Background and Support Materials for Peer Consultation Webinar Workshop on Model Averaging Methods for Dose-Response Analysis

6 November 2015

U. S. Environmental Protection Agency
Office of Research and Development
National Center for Environmental Assessment
Washington, D.C.

Disclaimer

This information is distributed solely for the purpose of pre-dissemination discussion under applicable information quality guidelines. It has not been formally disseminated by EPA. It does not represent and should not be construed to represent any Agency determination or policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

CONTENTS

1. INTRODUCTION

1.1. Purpose

The primary purpose of this document and associated software is to support discussions at the EPA model averaging workshop to be held December 10-11, 2015.[1] This support material is intended to assist in the evaluation of prevailing model averaging methods and options. The software package facilitates the analysis of continuous data, i.e., dose-response data that have responses measured (and reported) on a continuous scale (e.g., body weight or serum enzyme levels). The software and the test runs that have been completed using it are the first step in a process in which various model averaging techniques will be subject to peer consultation and comment, with an ultimate goal to identify model averaging approaches that are of greatest advantage in the context of dose-response analysis and health assessment.

While model averaging may be viewed as one of several approaches designed to address model uncertainty, it has been the primary focus of recent EPA research because it has been extensively vetted in the literature for this purpose. It also offers potential advantages over existing approaches that rely on selection of a single model, including the ability to take into account prior knowledge (e.g., biological and historical information) regarding models and parameters under consideration. While the other approaches (e.g., semi-parametric modeling) may turn out to be viable options in some circumstances, this document and the scope of the research/development process are limited to methods that employ averaging. No attempt is made at this time to compare the averaging methods described here to the alternative approaches.

The model averaging techniques described in this document have been amalgamated into a prototype software package that performs all of the methods in a single pass. At this stage, the benefit of having all methods computed simultaneously is that it allows for a comparison of the results, both with respect to run times and with respect to BMDL estimates. Thus, the software provides the means by which systematic and extensive testing can be performed by, and peer consultation can be obtained from, a variety of experts participating in the workshop.

In addition to describing the methods implemented in the associated software package, this document presents a set of test results of the methods applied to some real and some simulated datasets. Of primary interest here are run times (because some of the methods employ bootstrap-based calculations) and the benchmark dose (BMD) values estimated. These results are the start of

[1] For information on the planned December 10-11, 2015 model averaging peer consultation webinar workshop visit http://www2.epa.gov/bmds/model-averaging-webinar-workshop-announcement.

the process of evaluation of the proposed methods described below. Preliminary observations about the relative outputs of the model averaging methods are provided. Finally, additional steps and suggestions for enhancing the software's utility for testing the methods and assumptions are presented.

1.2. Background

Model development to describe available dose-response data and predict the sensitivity of specific species to specific toxic chemicals can be quite complex, largely because of the highly interdisciplinary nature of the underlying processes, broad variety of the molecular targets and modes of action (MOA), and a large number of diverse factors involved.

This complexity may (and often does) result in model uncertainty – situations where modeling outcome depends on the choice of a particular model and/or the assignment of values to its parameters. Different models may yield close but still different results, or small modifications in the data or the model may trigger a qualitative change in the modeling outcome. For example, when predicting low-dose response using high-dose data, the outcome may dramatically change depending on the choice of a particular model [1].

To address model uncertainty, concepts of model selection and model averaging were introduced in the 1970's [2]. Model selection methods, including those currently employed by EPA [3,4], provide means to identify the best model out of a given set of models, whereas model averaging methods employ weighted averaging of results from multiple models in hopes of improving the quality of an assessment and providing more accurate predictions [2].

The research areas of model selection and model averaging are broad and so a complete description of all available approaches is beyond this document's scope. Instead, the support material prepared for this workshop focuses on modern approaches to model averaging that are best suited for dose-response analysis for health assessments, specifically on Bayesian model averaging introduced in the 1990's. The latter has been proposed as an alternative to the single-model, selected benchmark dose (BMD) [5]. The literature on model averaging is large, and the approaches to it discussed in this document represent a synthesis (with variations) of many of the considerations presented in that literature. Therefore, rather than cite all potential sources for the methods presented here, we have provided a bibliography of citations (Appendix A) that form the background for the discussions to follow. There has been active discussion of the merits of model averaging with respect to improving inferences, and that discussion is reflected in the bibliography.

To broadly apply model averaging to health assessment tasks, one needs to identify a standardized approach. As a first step towards that goal, EPA initiated work on the research and development of model averaging methods in 2013. The materials developed and distributed in support of this workshop have undergone internal reviews by the Agency's Statistical Workgroup (SWG), which have inspired important enhancements to the methods, test procedures and prototype software such as the addition of exponential models to the suite of models being

averaged, enhancements to the bootstrap approach for derivation of BMD lower bound confidence limits (BMDLs) and modifications that allow users to apply non-equal prior weights to models being averaged (e.g., based on biological plausibility). The methods described here and implemented in the associated software package were developed for continuous response data, but may be applicable, with relatively minor modifications, to dichotomous response data (see discussion in Section 4.3).

The remainder of this document presents five methods (with submethods) that are proposed options for implementation in dose-response modeling contexts. Those methods have been implemented in a software package. The construction, testing, and proposed usage of the software package for evaluating the averaging approaches it implements are also described in the following sections.

2. METHODS

2.1. Model Averaging Methods

The methods investigated and described in this document have all been proposed in (or are simple extrapolations from) approaches that have been presented in the literature. While the majority of model averaging techniques are based on Bayesian statistics [6], and while a full Bayesian analysis may be possible in some instances, simpler approximate methods for averaging have been presented. The methods under consideration here fall within the set of those "simpler" approaches.

Those methods depend on computing model weights in order to define the weighted average to be applied. Kang et al. [7] defined weights based on the Akaike information criterion (AIC). Weights based on the Bayesian information criterion (BIC) have been examined and found to provide adequate results much faster (than the corresponding full Bayesian analysis), in that case for non-informative priors [6]. Use of the BIC has been compared to the use of other information criteria such as the AIC, its finite-sample corrected version (AICc), and the focused information criterion (FIC), or their modifications (see for example, [2], [6], and [8]]. The overall literature does not clearly indicate that any of these choices are categorically better than any other.

Nevertheless, the methods incorporated into the prototype software and evaluated here use model weighting (or model-estimate weighting) that is based on BIC (see [9] for a rationale for that choice based on Bayes factors). Additional approaches using other information criteria or other weighting methods altogether could be proposed and/or considered as part of the peer-consultation process. Other weighting schemes would be trivial to implement given the Model Averaging software version discussed here, though the implications may be different.

Model weights are defined as follows.

Let

$$m(i, j) = \exp(-BIC(i, j)/2)$$

for model i in bootstrap iteration j. BIC(i, j) is the BIC value being considered for that particular iteration, j, of any of the methods for model averaging described (2a, 2b, …, 5b, 5c). BIC(i,j) may differ from method to method (some use original-data BICs, others use iteration-specific BICs), even for the same model i.

Then the weights are given by

$$wt(i,j) = pw(i)*m(i,j) \,/\, \sum_{k=1}^{M} pw(k) * m(k, j) \,. \qquad \text{Eq. 1}$$

where pw(i) is the prior weight given to model i, with a total of M models.

In general, BIC is defined as

$$BIC(i, j) = -2L(i, j) + p_i \log(N)$$

where $L(i,j)$ is the log-likelihood maximized by maximum likelihood estimation (MLE) for the i^{th} model at the j^{th} bootstrap iteration, p_i is the total number of the parameters in the i^{th} model, and N is the experimental sample size.

Taking the current outputs available from dose-response models (as implemented in BMDS) and then considering the various approaches suggested by the synthesis of the literature mentioned above (particularly the publications by Wheeler and Bailer, [10] and [11], who specifically looked at averaging for dose-response models applied to dichotomous endpoints), five model-averaging methods were selected as a starting point[2] for implementation and then testing of their properties:

Method 1: A simple extension of the calculations normally done with dose-response modeling, e.g., as implemented in BMDS.

- Calculate lower statistical confidence limit of the benchmark dose (the BMDL) for each included model using MLE and profile likelihood methods;
- Calculate individual model weights using BIC (Eq. 1);
- Calculate the weighted averaged BMDL as a weighted sum of the individual BMDLs for each model.

[2] We realize that other methods could be defined, but those that have some history in risk assessment contexts have been the primary focus here.

Method 2: Moerbeek et al. [12] proposed using bootstrap methods for deriving BMDLs. Methods 2 – 5 are all based on versions of bootstrapping for BMDL calculation.

- Method 2 uses a semi-parametric bootstrap, sampling from normal distributions defined by the observed means and standard deviations in the dataset under consideration, to derive BMDL values.

- Three submethods (a-c) implemented for this method differ with respect to if and how the BMDL estimates so derived are used.

    a) Following the work of Wheeler and Bailer [10] an averaged BMDL was computed from the bootstrap-based, model-specific BMDLs with weights determined from Eq. 1 and BICs derived from the original dataset.

    b) In place of averaging the BMDLs, the weighted BMD for each bootstrap iteration is calculated (using Eq. 1). The 5^{th} percentile of the weighted BMDs over all iterations is selected as the BMDL.

        - Weights for averaging in each iteration are determined by the weights (BICs) calculated in that specific iteration.

    c) In place of averaging the BMDLs, the weighted BMD for each bootstrap iteration is calculated (using Eq. 1). The 5th percentile of the weighted BMDs over all iterations is selected as the BMDL.

        - Weights for averaging in each iteration are determined by the weights (BICs) calculated from the original dataset.

Method 3: The same as Method 2 except with respect to the basis for the bootstrap sampling.

- Method 3 uses parametric bootstrapping [6]. Each simulated dataset was generated from distributions defined by the means and variances predicted by one of the models fit to the original data. The model used was selected randomly in accordance with the BIC-based weights (Eq. 1) when fit to the original data. Note that those weights are fixed and constant once the original data have been fit by all the models. The means and variances for the normal response distributions associated with the dose groups were equal to the predicted means and variances from that model.

- The three submethods (a-c) implemented for this method differ with respect to if and how the BMDL estimates so derived are used.

    a) Following the work of Wheeler and Bailer [10] an averaged BMDL was computed from the bootstrap-based, model-specific BMDLs with weights determined from Eq. 1 with BICs derived from original dataset.

    b) In place of averaging the BMDLs, the weighted BMD for each bootstrap iteration is calculated (using Eq. 1). The 5th percentile of the weighted BMDs over all iterations is selected as the BMDL.

- Weights for averaging in each iteration are determined by the weights (BICs) calculated in that specific iteration.

c) In place of averaging the BMDLs, the weighted BMD for each bootstrap iteration is calculated (using Eq. 1). The 5th percentile of the weighted BMDs over all iterations is selected as the BMDL.

- Weights for averaging in each iteration are determined by the weights (BICs) calculated from the original dataset.

Method 4: Based on the "model averaging" concept [11], in which model predictions are averaged (for all doses) and those "average model" predictions are used to derive BMDs and BMDLs.

° Generate Bootstrap using the semi-parametric procedure (from normal distributions defined by the observed means and variances).

° Identify the dose such that the averaged response (using Eq. 1) at that dose is equal to the response that corresponds to the definition of the BMR, relative to the averaged response for dose equal to zero.[3] Do this for each bootstrap iteration. The 5th percentile of the identified doses, over the bootstrap iterations, is set to the BMDL.

° The two submethods b and c[4] implemented for this method differ with respect to the weighting used to average the responses in each iteration:

b) Weights for averaging in each iteration are determined by the weights (BICs) calculated in that specific iteration.

c) Weights for averaging in each iteration are determined by the weights (BICs) calculated from the original dataset.

Method 5: The same as Method 4 except for the basis for the bootstrap sampling.

° Generate bootstrap samples using the parametric procedure (using normal distributions defined by model-predicted means and variances; model chosen randomly, with model selection probability dictated by the original BIC-based model weights).

