Poisson Modeling for LOD assessment for BCRABL Testing

Zhonggai Li* Tianlei Chen* Brian Mullaney*

Abstract

The tyrosine kinase inhibitors Gleevec and Tasigna are highly efficacious therapies for Philadelphia chromosome-positive chronic myeloid leukemia (Ph+ CML). Monitoring of minimal residual disease of BCRABL transcript is standard of care and the BCRABL/ABL ratio (IS) decreases multi-log with chronic treatment. Therefore highly sensitive BCRABL assays are required which can detect the BCRABL/ABL ratio (IS) to 4.5log10 below the IS baseline (MR4.5). Most BCRABL assays utilize qPCR technology with ABL as control gene for relative quantification. In the routine BCRABL testing community, when the BCRABL is not detected (ND), the assay algorithm will substitute ND with 1 copy for calculation and use (1/ABL copy)*IS Conversion Factor to report the BCRABL/ABL ratio (IS). However, in IVD practice, a more optimal approach will substituted with an assay specific LOD, which is higher than 1 copy due to sampling error in these very low copy number samples. This paper will use Poisson modeling to construct the connection and separation between the qPCR assays and sampling error, which will offer quantifiable tool to evaluate the difference and appropriateness of substitute 1 BCRABL copy vs LOD of BCRABL copy in the clinical diagnostics.

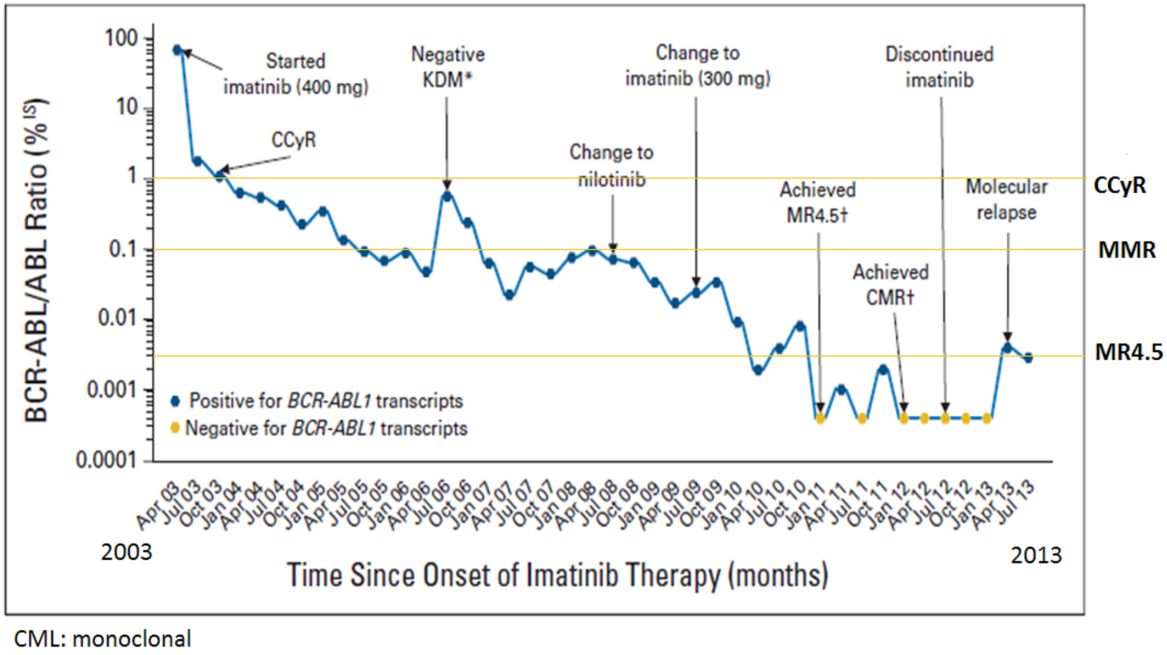
Key Words: BCRABL, CML, Poisson, LOD

1. Background

Chronic myeloid leukemia (CML) is a subtype of leukemia which is defined by the chromosomal translocation t(9;22) , also referred to as the "Philadelphia Chromosome". The translocation results in fusion of the BCR gene and the ABL kinase gene, ultimately leading to increased myeloid cell proliferation. Monitoring of minimal residual disease of BCRABL fusion transcript can easily be measured in blood by qPCR. Monitoring has become standard of care part of treatment and the BCRABL/ABL ratio (IS) decreases multi-log with chronic therapy. The disease burden for BCR-ABL positive chronic myelogenous leukemia (CML) patients is measured by comparing the expression level of the fusion gene BCR-ABL to a reference gene, such as ABL1. The level of BCR-ABL expression is typically reported using either a log reduction scale, established by the IRIS clinical trial laboratories (Hughes, T.P. et al., 2003), or an international scale (IS) designed to replace the log reduction scale (Branford, S. et al., 2008). Standard nomenclature has been established for reporting such that for example a 3-log reduction of transcript from baseline is described as Molecular Response (MR3.0). As Gleevec and Tasigna are highly efficacious therapies for CML, treatment will occur over months to years. The BCRABL/ABL ratio (IS) typically decreases over multi-log level during this time (See Figure (1) from Marin JCO (2014) 32:379); therefore,

