link
- R. Kilaru, S. Amodio, Y. Li, et al., “Evaluation of current statistical methods for implementing Quality Tolerance Limits”, Statistics in Biopharmaceutical Research, 1-14. Available at link
- EMA. Reflection paper on risk-based quality management in clinical trials. Published November 18, 2013
- Clinical Trials Transformation Initiative. Critical to Quality (CTQ) Factors Principles Document
- Risk-based quality management: quality tolerance limits and risk reporting. TransCelerate website: Risk-Based-Quality-Managment.pdf.
- Bhagat R, Bojarski L, Chevalier S, Görtz DR, Le Meignen S, Makowski M, Nadolny P, Pillwein M, Suprin M, Turri S. Quality Tolerance Limits: Framework for Successful Implementation in Clinical Development. Ther Innov Regul Sci. 2021 Mar; 55(2): 251-261
- Steven A. Gilbert. Implementing Quality Tolerance Limits at a Large Pharmaceutical Company. PharmaSUG 2020 – Paper SA-284 Accessed June 2023. Available at link
- Berry SM, Carlin BP, Lee JJ, Muller P. Bayesian Adaptive Methods for Clinical Trials. Boca Raton: CRC Press; 2010.
- Whitehead, M., Suprin, M., Mistree, T. et al. The Renovation of Good Clinical Practice: A Framework for Key Components of ICH E8. Ther Innov Regul Sci 58, 303–310 (2024)
- Keller, A., van Borrendam, N., Benner, P. et al (2024) Quality Tolerance Limits: A General Guidance for Parameter Selection and Threshold Setting. Ther Innov Regul Sci 58, 423-430. Available at link
- Poythress JC, Lee JH, Takeda K, Liu J. Bayesian Methods for Quality Tolerance Limit (QTL) Monitoring. Pharm Stat. 2024 Aug 9. doi: 10.1002/pst.2427. Epub ahead of print. PMID: 39119894

Merus

Thank you



Merus closing in on cancer

www.merus.nl

s.amodio@merus.nl