° Identify the dose such that the averaged response (using Eq. 1) at that dose is equal to the response that corresponds to the definition of the BMR, relative to the averaged response for dose equal to zero. Do this for each bootstrap iteration. The 5th percentile of the identified doses, over the bootstrap iterations, is set to the BMDL.

° The two submethods b and c implemented for this method differ with respect to the weighting used to average the responses in each iteration:

[3] For example, with the BMR being 10% change in response, the BMD is that dose at which the averaged response is 1.1 (or 0.9 for decreasing responses) times the averaged background response.
[4] There is no submethod a for Methods 4 and 5.

b) Weights for averaging in each iteration are determined by the weights (BICs) calculated in that specific iteration.

c) Weights for averaging in each iteration are determined by the weights (BICs) calculated from the original dataset.

Table 1 summarizes the approaches and differences among the five methods and their submethods. Diagrams 1a and 1b show the methods as flow charts delineating the steps in the processes. The methods will be referred to as Method 1, Methods 2a - 2c, Methods 3a – 3c, Methods 4b – 4c, and Methods 5b – 5c, as defined here and in Table 1.

Method 1 and the "a" submethods average BMDLs (or their bootstrap equivalents). The "b" and "c" submethods compute or average BMDs and then determine a percentile over all bootstrap iterations to be the BMDL estimate. These are two very different approaches. Both have been investigated in the literature on dichotomous dose-response models ([10, 11]).

No bias corrections or accelerating options (see [13]) for using bootstrap results have been employed here.

2.2. Implementation of Model Averaging Methods

2.2.1. Available Models and Specification of Model Modifications

The current version of the Model Averaging software implements the following six models:

- Linear
- Poly3 (Polynomial with degree 3)
- Power
- Hill
- Exponential 3 (Exp3)
- Exponential 5 (Exp5)

Equations for these models are given in the BMDS Help file (available at http://www2.epa.gov/bmds/benchmark-dose-software-bmds-user-manual). The Discussion Section offers suggestions for inclusion of additional models, but this set was chosen because they were considered to span the range of possible curve shapes available in BMDS.

For the purposes of the Model Averaging software, slight modifications were made to the currently available models in BMDS[5]:

[5] EPA plans to make future versions of BMDS consistent with these modifications, where appropriate.

- The Linear model is a limiting case of the Poly3 model. Hence, its likelihood should always be less than or equal to that from the Poly3 model. In any case where the likelihood for the Poly3 model was less than that for the Linear model (indicating a problem of fitting by the Poly3 model), Poly3 model results (log-likelihood, parameter estimates, model predictions) were set equal to the corresponding values from the Linear model.

- The Power model is a limiting case of the Hill model. Hence, its likelihood should always be less than or equal to that from the Hill model. In any case where the likelihood for the Hill model was less than that for the Power model (indicating a problem of fitting by the Hill model), Hill model results (log-likelihood, parameter estimates, model predictions) were set equal to the corresponding values from the Power model.

- The Exp3 model is a limiting case of the Exp5 model. Hence, its likelihood should always be less than or equal to that from the Exp5 model. In any case where the likelihood for the Exp5 model was less than that for the Exp3 model (indicating a problem of fitting by the Exp5 model), Exp5 model results (log-likelihood, parameter estimates, model predictions) were set equal to the corresponding values from the Exp3 model.

- For all models, if the BMD estimate was greater than 1000 times the maximum dose in the dataset under consideration, the BMD reported by the Model Averaging software is 9999 times that maximum dose. This provides an easily identifiable flag for when a model is extremely flat (including perfectly flat, with associated infinite BMD). That identifiability extends to the averaging: averaged values that include a BMD of 9999 times the maximum dose are still apparent after the averaging has been done. Moreover, the averaged values will be on the high end of the distribution of averaged BMDs, so percentiles of interest for defining the BMDL (e.g., the 5^{th} percentile for a 95% confidence interval) are insensitive to the exact value chosen to substitute for extremely large (perhaps infinite) BMD estimates.

- A BMD having the value of -9999 for a model run indicates failure to converge. The flag of -9999 is treated in the Model Averaging software as a number for the purposes of averaging the BMDs. Any negative values of averaged BMDs are ignored when percentiles across the bootstrap iterations are computed.

Other constraints currently imposed on the Model Averaging runs include the following (in some cases, options that could be added in a future version are listed):

- In the Model Averaging software distributed with this document, the user must specify the adverse direction rather than allowing the program to select it automatically. This is to ensure that, for the BMD computations performed for all bootstrap generated datasets, the BMD is consistently associated with the same magnitude of response in the same direction the user considers to be the adverse direction. Restrictions on model parameters (for the Linear, Poly3, Exp3, and Exp5 models in the current version) are linked to the specified adverse direction and cannot be changed. For example, if the adverse direction is up, then the dose coefficients of the Linear and Poly3 models, when restricted, are restricted to be non-negative.

- The BMR type must be relative deviation. Future versions could include other BMR options such as the absolute deviation, standard deviation, and point estimate. These are options offered in BMDS.

- The response data are assumed to be normally distributed. Future versions could also allow the assumption of a lognormal distribution for the underlying continuous responses.

- No parameter values for any of the models can be specified to be equal to a user-selected value. Future versions could allow users to inform model parameter estimations with prior information (e.g., regarding parameter ranges/boundaries).

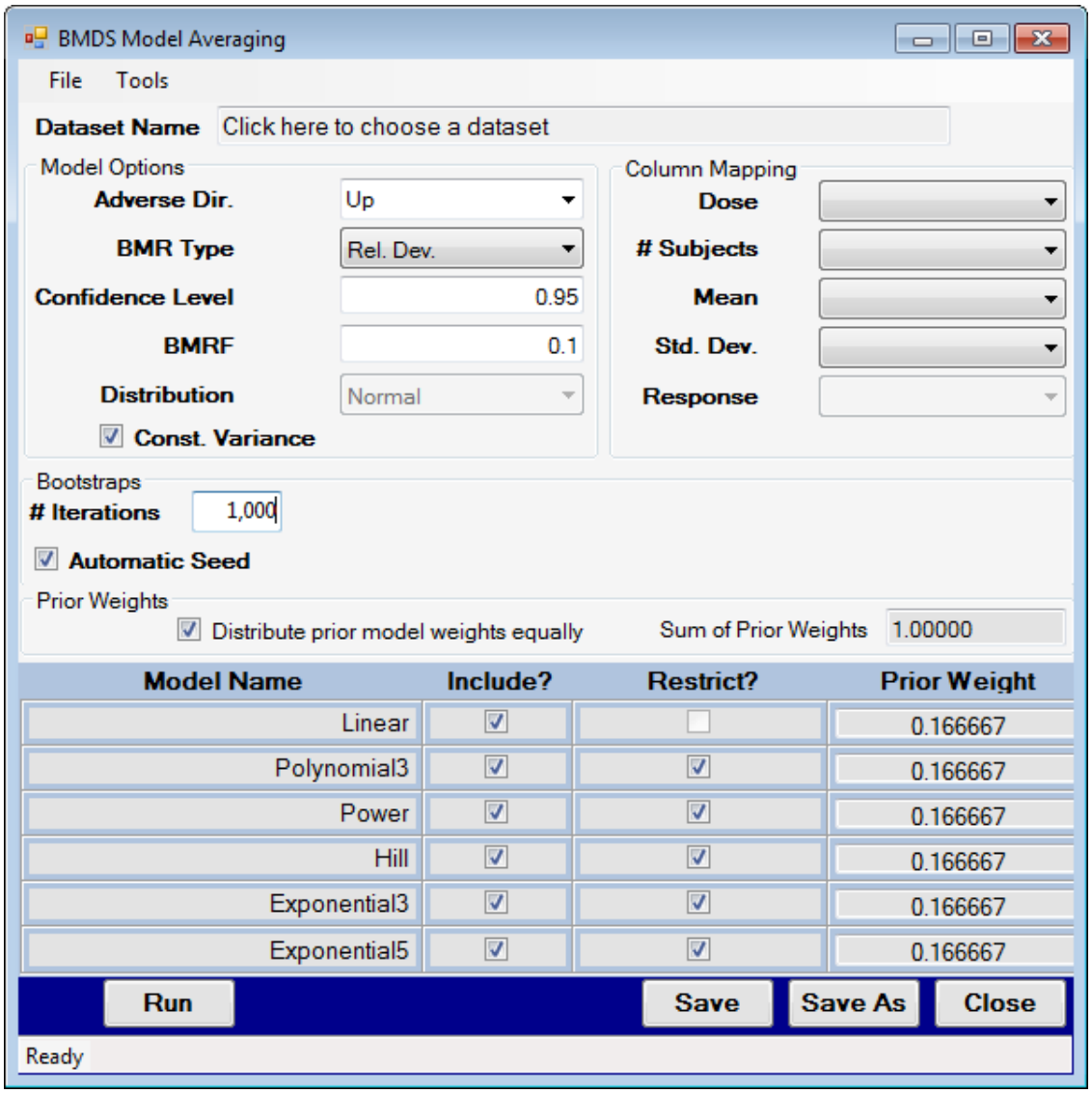
- For this version of the Model Averaging software, EPA used datasets with at least four dose groups. This is because important models that EPA wanted to test, such as the Poly3, Hill, and Exp5 models, have four parameters and will not run for datasets with fewer than four dose groups. In this version of the model averaging software, smaller datasets will result in null results because the models will quit when they determine that the number of parameters is greater than the number of observations. As discussed in Section 4.2, a later version could make adjustments for the number of dose groups in a dataset such as applying parameter constraints when the number of parameters exceeds the number of dose groups, or running a subset of the models on the dataset in question.

2.3. Running the Model Averaging Software

Model averaging software designed for purposes of the December 10-11, 2015 peer consultation workshop was distributed on November 10, 2015 as two zip files.[6] Software for performing model runs on a single dataset using a Windows graphical user interface (GUI) (Run Options 1) are contained in the file "ModelAvgGUI_20151106.zip." Software for performing model runs on a multiple datasets in a batch fashion (Run Options 2) are contained in the file "ModelAvgBatch_20151106.zip." The datasets used to perform the batch test runs described in this document are contained in a third zip file called "MA_data.zip."

Run Option 1: For single datasets, one can run a graphical user interface (GUI) which displays the various elements affecting the software's behavior. Instructions for using the GUI are contained the file "BMDS Model Averaging Quick Start.docx" distributed within the "ModelAvgGUI_20151106.zip" file. The following is a screenshot showing the default options of the model averaging GUI.

[6] Visit http://www2.epa.gov/bmds/model-averaging-webinar-workshop-announcement to download the software prior to the workshop. Contact Jeff Gift, Ph.D., NCEA, at gift.jeff@epa.gov or by phone at 919-541-4828 with questions concerning the use of the software.