*Novartis Oncology, 45 Sidney Street Cambridge, MA 02139

a highly sensitive BCRABL assay that can detect the BCRABL/ABL ratio (IS) to 4.5log10 below the IS baseline (MR4.5). In addition newer clinical trial designs are exploring molecular endpoints involving MR3.0, MR4.0, and MR4.5, etc.



CML: monoclonal

Figure 1: Marin JCO (2014) 32:379

The challenge with highly sensitive assays is that, it is difficult to precisely determine the BCRABL copy near MR4.5 and therefore compute a precise BCRABL/ABL ratio. Even if the assay is capable of single-molecule detection, BCRABL/ABL copy number cannot be exactly measured at limiting dilutions due to two sources of variation. One source is the quantitative polymerase chain reaction (qPCR), a technology that most of BCRABL assays use. A qPCR is a laboratory technique of molecular biology based on the polymerase chain reaction (PCR), which is used to amplify and simultaneously quantify a targeted DNA molecule. It is the method of choice for accurate estimation of gene expression. However, the qPCR test is a complex measurement system. It is not possible to directly measure the exact copy number in the tested sample. Besides the measuring error from the qPCR technology, another source of variation is the sampling process. Consider an experiment performed at limiting dilutions, suppose one creates a large bulk solution at this average concentration, and then dispenses from this bulk a large number of seemingly identical replicates. Each replicate contains λ copies of analyte on average. But in fact, the replicates are not identical, because they will contain varying numbers of analyte molecules. (The copy numbers of analyte molecules will follow a Poisson distribution with mean equal to the target average λ given

certain assumptions. More details about the appropriateness of assuming Poisson distribution are discussed in Section 2.)

In quantitative and qualitative molecular measurement procedures, the LOD is defined as the lowest concentration of analyte that can be consistently detected (typically 95% of samples tested under routine clinical laboratory conditions and in a defined type of sample). (refer to James F. Pierson-Perry, et. al (2012)) LOD is used in various ways in the result interpretation for BCRABL tests. When the observed BCRABL copy is less than LOD, it's not necessarily to conclude that no detection of BCRABL in the sample and use 0 as the final report value for the BCRABL copy. Instead, several options are available, as discussed in Section 3. The target of this report is not to discuss how to set a LOD in the BCRABL assay, but under a given LOD, we want to know which result interpretation option can minimize the difference or bias between the reported BCRABL copy and the true target value. Our purpose is to evaluate the bias of these options quantitatively and derive an overall most appropriate option for clinical use recommendation.

In Section 2, we will build a model which employs Poisson distribution to describe the distribution of analyte copy numbers in the sample replicates, and uses normal distribution or uniform distribution to approximate the variation from the qPCR assay. With this model, in Section 3 we will evaluate the bias generated by various result interpretation options. We will also discuss the suitability of the model assumptions in Section 4.

2. Model and Assumptions

As discussed in Section 1, if the target sample copy number λ is given, the variation of the qPCR output comes from two sources. One is the sampling error and the other is the qPCR measurement error. This implies that the observation k_i^* can be decomposed to two components, as shown in the following formula

$$k_i^* = \max(k_i + \epsilon_i, 0), \tag{1}$$

where k_i corresponds to the varying number of analyte molecules in the tested sample, which follows a Poisson distribution Poisson(λ), and ϵ_i corresponds to the qPCR measurement error with either a normal distribution $N(0, \sigma^2)$ or a uniform distribution $U[-d, d]$. We further assume k_i and ϵ_i are all independent. In practice, a negative observation does not exist. So if $k_i + \epsilon_i < 0$, the result would be simply reported as non-detection and 0 may be assigned as the observed value.

The rationale of assuming that k_i is Poisson distributed is as followed: suppose the analyte molecules are sparse in the container (human body), after proper mixing the material (blood) in the container, each analyte molecule could be present anywhere in the container (e.g. spot A, B, C ...); we can assume that an analyte molecule is present in any spot A and any spot B with equal probability. Now one takes a sample (e.g. 2.5 ml blood) from the container (human body). If this sample's volume is small enough compared to the container, the number of analyte molecules in that sample should follow a Poisson distribution. Mathematically we can prove it this way: suppose we have m molecules in the container in total, and each one can be in any spot (n spots in total), with equal probability. ($\lambda = m/n$ is fixed.) We randomly choose a spot to use as a sample

replicate, the number of the molecules in that spot can be approximated by a Poisson distribution with mean λ as $n \to \infty$. (It can be mathematically proved and demonstrated by simulation, see Figure (2) for an example). Another explanation is: the sample replicate's volume is $1/n$ of the container's, so each molecule has $1/n$ chance to exist in this sample replicate. The number of the analyte molecules in the sample have a Poisson distribution as $n \to \infty$ according to Poisson Approximation Theorem. See Papoulis, A. and Pillai S. U. (2002) p. 112.

In summary, two conditions must be satisfied to approximately assume Poisson distribution:

- An analyte molecule is present in any spot A and any spot B with equal probability;
- The sample taken from the container is small enough compared to the container.

The critical point of assuming model (1) is not to estimate the parameters λ, σ or d from the observations k_i^* to explain the assay, but to use this model to calculate the bias of different options when k_i^* <LOD and compare them to find the best option.

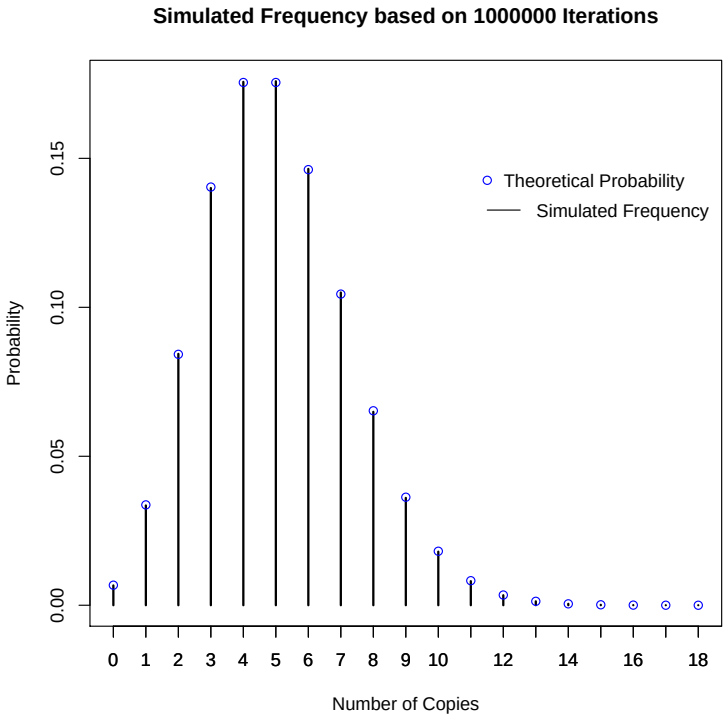


Figure 2: Simulation result by choosing $n = 200$ (the sample's volume is $1/200$ of the container's), $m = 1000$, $\lambda = 5$. The empirical frequency and the probability from the mass function of $Poisson(5)$ are almost identical.

3. Comparison of Options

After obtaining k_i^*, some practice is implemented to report BCRABL copy value. Some frequently used options are listed below:

Our target is to evaluate the clinical impact (bias) of the above options of test result interpretation. We denote the final report BCRABL value as Y^*, which is obtained from k_i^* through the above options. And the bias is defined as $Bias = E(Y^*) - \lambda$, where λ is the true target level. Assuming model (1), we can compare the bias from different options under different σ or d and under different LODs.

It's easy to mathematically prove that Option d is always inferior to Option c (to prove $Bias$(option c)$<$ $Bias$(option d)); Option f is always inferior to any other option unless the target level λ is very small. So from now on we are only concerned with the competitive candidates Option a, b, c, e.

When the measurement error is negligible ($\sigma = 0$ or $d = 0$), which means we can exactly observe the integer k_i, the absolute bias curves are in the following plots and the integral of the absolute bias curve corresponding to each option is summarized in Table (1). From the table, we can see that when LOD=3 or 4 Option c is the best overall choice under the criterion of cumulative absolute bias; when LOD=2 Option b and c are the same options and they have the same cumulative absolute bias as Option a; when LOD=1 level, Option a has the smallest cumulative absolute bias.