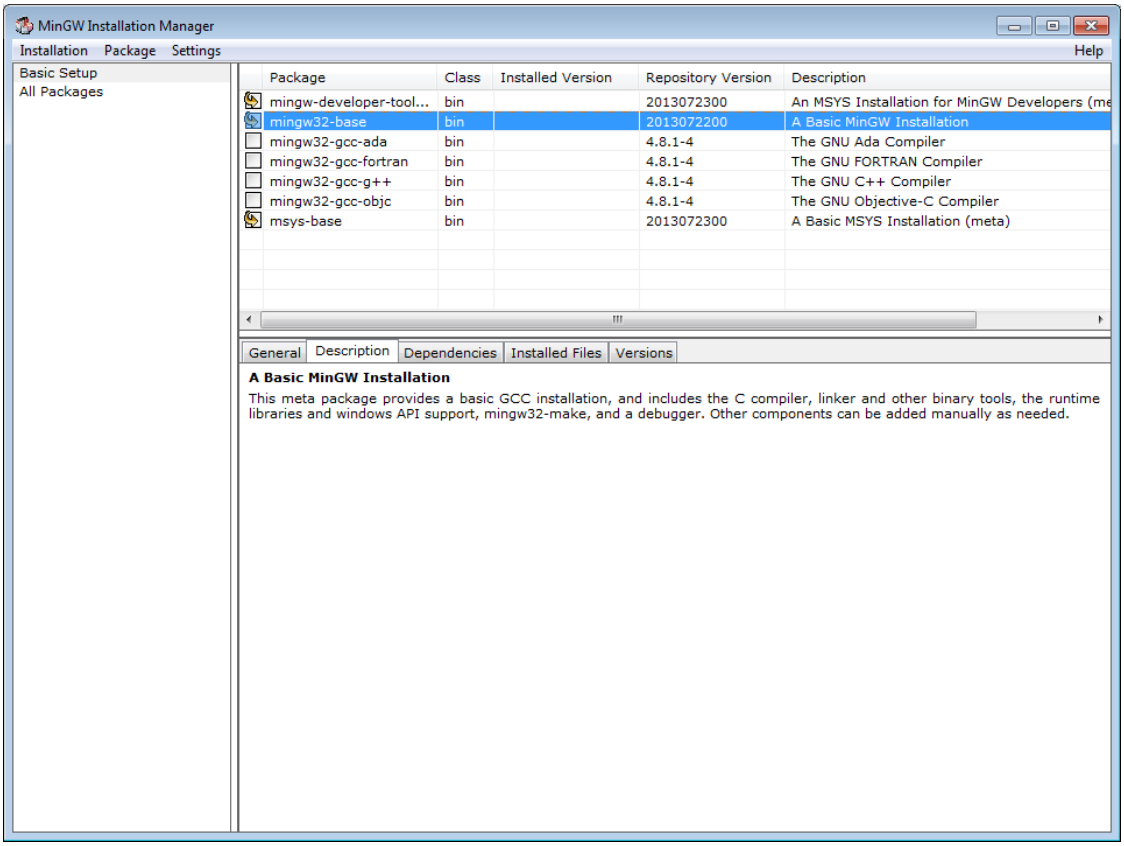
The GUI automatically creates and runs the avg text file that is required for input into the model averaging program. The format for an avg file is presented in Appendix B. These avg text files can also be created using any word processing program and run from the DOS command line by navigating within the DOS program of your computer to the folder that contains the model averaging executable (CModelAvg.exe) and the avg file of interest and entering the following command at the DOS prompt.

Prompt >> CModelAvg.exe *.avg

The asterisk in the above represents the base name of the selected avg file. The output file (having an extension of .log) will have that base name.

Run Option 2: The "matest2.sh" shell script distributed within the "ModelAvgBatch_20151106.zip" file, can be used, along with a csv dataset file, to create and run avg files automatically for multiple datasets. The "matest2.sh" program must be called from a MinGW command window. MinGW, which stands for "Minimalist GNU for Windows," provides an Open Source tool set that resembles Linux for building and running applications on Microsoft Windows.

To install MinGW you must first download the MinGW installer.[7] You will be asked to designate an installation directory with the default being "c:\MinGW." If you choose to change the install location, be sure to install into a directory for which you have administrative rights and no spaces in the directory path name. Select "…on the desktop" under the program shortcuts options. Once installed, double-click on the "MinGW installer" short-cut on your desktop and you should see the following screen.



In this window, select "Basic Setup" and choose "mingw-developer-toolkit," "mingw32-base" and "msys-base." Then click on "Installation," then "Apply Changes." Navigate within Windows Explorer to the installation location you designated, locate the "msys.bat" file within the "msys\1.0" folder and create a shortcut to it from your desktop. Double-click on the "msys.bat" short-cut. Using the "cd" command (type "help cd" for a definition of this command) to navigate in

[7] See the BMDS Source Code Download page (http://www2.epa.gov/bmds/download-benchmark-dose-software-bmds-source-code) for additional details and download links.

the MinGW32 window to the ModelAvgBatch folder that you extracted from the "ModelAvgBatch_20151106.zip" file and then type the following:

Matest2.sh <test set name> <adverse dir> <variance flag> <# iterations>

where:

- <test set name> = csv file name (**without the 'csv' extension**). This **MUST** match the test name in the top-left cell of the csv file (cell A1 when viewing the CSV file in Excel).
- <adverse dir> = -1 or 1 for down or up direction of adversity, respectively
- <variance flag> = 1 for constant variance, 0 for non-constant (modeled) variance
- <# iterations> = optional argument indicating number of bootstrap iterations to run for each dataset (defaults to 1,000; 100 is recommended for preliminary runs).

The matest2.sh file and the csv file to be run must be in the same folder (directory). Running matest2.sh in that folder will create a subfolder with the name of the csv file; that subfolder will contain the *.err, *.log, and *.stdout files (one of each type for each row in the csv file representing a dataset). A csv summary file (having the name of the data-containing csv file with "_Summary" appended to that base name) is also created. For example, running matest2.sh with the data-containing csv file named "real_data_up.csv" will create a subfolder named "real_data_up" (with .err, .log, and .stdout files) and the csv file "real_data_up_Summary.csv." The latter file will not be in the subfolder, but rather the 'parent' folder with the data-containing csv file (and matest2.sh).

An example of a csv file having the correct input format is shown in Appendix C. The csv files used for the batch test runs created for this document, including the "real_data_up.csv" file pictured in Appendix C are contained in the "MA_data.zip" file.

Run option 2 allows the user to run more than one dataset at a time. However, the following constraints are in place at this time:[8]

- The adverse direction for all the datasets in the csv file must be the same.
- The choice of variance model (constant or non-constant) must be the same for each model applied to each dataset.

[8] It has been observed that on some occasions matest2.sh will hang up on a dataset. Although this has apparently been fixed (has not reoccurred in testing after it was initially noted) the following can be used as work-arounds. In cases matest2.sh does not terminate because it hangs and fails to progress to the next dataset the user may:

- terminate matest2.sh and restart the run;
- terminate matest2.sh and then create a new csv file that has data lines only for the dataset on which matest2.sh hung and subsequent datasets;
- kill CmodeAvg.exe from the Windows Task Manager.

The latter option will abort the run on the dataset causing the problem and move processing on to the next dataset, albeit without getting results for the dataset that hung up.

- The confidence level, BMR type, and BMR factor must be the same for all models and datasets.

When matest2.sh is run, the .avg files necessary for CModelAvg.exe to run are created automatically from the csv file containing the data (one .avg file per line of data). Those .avg files will be in the newly created subfolder.

Matest2.sh and its associated awk files; a sample csv file; the CModelAvg executable and its associated subfolder ("SysData") and dll files; and a sample avg file are included in the zip file "ModelAvgBatch_20151106.zip."

2.4. Test Runs: Methods

Testing of the Model Averaging software was completed using two types of datasets: datasets of real data from toxicology experiments, and simulated datasets generated from known underlying dose-response relationships. These two types are henceforward referred to as the Real datasets and the Simulated datasets.

For all the runs included in this document, the following settings were always in place (unless specified elsewhere when results are presented):

- 95% confidence limit is used to define the BMDL
- BMR type is relative deviation
- BMRF is 0.10
- Response distribution is normal
- Seeding for random number generation is automatic
- All models were run "restricted:" power parameters were constrained to be greater than or equal to 1 (Power, Hill, Exp3, and Exp5 models) and dose coefficients were constrained to have the same sign, consistent with the chosen adverse direction (Linear and Poly3 models). The power parameters were also always constrained to be less than 18 (the same constraint imposed by BMDS).
- All 6 models were included in the averaging. The runs assumed equal model prior weights (1/6).

2.4.1. Real Datasets

The Real datasets come from a repository of data that have been retained and used for testing of BMDS models during model development. There are a total of 100 such datasets having continuous responses. The identities of the tested compounds and of the endpoints from which the data were obtained have been blinded.

As discussed previously, the test was restricted to datasets having four or more dose groups. With that restriction a total of 76 Real datasets were available for testing. Twenty-three of them had an adverse direction of "up;" 53 had an adverse direction of "down."

Electronic Attachment "MA_data.zip" contains the two csv files that list all the Real data, one for each adverse direction: real_data_up.csv, and real_data_down.csv.

2.4.2. **Simulated Datasets**

The disadvantage of a Real dataset is that the underlying, "true" BMD for the dose-response relationship generating the observations is unknown. The Simulated data provide a better means of judging the adequacy of the MA methods by comparing the estimated BMDLs to the known BMD.

The Simulated datasets were generated from one of 64 Templates. The Templates define the experimental design of the hypothetical toxicology study, the assumed response distribution, and the actual dose-response relationship (with known BMD) used to generate the data.

The 64 templates were defined by all combinations of the following features:

- Experimental design: 2 possible designs referred to as chronic or subchronic. Each chronic-design template had 4 log-spaced dose groups (0, 0.25, 0.5, and 1) and 50 animals per group. Each subchronic-design template had 5 log-spaced dose groups (0, 0.125, 0.25, 0.5, and 1) and 10 animals per group.

- Response distribution: 2 distributions, normal or lognormal. For the normal distribution, the standard deviation was constant at 14. For the lognormal distribution, the log-scale standard deviation was constant at 0.14.[9]

- Dose-response relationships: 16 dose-response relationships (4 for each of 4 model types) as defined in Table 2 and displayed in Figures 1 – 16. Two model types (Power and Hill) are models that are included in the set of dose-response models being averaged. The other two (Poly and Exponential 4) are not included among the fitted models, though the fitted model Exp5 contains Exp4 as a special case (power parameter = 1).[10]

The 64 csv files that contain the 1000 simulated datasets per Template are in the electronic Attachment "MA_data.zip."

[9] No units are attached to the standard deviations, just as there are no real units attached to the simulated data.
[10] The Poly templates are from a non-fitted model for two reasons: the Poly models used to generate the simulated data have degree 4 (only a 3rd degree polynomial is included in the set of fitted, averaged, models); and those 4th degree generating polynomials have some negative coefficients (the fitted polynomial model restricted parameters to be non-negative).

3. RESULTS

3.1. Timing

Timing of the various methods under consideration was evaluated with respect to the real datasets. For that purpose, EPA ran each dataset through the methods a total of five times; i.e., there are five replicates of the analysis per dataset. Each replicate invoked 10,000 bootstrap iterations.

The times associated with Method 1 are independent of bootstrap method or number of iterations. They do include time for calculating the profile likelihood BMDLs for each model, which is not included in the times computed for Methods 2 through 5. Methods 2 through 5 substitute the time for bootstrapping in lieu of the time for profile likelihood methods. Methods 2a and 3a, are most similar to Method 1 in that they all get a BMDL for each model separately and then average them to get the final result. Methods 2b-c, 3b-c, 4b-c, and 5b-c all involve weighted averaging for each iteration.

The Method 1 run times are summarized for the real data, fit with constant-variance and non-constant-variance models, in Tables 3a and 3b. When fit assuming constant variance, a large majority of the runs took well less than a second to run; the 95th percentile of all 76 real datasets was 0.14 seconds on average. Run times were notably longer when the models allowed for a non-constant variance; the 95th percentile for average run time was about 5 seconds.

Despite the fact that no bootstrap iterations affect Method 1 run times, there was some variability in observed run time for that method across runs. With constant variance, the CVs (within dataset, across the five replicate runs) ranged up to 0.17, with 95% being less than 0.112. The CVs for run times with a non-constant variance had a maximum of 0.14.

Given the relatively consistent Method 1 runs times within dataset, and an even smaller magnitude for variation in Methods 2 through 5 (maximum CV was less than 0.09 for all non-constant variance and constant variance runs), the following comparisons of run times across Methods are based on the average times (of 5 runs) for each Method (e.g., as summarized in the 'Mean' column of Tables 3a and 3b). Tables 4a and 4b summarize Method 2 through 5 run times relative to the corresponding Method 1 run time for each dataset; (Method 'x' run time for dataset y) / (Method 1 run time for dataset y).