	Option a	Option b	Option c	Option e
$LOD = 1$	0	1	1	1
$LOD = 2$	1	1	1	3
$LOD = 3$	3	1.17	1	6
$LOD = 4$	5.96	3.01	1	9.98

Table 1: Cumulative absolute bias when measurement negligible

When the measurement error is not negligible ($\sigma > 0$ or $d > 0$), we further consider two conditions: measurement error is either normal distributed or uniform distributed. First, we consider normal distribution for ϵ_i in model (1). Denote $k_i' = k_i + \epsilon_i$. The density function of k_i' becomes:

$$f_{k_i'}(x) = \sum_{j=0}^{\infty} f_0(x - j)P(k_i = j|\lambda), \tag{2}$$

where f_0 is the density function of $N(0, \sigma^2)$. So the random variable k_i^* has a distribution of mixed type, with $P(k_i^* = 0) = \int_{-\infty}^{0} f_{k_i'}(x)dx$, and $f_{k_i^*}(x) = f_{k_i'}(x)$ when $x > 0$. From this distribution function it's easy to find the distribution of Y_i^*, and the calculation of the absolute value of the bias is straightforward. The absolute value of the bias bears a similar shape as in Figure (3).

When σ is very close to 0, the numerical method for integral when computing $E(Y^*)$ is not stable, but we can use theoretical method to obtain an approximation and the conclusion would be similar to the $\sigma = 0$ case. When σ is large enough (e.g. $\sigma > .17$), the numerical method is applied and Figure (4), Figure (5) and Table (2) give a summary of the results.

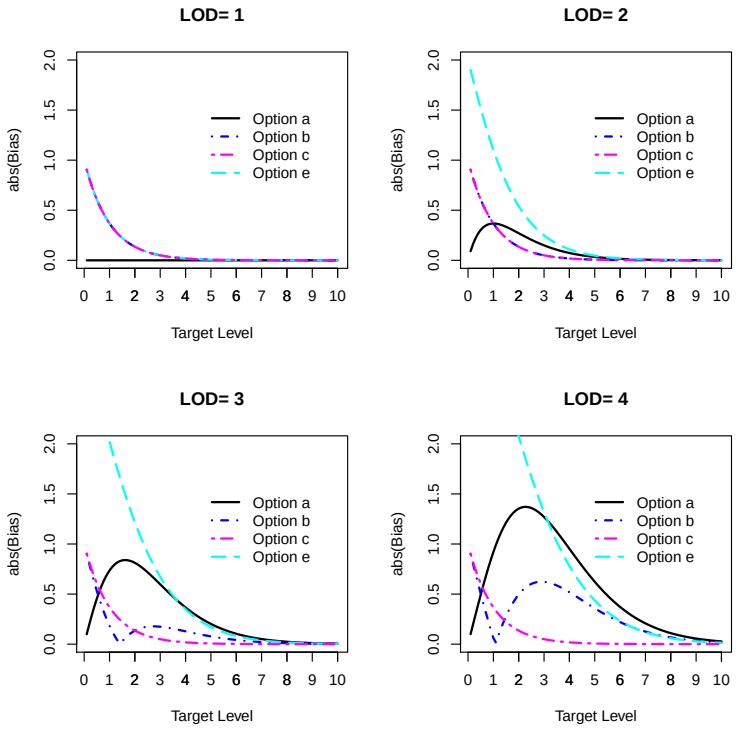


Figure 3: Bias curves when the measurement error negligible

	$\sigma = 0.3$				$\sigma = 2.5$			
	a	b	c	e	a	b	c	e
$LOD = 1$	0.40	1.06	0.59	1.06	1.72	3.58	3.28	3.58
$LOD = 2$	1.88	0.82	0.59	3.02	1.02	3.19	3.28	5.82
$LOD = 3$	4.36	1.92	0.59	5.97	2.01	2.24	3.28	8.87
$LOD = 4$	7.79	4.29	0.59	9.90	4.73	2.67	3.28	12.78

Table 2: Cumulative absolute bias

From Figure (4), Figure (5) and Table (2), we can see that Option a, Option b and Option c are three competitive candidates for the best option. If σ is small, absolute bias curves are similar to the curves in $\sigma = 0$ case (compare Figure (3) and Figure (4)). When LOD=1, Option a is always the best option under all true target BCR-ABL level (λ from 0 to 10). If LOD=2, 3, 4, Option a, Option b and Option c are the best options under different true target BCRABL level. In the low end of target BCRABL level ($0 < \lambda <$a certain level) Option a gives the smallest absolute bias,