The run times across the submethods (a-c) within Methods 2 through 5 were essentially identical within each dataset. In fact, the run times of corresponding submethods for Method 2 and Method 4 were essentially identical, as were the corresponding times for Method 3 and Method 5. Using Methods 2 and 3 to gauge the difference in times contributed by the bootstrapping method (semi-parametric for Methods 2 and 4; parametric for Methods 3 and 5) it appears that there is no consistent difference in run times (Table 5). Regardless of the variance model used in the fitting,

Methods 2 and 3 run times differed no more than by a factor of 2.22. Interestingly, for the constant variance modeling 45% of the datasets were fit faster with Method 3 than with Method 2 (59% for the non-constant variance modeling), despite the fact that Method 3 involves an extra step of selecting a model to generate the bootstrap samples (at each iteration). Presumably, the shorter run times for Method 3 in those cases are due to the faster fitting (maximum likelihood estimation) achieved because at least one of the models being fit was the one used to generate the simulated data).

3.2. BMDL Estimates

3.2.1. Real Data

Similar to the timing estimates presented above, the real datasets can inform us about the relative values obtained among the methods. Moreover, the adequacy of using 10,000 bootstrap iterations can be evaluated through examination of the CVs associated with the BMDL estimates, by dataset, across the 5 replicates of analyses that were done. 'Adequacy' in this case refers to the precision of the estimates; if 1000 bootstrap iterations are adequate, then there should be little variation (small CVs) across those replicates.

Precision, in terms of CVs, is summarized in Tables 6a and 6b.[11] The median CVs are similar across methods (ranging between <0.1% and 0.8% regardless of the variance modeling approach). However, there is a tendency for the 'b' submethods (those that re-weight the average each iteration) to be more sensitive to some aspects of data differences, in the sense that those methods can attain CVs that are notably larger than the other methods (see the 90th, 95th and maximum CV values for the 'b' submethods compared to the others). Methods 2c and 4c look particularly good in this comparison, and might be judged more efficient, in the sense that the variability in the BMDL estimates for 10,000 bootstrap iterations is notably less than that for the other methods. In the future, if running the 'b' submethods, one might be better served to run more than 10,000 bootstrap iterations. It is expected that CVs would decrease with increasing number of bootstrap iterations per run. A previous set of runs using only 1000 iterations had median BMDL CVs in the range of 1% to 2% (data not shown) as opposed to the maximum of 0.8% observed with 10,000 bootstrap iterations.

Tables 7a and 7b summarize the Method 2a-c, 3a-c, 4b-c, and 5b-c BMDL estimates relative to the Method 1 BMDL estimate, for each dataset. The ratios presented for relative BMDLs are based on the average BMDL (over 5 replicates) for each dataset divided by the Method 1 BMDL.

The median relative BMDL for every method (with the exceptions of Method 4b and, to a lesser extent, 5b) is very close to 1, indicating that "on average" the BMDL from Methods 2-5 did not differ greatly from that from Method 1. There are some highly divergent values, however. These are

[11] No results are shown for Method 1 because all BMDL estimates for that method are the same across replicates (it does not require any bootstrapping).

in some cases "artifacts" (or perhaps "indicators") of the fact that some of the datasets showed little or no dose-related response. In such cases, one or more of the models will yield a very large (or infinite) BMD (or, for Method 1, BMDL); such large (or infinite) estimates are set to 9999 times the maximum dose. Limiting attention to datasets for which there was a significant dose-response relationship would tend to eliminate such extreme values.

The true BMDs for these Real datasets are not known, thus, this approach cannot be used to evaluate the accuracy of the methods. That is not the case for the simulated data, for which the true BMD is known. Issues of accuracy and coverage are discussed in relation to those datasets in the next section.

3.2.2. **Simulated data**

For the Simulated data, the true BMD is known. Therefore, the performance of the various methods and submethods can be evaluated with respect to two metrics.

First, the medians and inter-quartile ranges for the BMD estimates are presented for the two main categories of methods, BMD-averaging methods (Methods 1, 2a-2c, and 3a-3c) and model-averaging methods (Methods 4b-4c, 5b-5c). It may be more typical to present expected values (means) and variances of the BMDs (corresponding to bias and precision metrics), but EPA hesitates to do so in this analysis because of the arbitrary decision to set "large" BMD estimates (greater than 1000 and possibly infinite) equal to 9999. Percentiles (including the median) will be predominantly unaffected by that decision; means and variances may be greatly affected.

Second, methods are presented and evaluated according to "coverage" of the BMDLs. The estimated BMDLs can be compared to the known BMD value. Because all the BMDLs calculated are intended to correspond to a 95% one-sided confidence limit, the ideal distribution of BMDLs for each method would have 95% of that distribution less than or equal to the true BMD (i.e., to have the "advertised" 95% coverage). Because each submethod evaluated here differs from the others in some respect (weights used, bootstrap approach, or whether it is a BMD-averaging or model-averaging method) plots displaying coverage (distributions of BMDLs across the 10,000 simulated datasets per template) show the performance of each submethod as a separate curve.

The median and inter-quartile range values discussed above are presented with the corresponding plot of the BMDL distributions for each template.

Case 1: Response Distribution Assumptions All Correct; Data Generating Model Included

The results of the runs on the Simulated data are organized by first examining behavior for those templates expected to have the best performance: those with a data-generating model included in the averaged models (Power and Hill, or "w" and "h," templates) having response data distributed normally around the median values. These templates have a constant variance, so models fit assuming constant variance are shown. Performance for those templates is illustrated in Figures 17 – 32. In all but one case (Figure 26, Template h1_normal_subchronic) the coverage for

each submethod is at or near the advertised level (95%) and in some cases somewhat exceeds the advertised level. These results do not strongly favor any method over another, since they all are fairly good (as expected). In the one exceptional case (Figure 26), the bias for the two main approaches was notable; the true BMD was just about at the 25th percentile of the distribution of BMDs estimated by the BMD-averaging methods, and only slightly better for the model-averaging methods. Interestingly, that degree of bias and insufficient coverage was not in evidence for the corresponding chronic design template (Figure 25).

Case 2: Response Distribution Assumptions All Correct; Data Generating Model Not Included

When the data-generating model is *not* in the set of models being averaged, the results are somewhat different (Figures 33 – 40). In fact, the performance of each averaging method is determined by the bias associated with its BMD estimates (here represented by the difference between the median BMD and the true BMD).

Thus, for the "p" templates (Figures 33 – 40), when the BMD is higher (around 0.5) the biases are negative and the coverage tends to be adequate, at the cost of having some extremely low BMDL estimates possible. Conversely, when the BMD is lower (around 0.14, less than the lowest chronic dose and close the lowest subchronic dose) the biases are uniformly positive, greatly so in some cases. This makes the coverage very poor.

Case 3: Response Distribution Assumptions All Correct; Data Generating Model Bounds an Included Model

For the "e" templates (Figures 41 – 48), all methods resulted in positive biases, regardless of the relative magnitude of the BMD. Therefore, BMDL coverages tended to be very poor (less than advertised). The Methods 4b, 2b, 5b, and 3b (roughly in that order) tended to do better with coverage despite the bias. These are methods that redefine the model weights at every bootstrap iteration. Methods 4b and 2b use a semi-parametric bootstrapping technique (data-driven); 3b and 5b use a parametric bootstrapping technique (model-driven).

The Exp4 model generated the data for these templates; it is a nested submodel of the Exp5 model (nested because the power in Exp4 is fixed at 1) that included in the set of averaged models. In fact, it bounds the set of Exp5 models since the latter are constrained to have power ≥ 1. Nevertheless, the coverage of the model averaging methods under consideration was still poor, due to biases in the estimates, as noted above.

With respect to including Exp5 (and Exp3) in the set of averaged models, note the following. Previous runs were performed without the Exp models. For those runs, the coverage for Method 4b (representing the best achieved by the methods under consideration) are tabulated here:

Template (corresponding Figure Number)	Coverage for e Templates	
	Without Exp models	With Exp models
e1_normal_chronic (41)	0.731	0.76
e1_normal_subchronic (42)	0.678	0.713
e2_normal_chronic (43)	0.632	0.665
e2_normal_subchronic (44)	0.806	0.823
e3_normal_chronic (45)	0.873	0.873
e3_normal_subchronic (46)	0.752	0.766
e4_normal_chronic (47)	0.663	0.702
e4_normal_subchronic (48)	0.792	0.827

The addition of the Exp models did improve coverage of averaging Method 4b (and the others) slightly. The Exp5 models itself did have excellent coverage, as, quite often, did the Hill model (see Figures 41 – 48). But enough weight was still given to the other models to "degrade" the performance of the averaging.

Case 4: All Response Distribution Assumptions Incorrect; Data Generating Model Included

In this case, the underlying response distribution is lognormal. Thus, the assumptions made during the course of model fitting are wrong on two counts. First, the response data are truly lognormally distributed, but the models assume normality. Second, the fitting is done assuming constant variance whereas the true variances differ across dose levels (as a consequence of assuming a constant log-scale variance for the data generation).

However, the models do include the data generating model in the sense that the dose-related median values are known to be described by either a Power or a Hill model.

Figures 49-64 display the behavior of the various methods under these conditions. It appears that the incorrect specification of the underlying response distribution and variance structure had little impact on the performance of the methods. Here, as in the Case 1 where everything was specified correctly (Figures 17 – 32), the biases are not great and the coverage is generally close to the desired 95%. If there is a slight difference between the sets of figures, it is that with the lognormal data the separation between BMDL distributions for Methods 2 and 4, on the one hand, and 3 and 5, on the other hand, is even more distinct, for templates in which the separation was not great in the normal-data case (e.g., compare Figures 17 and 49). Methods 2 and 4 (semi-parametric bootstrap) appear to be a bit less conservative than Methods 3 and 5 (parametric bootstrap).

Case 5: Response Distribution Assumption Incorrect, but Variance Model Correct; Data Generating Model Included

If the previous case is "corrected" slightly (Figures 65 – 80), where the variance of the underlying data could be modeled correctly (but the type of response distribution is still mis-specified), results look even more like the original case (where all assumptions were correct, Figures 17-32), in the sense that when results are good, all the methods perform similarly well (e.g., Figures 65 – 72). When the methods differ (as for most of the templates with a Hill data-generating model) the pattern observed for Case 4 is maintained; Methods 3 and 5 tend to be a bit more conservative than Methods 2 and 4. Submethod 'b' within Methods 2 and 4 appears to be more conservative than submethod 'c;' sometimes this leads to better coverage (e.g., Figures 70, 74, 76, 78, and 80 – i.e., for the subchronic designs, regardless of the magnitude of the BMD).

Case 6: Response Distribution Assumption Incorrect, but Variance Model Correct; Data Generating Model Not Included or Is a Bounding Model

Finally (Figures 81 – 96), one can consider the case analogous to Case 2, where the data-generating model is *not* included in the set of models being averaged or is a bounding case for one of the averaged models. In this case, however, the incorrect type of response distribution is assumed, although the variance structure could be estimated (non-constant variance is allowed). As was true of Case 2, the biases were often substantial and coverage was consequently adversely affected. Bias was worse when the BMD was low (near or below the lowest dose, even in the subchronic design). Overall, when bias and coverage tended to be particularly poor, Models 2b and 4b provided better coverage (though still coverage that was too low) than the other methods (Figures 85 – 86, 89 – 96).

4. DISCUSSION

The primary purpose of this document and associated software is to facilitate discussions at the EPA model averaging workshop to be held December 10-11, 2015. This workshop support package allows for the evaluation of prevailing model averaging methods and options. Within the constraints of the options that are included to date (see below), the software has been successfully run on a variety of datasets, both real and simulated, and the model average methods have been satisfactorily implemented.

It should be noted that all individual steps in the Model Averaging methods (fitting, weighting, and bootstrap simulation) have been examined separately and determined to be returning correct values. That is, EPA employed a separate and independent investigator to

implement each component (unit) of the process that yields the BMD and BMDL estimates for each method. The results of the independent unit tests matched those obtained from the developed Model Averaging software for the various test data sets.

At this point, the Model Averaging software is deemed ready for follow-up investigations by the workshop peer consultants. Additional considerations for these follow-up investigations are discussed in Section 4.4.