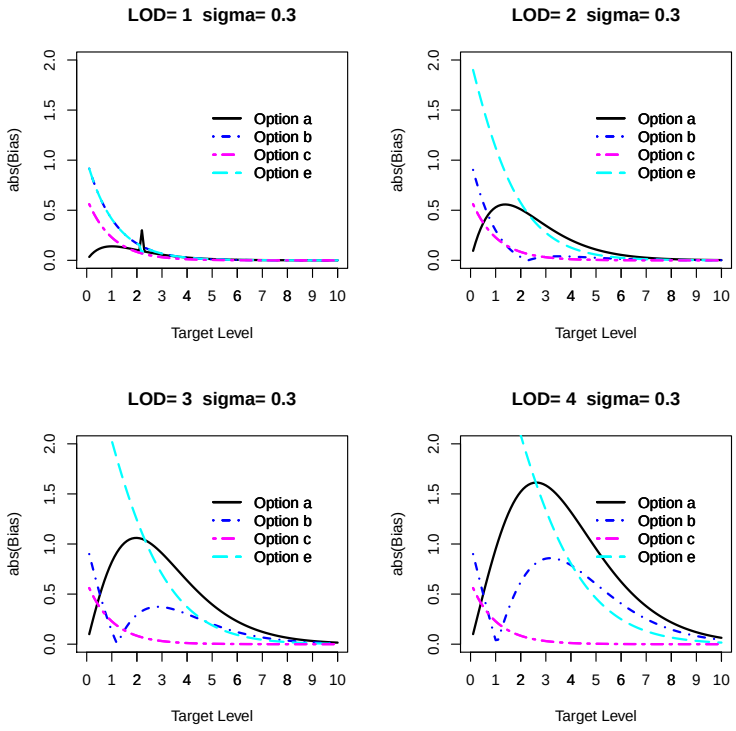


Figure 4: Bias curves when $\sigma = 0.3$

while in the high end of target BCRABL level (a certain level$< \lambda < 10$) Option c performs best. Option b is the best in a small range in the middle of the BCRABL level (e.g. $1 < \lambda < 1.5$ when LOD=3 and $\sigma = 0.3$). Also notice that as σ increases, the interaction point of the blue line (Option b) and the black line (Option a) moves to the right under each LOD level (LOD> 1), which means Option a's performance improves under larger measurement error.

If we use the cumulative absolute bias as a criterion, we can obtain Figure (6). When LOD=1, Option a is the best option no matter how λ changes. When LOD=2, the absolute bias curves of Option b and Option c bear similar shapes and their cumulative absolute bias are also close. If the variation of measurement error is not very large ($\sigma < 1.2$), Option b and Option c are better than Option a. If $\sigma > 1.2$, Option a is the best one. When LOD=3, Option c is the best if $\sigma < 1.3$; Option b is the best if $\sigma > 1.3$. When LOD=4, Option c is the best option ($\sigma < 2.2$). Overall speaking, Option c is the recommended option because LOD in BCR-ABL assay will be from 2 to 3 and the variation from the qPCR is not very large ($\sigma < 1.2$).

If we assume ϵ_i has the uniform distribution, we obtain similar figures and similar tables. See Figure (7), Figure (8), Figure (9) and Table (5) in Appendix B.

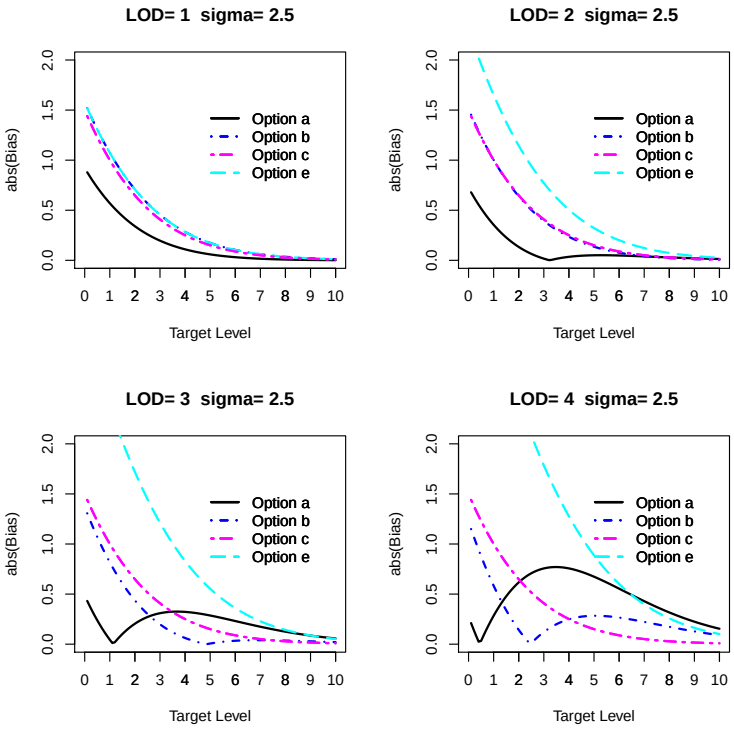


Figure 5: Bias curves when $\sigma = 2.5$

4. Goodness of Fit

In this section we will discuss whether the assumption of model (1) is appropriate. Although it is reasonable to assume this model theoretically and mathematically, it is always better to find more evidence from the real world assay data to support the model. Under given parameter λ and σ(or d), we can derive the close form of the distribution of k^*. And we can apply KS-test to a list of observed k_i^*s, to see if the observations give any evidence against the distribution assumption. But the problem is that we need to estimate the parameters λ and σ(or d) first, which is not realistic with the limited information only from the observations. The problem is that it seems not possible to separate k_i and ϵ_i from k_i^* without further experiment, if we only have a table of independent observations k_i^*. So it's not possible to directly test whether k_i comes from a Poisson distribution and whether ϵ_i has a normal/uniform distribution or further estimate λ and σ(or d).

As discussed above, we need stronger assumption to do the goodness of fit test. Considering that usually the variation from the sample is more dominant rather than the error from the qPCR assay, we assume that the error is small enough such that $P(|\epsilon_i| > .5) \approx 0$, which means $\sigma < .17$

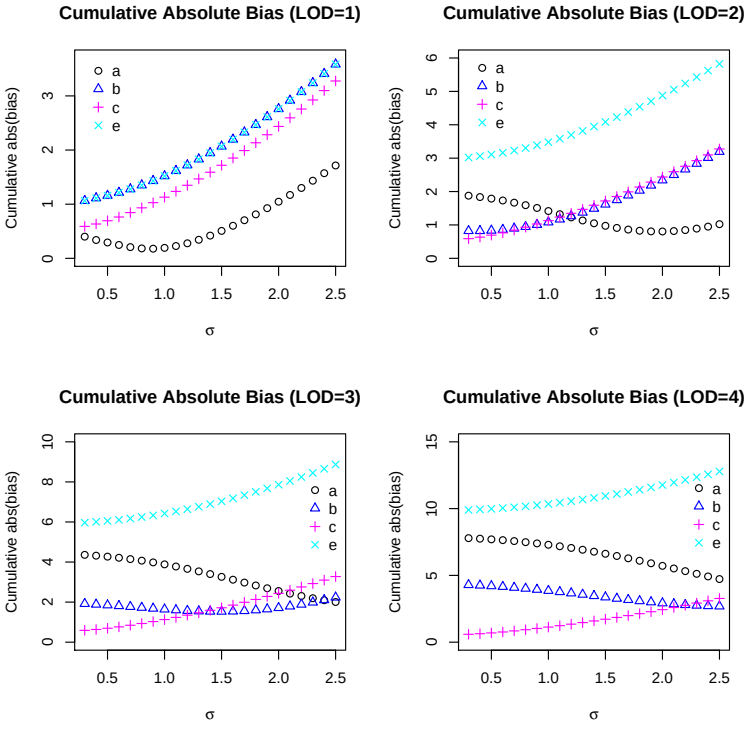


Figure 6: Cumulative absolute bias

or $d < .5$. With this assumption, we can directly decompose k_i^* into two known components k_i and ϵ_i and estimate λ and σ separately. Denote κ_i the largest integer less than k_i. If $k^* > 0$ we have $\epsilon_i = \min(k_i^* - \kappa_i, \kappa_i + 1 - k_i^*)$, and $k_i = k_i^* - \epsilon_i$. If $k_i^* = 0$, then $k_i = 0$ and $\epsilon_i < .5$ but still unknown.

If we have the above assumption, we can try the following data to see if it supports the validity of model (1). A detailed experiment was conducted at MolecualrMD testing laboratory (Portland, OR) in Q1 2013, using test materials provided by Novartis. [03-april-2013, CTA LoD data, Julie Toplin; TY folder # IVDD-3109] Some key results from this experiment are summarized in Table (3), (4) in the appendix. The study involved the serial dilution of a patient PAXgene blood sample from Novartis into healthy donor PAXgene blood, to create replicates at each of a series of targeted dilutions. Each sample was extracted by a PAXgene manual process, and the resultant purified RNA was assayed in duplicate reactions with the MRDx test.

Since the observations can be decomposed into two components, we can check the goodness of fit for both Poisson assumption and normal distribution/uniform distribution. We use Chi-Square goodness of fit test to check Poisson distribution assumption. We use QQ plot/Shapiro-Wilk

Test/KS Test to check normality assumption or uniform distribution assumption. All of these tests do not give a significant p-value against the distribution assumptions.

5. Discussion and Conclusion

5.1 Discussion

The construction of model (1) is not natural when k_i is close to 0. Model (2.1) implies that a negative value $k_i + \epsilon_i$ may first exist, then be replaced with 0, which may not be very realistic. Instead, the ϵ_i may have an asymmetric distribution with a longer right tail and cannot take any value less than $-k_i$. In fact this implies that the distribution of ϵ_i is related to k_i. But a new distribution assumption for ϵ_i will make the model very complicated, so model (1) is still preferred for comparison for options.