4.1. Overall Observations

The main conclusions of the testing completed to date are the following:

- By far the biggest impact on bias and coverage is due to whether or not the data-generating dose-response model is included in the set of models being averaged. For the cases investigated so far, this factor dominated any other consideration. From that perspective, one of the priorities for developing a model averaging procedure to apply in health assessments should be the inclusion of as many reasonable dose-response relationships as possible. While more computationally involved, the inclusion of additional model shapes should add no "interpretation burden." This is because the evidence so far suggests that there is no onus on selecting a model from the set of models that have been fit to a dataset; it appears the averaged value can be used without worrying about which model(s) did or did not fit the data well. In other words, it appears that if one is fairly certain that the true dose-response is among the models being averaged, then just using the averaged BMD and BMDL values should suffice.

  This conclusion may appear to be at odds with recent analyses that have reported on the adequacy of the Hill and/or Exponential models [14]. However, it should be noted that in that investigation, it was not simply the full Hill model (as used here) or the Exp5 model alone that were evaluated. Rather an entire family of "Hill" and Exponential models were considered, stepping up to the more complex versions only if needed (via a series of likelihood ratio tests). That investigation did not consider coverage probabilities or bias of the BMD and BMDL estimates for the selected Hill or Exponential models. Note also that there has been extensive discussion in the literature about the inability to correctly estimate confidence intervals for estimators once model selection (as was done by Slob and Setzer [14]) has occurred (see Leeb and Pötscher [15] for a relatively recent discussion). That has been one of the driving factors for development of model averaging approaches.

  Note also that having an array of possible models for averaging may enhance the role of biological/toxicological considerations in dose-response analyses. Suppose prior information is available that suggests that certain curve shapes (e.g., low-dose linear) are more biologically plausible than others. If a subset of the models included in the set of models to be averaged reflects the biologically based assumptions, greater *a priori* weight can be assigned to that subset of models. The software developed for this workshop has the capability to assign such prior weights and to therefore transparently reflect assumptions that are made in any dose-response modeling exercise.

- The comments in the immediately preceding bullet should be tempered to some degree, perhaps based on the results for the "e" templates. Those templates are generated from a

model that is nested within, and is a bounding case for,[12] one of the averaged models. Even with the more general model included in the averaging set, the coverage for all the averaging methods was not close to the advertised level of 0.95. This points up a need to consider bounding cases when determining which models to average. Aside from the addition of the Exp2 and Exp4 models, one might consider the addition of a Michaelis-Menten model, even though its more general form (the Hill model) is already included. Wheeler and Bailer [11] noted similar difficulties for model averaging of dichotomous response models when the true model is a bounding case.

Considered in relation to the first bullet item above, the proposed approach may be characterized as one that expands the model space, especially inclusion of edge cases (bounding models) rather than selecting from a smaller set of nested models and then making inferences from the selected model(s) as in [14].

That approach is consistent with another decision made in the current implementation. That is, for three models (Power, Hill and Exp5) there are corresponding nested models (Linear, Power, and Exp3) that impute their values (log-likelihood and BMD estimates) when the more complicated models "fail to perform." This is not a selection process, but rather reflects the known difficulty of fitting models (particularly those with a response asymptote) to certain data sets. Moreover, it is based on the known, logical relationship within each pair of models, i.e., that the log-likelihood for the more complex model must be at least as great as that for the simpler model. It is worth remembering that when the Exp5 or Hill model results (e.g., with respect to coverage or bias) are good, some of that may be attributable to the fact that those models can default to the simpler forms automatically (in the current software).

- Surprisingly, if the distribution of the underlying response data is misrepresented (as in the cases where a normal distribution is assumed but the response data are lognormally distributed), the effect on BMDL estimation was relatively small. This remained the case even if the variance was constrained to be constant even when the variance actually changed as a function of dose (through the change in the median). This may be largely due to the fact that the BMR for these test cases is based on relative change in the median response ($BMR_{10\%}$). Had the BMR been examined based on change relative to the standard deviation, or a "hybrid" definition for the BMD [4, Section 2.3.3.1], then this mis-specification may have been of greater importance.

- There were no large and over-riding differences among the methods that have been investigated. There was a suggestion that Methods 2b and 4b might perform better (though not all that well) than other methods should the data-generating model not be among those averaged. This was true even in preliminary runs that did not include the Exp models, suggesting that merely widening the set of models included in the averaging set does not appear to reduce or eliminate the performance margin for Methods 2b and 4b.

- A minor conclusion is that one probably ought to fit nonconstant variance models as a matter of course for model averaging. The effect of the variance assumption on the model-averaging results was minor. Therefore, at the cost of somewhat longer run times (but still with run times on the order of 10 seconds or less per 1000 bootstrap samples) one can

[12] By "bounding case" we mean a model that is nested within another, "larger" model and is obtained when a parameter value for the larger model is on the boundary of the parameter space allowed for that parameter.

cover all possible variance models (constant variance being a special case of the nonconstant variance models considered in BMDS). It may actually be desirable to run, together, both constant and nonconstant variance versions of each model included in the averaging set. If that was done, one would be averaging both over the possible median dose-response curve *and* over the variance model options.

- The biggest effect associated with experimental design (chronic vs. subchronic) was the greater spread in BMD estimates obtained from the subchronic design (compare the IQR values between corresponding plots that differ only with respect to experimental design). The total sample size for the subchronic design was 50 units (spread equally over 5 dose groups) whereas the chronic design included 200 units (50 in each of four dose groups). Peer-reviewers may be interested in other designs.

- Within the limitations of the designs considered here, there was a tendency for there to be greater (positive) bias in BMD estimates when the BMD was lower. That bias adversely affected coverage and was in the direction that leads to less health-protective estimates. Compare, for example the biases and coverage differences between Figures 26 and 28; between 34 and 36, between 37 and 39, and between 42 and 44. In all these cases the positive bias was greater for the low-BMD case than for the corresponding high-BMD case.[13] The low-BMD templates considered here had BMDs that were about half the lowest positive dose for the chronic design and close to the lowest positive dose in the subchronic design. Cautions about extrapolating far below tested doses and their responses appear to be applicable also to model averaged results. The degree to which averaging assists in that respect needs further investigation. The same applies to the individual models, except when the underlying dose-response pattern was one of the averaged models, in which case that model showed better coverage for the low-BMD cases. Sometimes this also resulted in improved model-averaging coverage.

4.2. Extensions to Continuous Data Model Averaging Software/Methods

There were no issues associated with implementation of the computational aspects of the software. However, there are several items that may be addressed as the software is developed in the next round. These include the following:

- Inclusion of additional models: Workshop peer consultants may be able to offer recommendations on whether other dose-response patterns should be included. The results presented here suggest that the Exp4 and Exp2 models, as well as a Michaelis-Menten model, could be added because they are bounding cases for models already considered.

- Additional BMR values (e.g., 1% and 5% relative risk) and types (e.g., absolute deviation, standard deviation, and point estimate) could be investigated. Such an investigation could help clarify the extent to which adding capabilities to fit non-normal models will add to the accuracy and coverage of the model-averaging predictions. Of particular importance would

[13] Note, however, that these "corresponding figures" were obtained with different underlying dose-response relationships; that is why the BMD values are different. In these cases, however, the same dose-response function is common to each pair (e.g., a low-BMD power model is compared to a high-BMD power model) and the experimental design is the same within the compared pairs.

be BMRs defined in terms of standard deviation changes or via a hybrid (risk-like) designation.

- Model runs which return a large BMD estimate (greater than 1000 times the highest dose) have been flagged by setting the BMD estimate to 9999 times the highest dose. While this is partial "protection" against infinite values or nonconvergence, a user-friendly addition might be added that provides an explicit warning when this has occurred. Safeguards or warnings against the use of unsuitable data should also be incorporated.

- The Windows-based graphical user interface (GUI) might be re-envisioned so that both constant and nonconstant variance models can be included in the set of models to be averaged.

- Addition of computational capabilities to allow the assumption of a lognormal distribution for the response observations. Currently all models assume the response data are normally distributed. As noted above, the priority for this may depend on further investigation related to different BMR types.

- It has been suggested that it might be useful to have an averaging method that would be applied to obtain values of 1/BMDL rather than BMDL itself. This might be useful in a context of deriving a cancer slope factor, for example, where one of the terms in that calculation could be viewed as 1/BMDL (the slope factor being BMR *(1/BMDL). Modifications to the existing code should be relatively straight-forward, requiring only the computation of the model-specific 1/BMDL values and then application of any (or all) of the averaging methods.

- Similarly, a possible extension would be to apply these methods to estimates of risk at a specified dose. Currently, we are specifying the risk of interest (the BMR) and determining the doses corresponding to that risk. Particularly in cancer assessments, one sometimes wants to estimate the "risk at a dose." For Methods 2 and 4, the observed data are used to generate the bootstrap samples; for Methods 3 and 5, a randomly chosen model (based on the model weights) is chosen as the basis for generating a bootstrap sample for each iteration. Given the patterns seen in the test runs to date, and the similarity of the Method 2/4 and Method 3/5 results (e.g., in terms of bias and coverage for the simulated datasets) it may be desirable to combine them to define a "unified" approach to bootstrap sample generation. That would be done by treating the saturated model (basically the observed means and variances, i.e., the basis for Methods 2 and 4) as another model that gets a weight (in relation to the fitted models) for model selection when generation of a bootstrap sample is required, as in Methods 3 and 5. This may prove particularly beneficial when none of the models fits particularly well. In that case, the random selection would favor the observed data, so the bootstrapping would predominantly look like Methods 2 and 4. On the other hand, if the models fit well, because they have fewer parameters than the saturated model their BICs would be less than that of the saturated model, and they would tend to be selected for bootstrapping. In that case, the method would be more like Methods 3 and 5. It appears that the proposed method could reap the benefits of both Methods 2/4 and 3/5 and perhaps even out the differences in coverage, minimal as they appear to be. This suggestion for alternative bootstrapping assumptions might be considered during the peer consultation workshop.

- On a related note, one might consider fitting models that relax the assumptions on the variance. Currently, one has to assume either constant variance or a variance model that is parametrically constrained by the estimates of the means. Neither needs to be the case. The alternative would be to fit models with a saturated variance model, allowing each variance (across dose groups) to be estimated independently of the other variances. This is the model that allows greatest flexibility for fitting models of the mean (or median) dose-response relationship. It is, however, not one of the options currently available in the BMDS models and would require modifications to those models for inclusion in the Model Averaging software.

- Some recent work has focused on situations where the true model was "on the edge of" or outside the range of averaged models [10, 11]. These are cases that were problematic for the averaging methods investigated to date and reported above. See, especially, the results for the "e" templates. Wheeler and Bailer [16] have proposed semi-parametric approaches to address these concerns; it may be desirable to investigate such approaches as part of follow-on analyses.

- For templates where a more complex model (Hill or Exp4) was the data-generating model (the "h" and "e" templates) the averaging appeared to be adversely affected by the BMDs estimated by the simpler models (e.g., linear or power models). It appears that too much weight is being given to those simpler models. One might consider using, as the basis for defining model weights, an alternative to the BIC. The AIC is a natural candidate since it would have just the desired effect. Its penalty for additional parameters is less than that for the BIC; it has been noted that model selection based on AIC tends to accept more highly parameterized models that does selection based on BIC. Investigation of such alternatives to the BIC-based weights may be desirable.

- Decisions need to be made about what to do in relation to data sets with 3 dose groups. One option is to run only those models with 3 or fewer parameters. Another option would be to define parameters constraints (equality constraints) for one or more of the parameters in more-highly parameterized models. Choices for which parameters to constrain need to be evaluated.

- In the runs performed to date, parameters subject to constraints (e.g., power parameters) have been constrained. It is not apparent if, or how, imposition of those constraints has affected the results shown above. It is possible that removing some of the constraints might affect performance for the "edge case" templates discussed above. This may be something that is explored further by the workshop consultants recruited to evaluate the proposed model averaging methods and options.

4.3. Applicability to Dichotomous Data

The analyses described here are for continuous endpoints and continuous dose-response models. A logical extension would be to apply the lessons learned here to, or to do additional testing on, dichotomous endpoints. Model averaging has been investigated extensively for dichotomous models by Drs. Wheeler and Bailer. Any extension of model averaging to dichotomous dose-response should borrow heavily from their investigations and, to the extent possible, their software.

Nevertheless, one may need to consider issues like the relationships between the models (simpler models nested within more complex models that are known to have certain log-likelihood relationships) as has been done for the continuous model software. Extensions of the software that EPA has developed for continuous endpoints may need to reflect any averaging issues specific to dichotomous responses, if any. The input of the workshop consultants and beta-testers will be helpful in identifying potential difficulties associated with a transition to dichotomous endpoints.

4.4. Further Examination and Selection of Model Averaging Methods

This round of testing suggests that, once the possible additions to the software have been agreed upon and successfully implemented, it will be ready to use as the basis for a more complete and thorough examination of the Model Averaging methods described above. It is anticipated that that process will involve a number of expert workshop consultants (and beta-testers) who can use the software to apply the methods in an automated manner. That automation will allow a much wider set of test datasets to be examined. Those datasets can differ with respect to experimental design as well as dose-response. It would be ideal if the datasets varied as widely as possible to get a complete picture of the performance of the Model Averaging methods.

Another aspect of model averaging that has not been exercised in the test runs to date concerns the prior model weights. All runs to date have been completed assuming equal model prior weights. Prior weighting schemes that reflected beliefs about biological plausibility might be considered. The translation of prior knowledge about underlying biology and toxicology into model weights is an important and potentially highly influential aspect that can be explored by the workshop consultants. From an implementation standpoint (i.e., regardless of biological considerations) the workshop consultants might want to explore the impact of alternative prior weightings on the BMD estimates. A particularly attractive option would be to run the datasets generated from the templates used in this report, ones with known underlying dose-response relationships and BMD values. While initial examinations of the effect of prior weighting for such datasets can be done using the GUI (i.e., for an individual dataset or two) a fuller exploration of differences in BMD estimates as a function of prior weighting could be obtained via a batch run over the sets of 1000 realizations for selected templates. Procedures for batch running the model averaging software are described in Section 2.4. The results presented in this document (i.e., all the Figures from 17 onward) were generated via such batch running. The results of altering prior weights could be summarized in the same manner and compared to the figures included here. [14]

[14] When comparing results for different prior weights, consider using the same number of bootstrap iterations and a fixed random number seed value. A different number of iterations and random seed values, with input settings otherwise identical, can produce differing results.

Some potential model averaging methods and areas of exploration have been discussed in this workshop support document. There are likely to be other alternative and complementary options worth exploring. EPA anticipates that the options and approaches discussed here will spark additional suggestions from expert analysts at the planned December 10-11, 2015 peer consultation workshop.

5. REFERENCES

1) H. Moon, H.-J. Kim, J. J. Chen, and R. L. Kodell. Model Averaging Using the Kullback Information Criterion in Estimating Effective Doses for Microbial Infection and Illness. *Risk Analysis* **25**, 1147-1159 (2005).

2) G. Claeskens and N. L. Hjort. *Model Selection and Model Averaging*. Cambridge University Press, 2008.

3) J.A. Davis, J.S. Gift, Q.J. Zhao. Introduction to benchmark dose methods and U.S. EPA's Benchmark Dose Software (BMDS) version 2.1. *Toxicol Applied Pharmacol*. 254: 181-191 (2010).

4) U.S. EPA. Benchmark dose technical guidance document. Risk Assessment Forum, Washington, DC; EPA/630/R-00/001F. (available at http://www2.epa.gov/risk/benchmark-dose-technical-guidance). (2012).

5) K. Shao and M. J. Small. Potential Uncertainty Reduction in Model-Averaged Benchmark Dose Estimates Informed by an Additional Dose Study. *Risk Analysis* **31**, 1561-1575 (2011).

6) K. P. Burnham and D. R. Anderson. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociol. Methods Res.* **33**, 261-304 (2004).

7) S-H. Kang, R.L. Kodell, and J.J. Chen. Incorporating model uncertainties along with data uncertainties in microbial risk assessment. *Reg Tox Pharm.* 32, 68-72 (2000).

8) H. de-G. Acquah. Comparison of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in Selection of an Asymmetric Price Relationship. *J. Develop. Agricult. Econom.* **2**, 1-6 (2010).

9) R.E. Kass and A.E. Rafftery. Bayes Factors. *J. Am Stat Assoc*. **90**, 773-795 (1995).

10) M. W. Wheeler and A. J. Bailer. Comparing Model Averaging With Other Model Selection Strategies for Benchmark Dose Estimation. *Environ. Ecol. Stat.* **16**, 37-51 (2009).

11) M. W. Wheeler and A. J. Bailer. Properties of Model-Averaged BMDLs: A Study of Model Averaging in Dichotomous Response Risk Estimation. *Risk Analysis* **27**, 659–670 (2007).

12) M. Moerbeek, A. H. Piersma, and W. Slob. A Comparison of Three Methods for Calculating Confidence Intervals for the Benchmark Dose. *Risk Analysis* **24**, 31-40 (2004).

13) B Efron. Better Bootstrap Confidence Intervals. *Journal of the American statistical Association*, **82**(397), 171-185 (1987).

14) W. Slob & R. W. Setzer. Shape and steepness of toxicological dose-response relationships of continuous endpoints. *Crit Rev in Toxicol*. **44**(3), 270-297 (2014).

15) M. Wheeler and A. J. Bailer. Monotonic Bayesian Semiparametric Benchmark Dose Analysis. *Risk Analysis* **32**, 1207-1218 (2012).

6. TABLES

Table 1: Definition of Model Averaging Approaches

Weighting		Bootstrap Technique		
		None (use profile likelihood methods)	**Semi-parametric (use original means and variances)**	**Parametric (use model-predicted means and variances)**
From original model fits	Average model-specific BMDLs	Method 1	Method 2a	Method 3a
	Average iteration-specific BMDs[1]	--	Method 2c	Method 3c
	Compute iteration-specific dose where average response = BMR[2]	--	Method 4c	Method 5c
From iteration-specific model weights	Average iteration-specific BMDs[1]	--	Method 2b	Method 3b
	Compute iteration-specific dose where average response = BMR[2]	--	Method 4b	Method 5b

[1] BMDLs are derived from percentiles of resulting iteration-specific averaged BMDs.
[2] BMDLs are derived from percentiles of resulting iteration-specific doses for which average response = BMR.

Table 2: Models Used to Generate Simulated Data for the Templates and Associated BMD Values

D-R Model Type	Template Label	Parameter Values					BMD
		a	b	c	g	e	
Polynomial	p1	80	66	-55	51	-11	.1345
	p2	80	26	-55	88	-28	.4775
	p3	100	-70	40	-50	35	.1541
	p4	100	-20	20	-40	5	.5112
Exponential M4	e1	80	4.5	1.2			.1540
	e2	80	2.1	1.17			.4225
	e3	120	3.05	0.75			.1675
	e4	120	1.68	0.8			.4126
Power	w1	80	167		1.9		.2021
	w2	80	111		3.1		.4281
	w3	135	-100		1.4		.2392
	w4	120	-86		2.9		.5071
Hill	h1	80	30	0.2	3.1		.1443
	h2	80	40	0.65	4.5		.4777
	h3	120	-40	0.2	5		.1688
	h4	120	-40	0.55	4.5		.4556

Model equations: m(d) is the dose-dependent median of the distribution of responses.
Polynomial: $m(d) = a + b*d + c*d^2 + g*d^3 + e*d^4$
Exponential M4: $m(d) = a *(c + (1-c)*\exp(-b*d)))$
Power: $m(d) = a + b*d^g$
Hill: $m(d) = a + b*d^g/(c^g + d^g)$
When the response distribution was assumed to be normal, then mean = median and std = 14
When the response distribution was assumed to be lognormal, then log-scale std = 0.14

Table 3a: Summary Distribution of Method 1 Run Times (milliseconds); Across 76 Real Datasets; Models Fit Assuming Constant Variance

Percentiles for Run Time (ms)	Run Number[1]						
	1	2	3	4	5	Mean	CV
minimum	7	6	6	6	6	6.2	0.000
5th	7	7	7	8	7	7.54	0.000
10th	7	8	8	8	8	7.8	0.000
25th	10	10	10	10	10	9.8	0.025
50th	13	15	14.5	14.5	14	14.5	0.051
75th	22	24.75	24	24	24	23.95	0.071
90th	63.6	75.5	76.1	75.2	75.5	72.76	0.091
95th	138.55	144.2	144.35	143.35	143.5	142.79	0.112
maximum	5020	6058	6048	6053	6071	5850	0.167

[1]Runs 1-5 are exactly the same (except for the bootstrap sampled values). Percentiles for 'Mean' and 'CV' values are for the within-dataset means and CVs (i.e., means and CVs across replicates).

Table 3b: Summary Distribution of Method 1 Run Times (milliseconds); Across 76 Real Datasets; Models Fit Assuming Non-Constant Variance

Percentiles for Run Time (ms)	Run Number1						
	1	2	3	4	5	Mean	CV
minimum	9	9	9	9	10	9.6	0.000
5th	11	11	11	11	11.85	11	0.000
10th	13	12	12	12.7	12	12.34	0.000
25th	17	17	17	17	17	17	0.000
50th	30	30	30	30	30	30	0.003
75th	75.5	77	74.5	74.5	74.5	75.2	0.012
90th	2674.9	2670.8	2675.5	2678.6	2677.3	2675.42	0.026
95th	5004.95	4967.9	5059.85	5013.05	4977.5	5004.65	0.040
maximum	10016	9944	9934	9948	9990	9966.4	0.140

[1]Runs 1-5 are exactly the same (except for the bootstrap sampled values). Percentiles for 'Mean' and 'CV' values are for the within-dataset means and CVs.

Table 4a: Summary Distribution of Relative Run Times (Method x Divided by Method 1, unitless); Across 76 Real Datasets; Models Fit Assuming Constant Variance

Percentiles for Relative Run Time	Method Run Time Divided by Corresponding Method 1 Run Time									
	2a	**2b**	**2c**	**3a**	**3b**	**3c**	**4b**	**4c**	**5b**	**5c**
minimum	25.5	25.5	25.5	17.9	17.9	17.9	25.6	25.6	18.0	18.0
5th	1190	1190	1190	943	943	943	1193	1192	946	945
10th	3000	2999	2999	3187	3187	3187	3011	3009	3192	3191
25th	4991	4991	4991	4964	4964	4963	5000	4999	4982	4979
50th	7447	7447	7446	8202	8202	8201	7472	7465	8229	8221
75th	10378	10378	10376	12756	12756	12755	10388	10385	12800	12780
90th	12328	12328	12326	17319	17319	17317	12366	12351	17371	17362
95th	14334	14334	14332	20249	20250	20248	14366	14360	20269	20265
maximum	34246	34246	34244	23233	23233	23232	34259	34257	23247	23243

Percentiles are of average run time for the Method in question (five runs per dataset) divided by the average run time for Method 1, for 10,000 iterations

Table 4b: Summary Distribution of Relative Run Times (Method x Divided by Method 1, unitless); Across 76 Real Datasets; Models Fit Assuming Non-Constant Variance

Percentiles for Relative Run Time	Method Run Time Divided by Corresponding Method 1 Run Time									
	2a	**2b**	**2c**	**3a**	**3b**	**3c**	**4b**	**4c**	**5b**	**5c**
minimum	11.7	11.7	11.7	11.3	11.3	11.3	11.7	11.7	11.3	11.3
5th	32	32	32	34	34	34	32	32	34	34
10th	108	108	108	79	79	79	108	108	79	79
25th	3895	3895	3895	3628	3628	3627	3900	3899	3632	3631
50th	8111	8111	8110	8112	8112	8112	8122	8120	8120	8119
75th	11837	11837	11837	11845	11845	11845	11846	11844	11860	11857
90th	15624	15624	15623	15742	15742	15741	15650	15643	15796	15785
95th	17850	17850	17849	18911	18911	18910	17859	17857	18922	18920
maximum	19903	19903	19903	22281	22281	22281	19909	19908	22296	22291

Percentiles are of average run time for the Method in question (five runs per dataset) divided by the average run time for Method 1, for 10,000 iterations

Table 5: Summary Distribution of Relative Run Times (Method 3x Divided by Method 2x, unitless); Across Real Datasets

	Method 3 Submethod Time Divided by Corresponding Method 2 Submethod Time					
Percentiles for Relative Run Time	**Assuming Constant Variance**			**Assuming Non-Constant Variance**		
	3a/2a	**3b/2b**	**3c/2c**	**3a/2a**	**3b/2b**	**3c/2c**
minimum	0.517	0.517	0.517	0.511	0.511	0.511
5th	0.630	0.630	0.630	0.721	0.721	0.721
10th	0.771	0.771	0.771	0.757	0.757	0.757
25th	0.924	0.924	0.924	0.883	0.883	0.883
50th	1.041	1.041	1.041	0.965	0.965	0.965
75th	1.351	1.351	1.351	1.103	1.103	1.103
90th	1.611	1.611	1.611	1.328	1.328	1.328
95th	1.776	1.776	1.776	1.422	1.422	1.422
maximum	2.219	2.219	2.220	2.066	2.067	2.067

Percentiles are of average run time for the Method 3 submethod (five runs per dataset) divided by the average run time for the corresponding Method 2 submethod.