The Poisson distribution and the normality assumption are checked at the target level $\lambda = 1$ and $\lambda = 3$ in Section 4. It is better to further check these assumptions at different target levels. We can also compare the options using the criterion of cumulative MSE but since the main concern is the bias, we don't pursue the option choice this way. The suggestion of choosing option b is based on the condition that the observation k^* is a single observation. If the assay takes a few observations and combine them to obtain a final score (e.g. take the average of two observations), then we need to do further investigation.

5.2 Conclusion

The absolute bias curves and the cumulative absolute bias are similar under normal assumption and uniform distribution assumption for the measurement ϵ_i. When LOD=1, Option a always temps to be the best option, but in the practical BCRABL assays LOD is usually between 2 and 3. When LOD=2, Option b and Option c are the best if the variation from qPCR is not too large, otherwise Option a is the best. When LOD=3, Option c is still the best option if the variation from qPCR is not too large, but as σ or d increases, Option b is becoming to perform better than Option c. When LOD=4, Option c is the best one.

Overall, reporting BCRABL = 1 when BCR-ABL is not detected is the best option under the criterion of cumulative absolute bias among the several options that are used in different communities, because the actual LOD is from 2 to 3 and the measurement errors from qPCR are usually not very large. Since the bias is the most important concern in this area, this option is recommended to use.

REFERENCES

Pierson-Perry, J. F., et. al (2012), "Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; approved Guideline - Second Edition,".

Papoulis, A., Pillai, S. U. (2002), "Probability, Random Variables and Stochastic Processes," The McGraw-Hill Companies, Inc. New York. p. 112

A. MolecualrMD Data

Extraction	Sample	Ratio	BCR-ABL	ABL
Ext1	Q958R1	0.0018	2.585	143897.1
Ext1	Q958R2	0.000779	1.007	129336
Ext1	Q958R3	0.00299	3.745	125314.9
Ext1	Q958R4	0.00349	4.39	125880.5
Ext2	Z531R1	0.00106	1.18	111572.7
Ext2	Z531R2	0.00308	3.27	106239.1
Ext2	Z531R3	0.00294	3	102148.1
Ext2	Z531R4	0.00241	2.43	100759.9
Ext3	U478R1	0.00319	4.93	154374.9
Ext3	U478R2	0.00405	6.36	156962.2
Ext3	U478R3	0.00174	2.92	167349.1
Ext3	U478R4	0.00262	4.18	159756.9
Ext4	N519R1	0.00411	5.55	134959.8
Ext4	N519R2	0.0015	2.03	135028
Ext4	N519R3	0.00331	4.66	140919
Ext4	N519R4	0.00192	2.89	150160.2
Ext5	Z988R1	0.00155	2.33	149879.2
Ext5	Z988R2	0.00242	3.32	136986.2
Ext5	Z988R3	0.00272	4	147159
Ext5	Z988R4	0.00155	2.28	146737.7

Table 3: BCR-ABL copy number/rx targeted to mean value = 3. Observed mean value = 3.4, observed median value = 3.1 (N=20). Yellow highlight indicates 1 of 2 post-extraction replicate reactions failed, in which case the reported value is for the other reaction of the pair.

B. Bias under Uniform Measurement Error

Extraction	Sample	Ratio	BCR-ABL	ABL
Ext1	N790R1	0.00186	2.61	140037.8
Ext1	N790R2	0.000713	9.61E-01	134777.8
Ext1	N790R3	0.000719	0.9745	135443.1
Ext1	N790R4	0.000684	1.004	146865.3
Ext2	W396R1	0.00208	2.28	109846.3
Ext2	W396R2	0.000899	9.99E-01	111121
Ext2	W396R3	0.000549	6.22E-01	113345.4
Ext2	W396R4	0.00216	2.49	115301.3
Ext3	F675R1	0.000689	1	145127.6
Ext3	F675R2	0.0021	3.34	159303.8
Ext3	F675R3		0	149454.5
Ext3	F675R4	0.00114	1.67	146786.4
Ext4	L557R1	0.000562	7.68E-01	136649.1
Ext4	L557R2	0.000619	7.34E-01	118608.2
Ext4	L557R3	0.000674	8.72E-01	129338.8
Ext4	L557R4	0.000901	1.24	137598
Ext5	S687R1	0.00117	1.5	128282
Ext5	S687R2	0.00101	1.42	140356.3
Ext5	S687R3	0.00129	1.9	147314.4
Ext5	S687R4		0	139568.2

Table 4: BCR-ABL copy number/rx targeted to mean value 1. Observed mean value = 1.4, observed median value = 1.0 (N=20). Yellow highlight indicates 1 of 2 post-extraction replicate reactions failed, in which case the reported value is for the other reaction of the pair. Red highlight indicates that both of the post-extraction replicate reactions failed, in which case the reported value is 0 (zero).