Table 6a: Summary Distribution of Coefficient of Variation (CV) Values of BMDLs (unitless); Across 5 Replicates of the Real Datasets; Fit with Constant Variance Models

Percentiles for CV Values	**Averaging Method**									
	2a	**2b**	**2c**	**3a**	**3b**	**3c**	**4b**	**4c**	**5b**	**5c**
minimum	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
5th	0.000	0.002	0.000	0.001	0.002	0.001	0.001	0.000	0.002	0.001
10th	0.001	0.003	0.001	0.001	0.002	0.001	0.001	0.001	0.003	0.001
25th	0.002	0.004	0.002	0.002	0.004	0.003	0.003	0.002	0.004	0.003
50th	0.003	0.008	0.004	0.004	0.007	0.005	0.008	0.004	0.007	0.005
75th	0.006	0.024	0.006	0.006	0.016	0.008	0.014	0.007	0.017	0.008
90th	0.011	0.053	0.010	0.012	0.043	0.036	0.034	0.012	0.047	0.017
95th	0.029	0.082	0.012	0.559	0.665	0.562	0.069	0.021	0.070	0.034
maximum	0.559	0.559	0.031	1.369	1.495	0.921	0.303	0.032	1.267	2.236

Table 6b: Summary Distribution of Coefficient of Variation (CV) Values of BMDLs (unitless); Across 5 Replicates of the Real Datasets; Fit with Non-Constant Variance Models

Percentiles for CV Values	Averaging Method									
	2a	2b	2c	3a	3b	3c	4b	4c	5b	5c
minimum	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000
5th	0.000	0.002	0.001	0.000	0.002	0.001	0.001	0.001	0.002	0.001
10th	0.001	0.002	0.001	0.001	0.003	0.001	0.002	0.001	0.002	0.001
25th	0.002	0.004	0.002	0.002	0.004	0.002	0.003	0.002	0.003	0.002
50th	0.003	0.007	0.004	0.004	0.007	0.004	0.007	0.004	0.007	0.004
75th	0.006	0.024	0.007	0.005	0.012	0.007	0.017	0.007	0.015	0.007
90th	0.012	0.049	0.013	0.008	0.037	0.011	0.033	0.018	0.036	0.021
95th	0.015	0.072	0.016	0.018	0.057	0.020	0.073	0.020	0.052	0.043
maximum	0.559	0.561	0.038	0.913	0.191	0.559	1.970	0.460	0.672	0.073

Table 7a: Summary Distribution of Relative BMDLs (Method x Divided by Method 1, unitless); Across Real Datasets; Models Fit Assuming Constant Variance Models

Percentiles for Relative BMDL Values	Averaging Method									
	2a	2b	2c	3a	3b	3c	4b	4c	5b	5c
minimum	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5th	0.0004	0.0001	0.0004	0.0003	0.0003	0.0003	0.0000	0.0001	0.0001	0.0000
10th	0.0022	0.0020	0.0023	0.0023	0.0023	0.0024	0.0009	0.0011	0.0008	0.0006
25th	0.9419	0.3107	1.0091	0.7397	0.4358	0.9225	0.0501	0.5298	0.2213	0.2636
50th	1.0084	0.8998	1.0515	0.9991	0.9735	1.0331	0.7185	1.0124	0.8768	0.9779
75th	1.0856	1.0395	1.1612	1.0323	1.0152	1.1068	1.0081	1.0549	0.9954	1.0363
90th	1.2420	1.1989	1.3907	1.0503	1.0580	1.2282	1.0897	1.1154	1.0198	1.0719
95th	1.8525	1.5165	1.9888	1.1015	1.3160	1.4403	1.3282	1.3410	1.0533	1.1451
maximum	20.9798	39.7791	42.5290	3.6645	7.6062	8.2557	23.6203	15.8380	7.8012	7.3527

Values for relative BMDL are (BMDL for Method in question) / (BMDL for Method 1).

Table 7b: Summary Distribution of Relative BMDLs (Method x Divided by Method 1, unitless); Across Real Datasets; Models Fit Assuming Non-Constant Variance Models

Percentiles for Relative BMDL Values	Averaging Method									
	2a	**2b**	**2c**	**3a**	**3b**	**3c**	**4b**	**4c**	**5b**	**5c**
minimum	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5th	0.0002	0.0001	0.0002	0.0002	0.0001	0.0003	0.0000	0.0003	0.0000	0.0004
10th	0.0033	0.0031	0.0034	0.0034	0.0035	0.0035	0.0002	0.0034	0.0021	0.0034
25th	0.9755	0.2231	1.0040	0.8361	0.4025	0.9298	0.0645	0.4819	0.2013	0.4904
50th	0.9985	0.8699	1.0412	0.9944	0.9356	1.0238	0.7716	0.9960	0.8342	0.9695
75th	1.0292	1.0200	1.1070	1.0287	0.9972	1.1027	1.0033	1.0400	0.9783	1.0334
90th	1.1202	1.1643	1.4829	1.0710	1.0288	1.3225	1.0876	1.1523	1.0097	1.0968
95th	1.9730	1.9951	2.0901	1.1359	1.2944	2.0255	1.8612	1.9597	1.2120	1.3236
maximum	5.0722	5.4429	5.6469	5.2681	6.5563	6.6870	3.9667	4.0137	4.0927	4.1096

Values for relative BMDL are (BMDL for Method in question) / (BMDL for Method 1).