	$d = 0.5$				$d = 4.0$			
	a	b	c	e	a	b	c	e
$LOD = 1$	0.38	1.08	0.59	1.08	1.52	3.35	3.05	3.35
$LOD = 2$	1.87	0.82	0.59	3.03	1.18	2.99	3.05	5.55
$LOD = 3$	4.35	1.91	0.59	5.97	2.13	2.28	3.05	8.54
$LOD = 4$	7.79	4.28	0.59	9.91	4.96	2.65	3.05	12.42

Table 5: Cumulative absolute bias

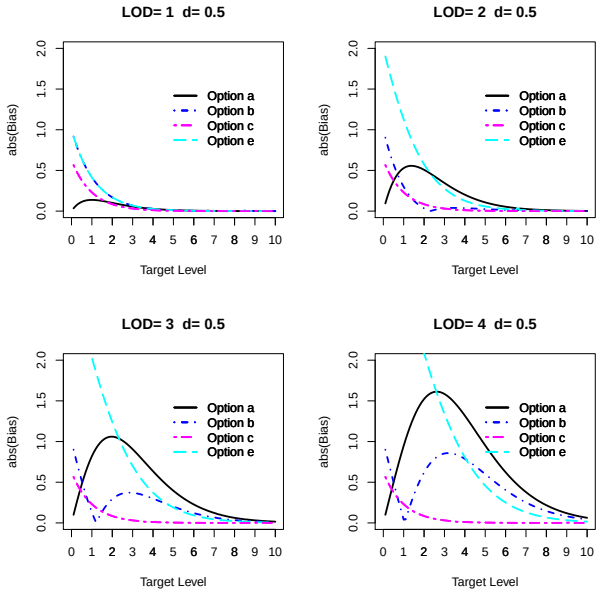


Figure 7: Bias curves when $d = 0.5$

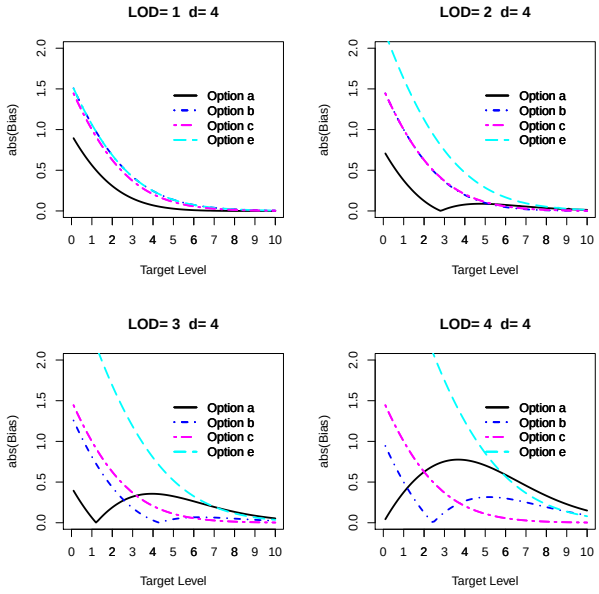


Figure 8: Bias curves when $d = 4.0$

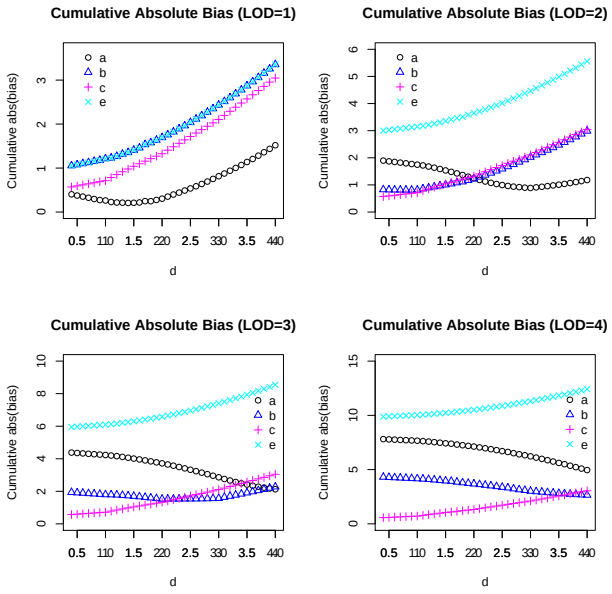


Figure 9: Cumulative absolute bias