7. DIAGRAMS

Diagram 1a: Flow Diagram for Methods 1 – 3

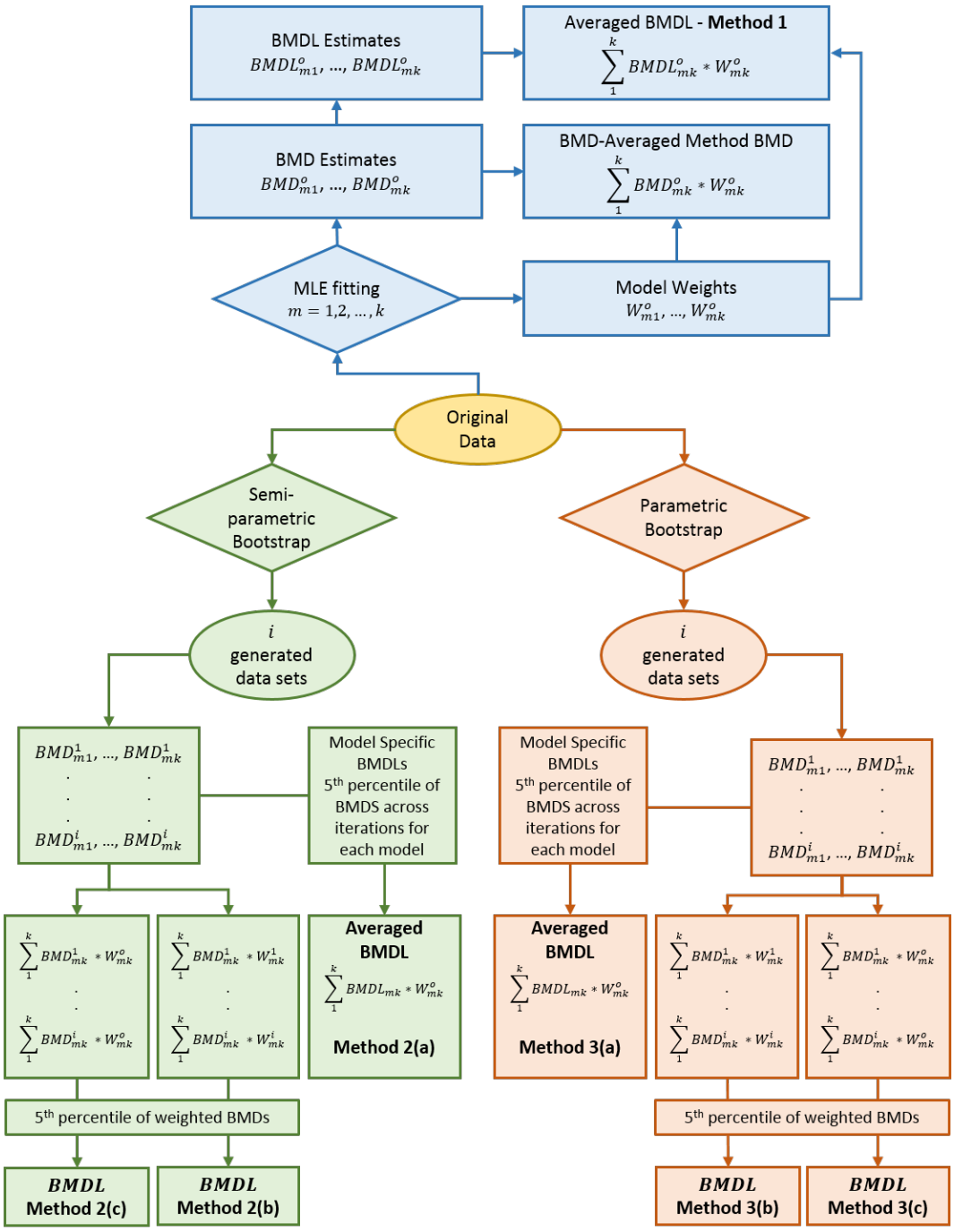


Diagram 1b: Flow Diagram for Methods 4 – 5

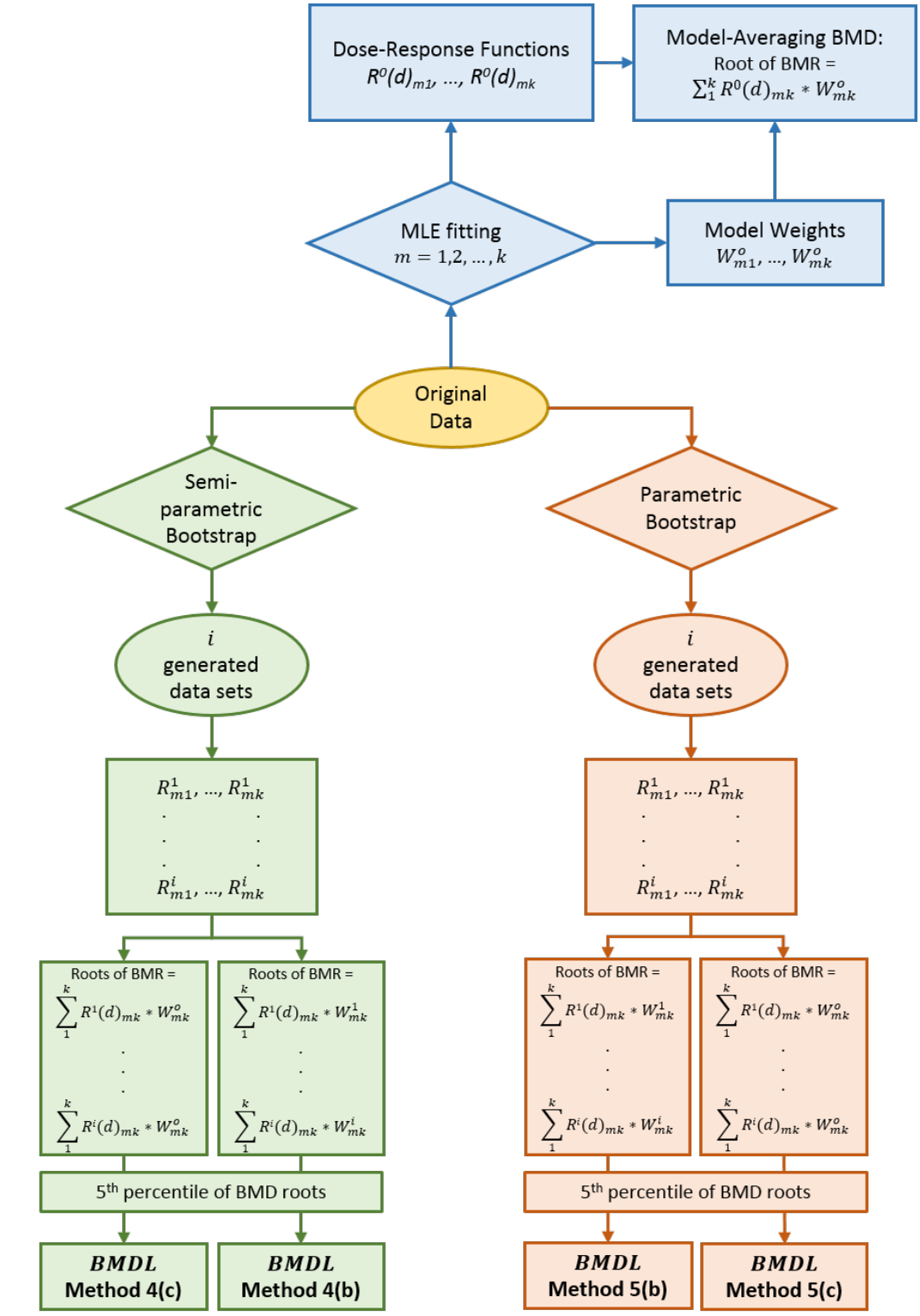


Diagram Abbreviations:

BMD_{mi}^{j}: BMD for model i from the jth bootstrap-generated data set.
$BMDL_{mi}^{0}$: BMDL estimate for model i, model fit to original data.
W_{mi}^{0}: Weight for model i, based on fit (BIC) of that model to original data.
W_{mi}^{j}: Weight for model i, based on fit (BIC) of that model to jth bootstrap-generated data set.

$R^{0}(d)_{mi}$: Fitted dose-response function for model i, fit to original data.
$R^{j}(d)_{mi}$: Fitted dose-response function for model i, fit to jth bootstrap-generated data set.

k models are included in the averaging.

8. FIGURES

Figure 1: Dose-response for p1 Templates

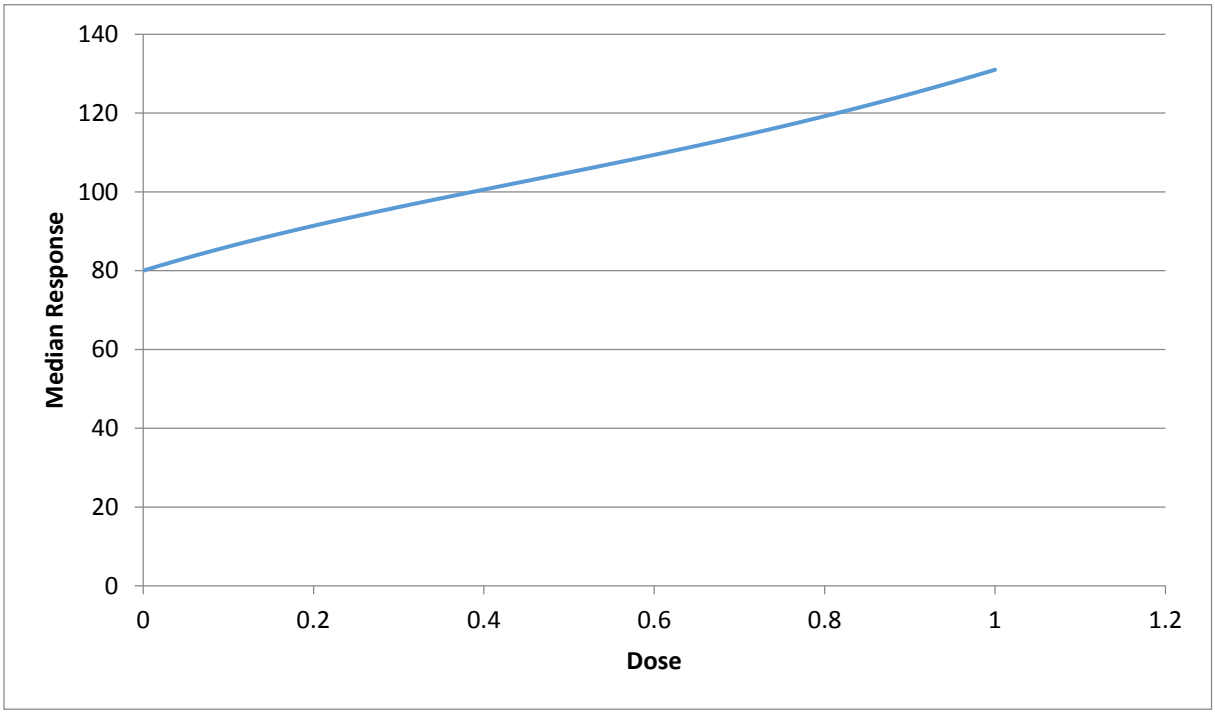


Figure 2: Dose-response for p2 Templates

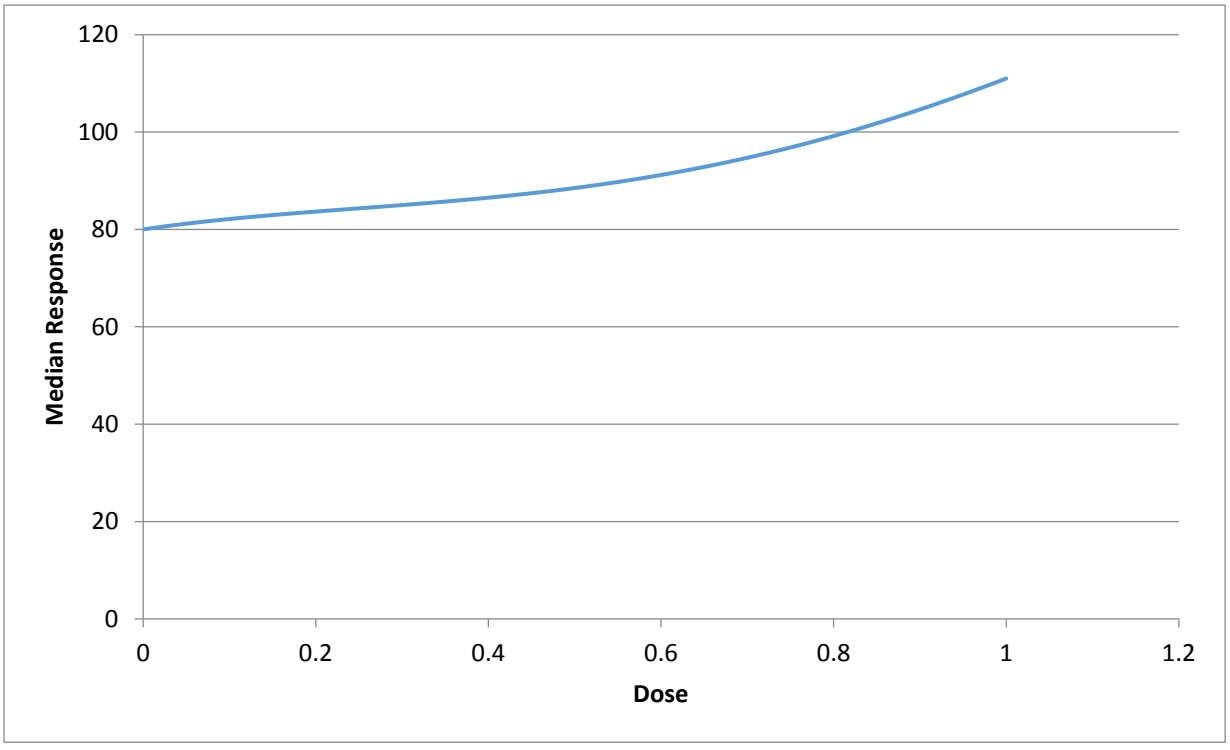


Figure 3: Dose-response for p3 Templates

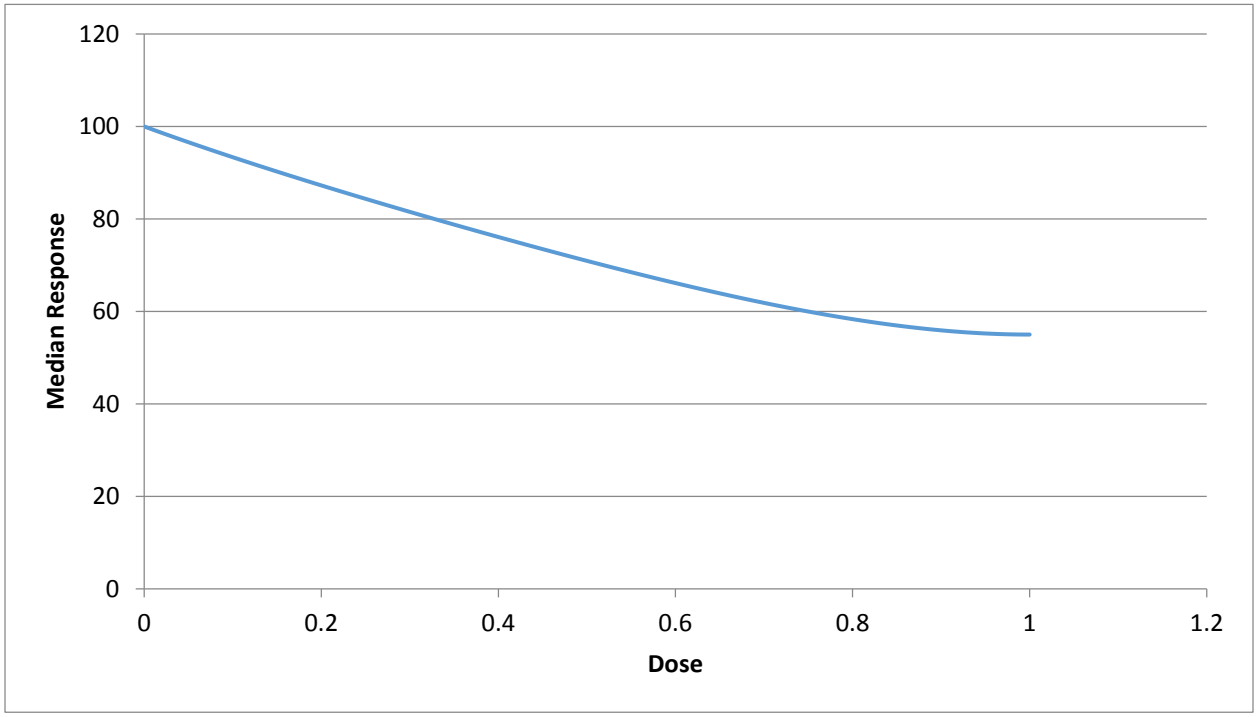


Figure 4: Dose-response for p4 Templates

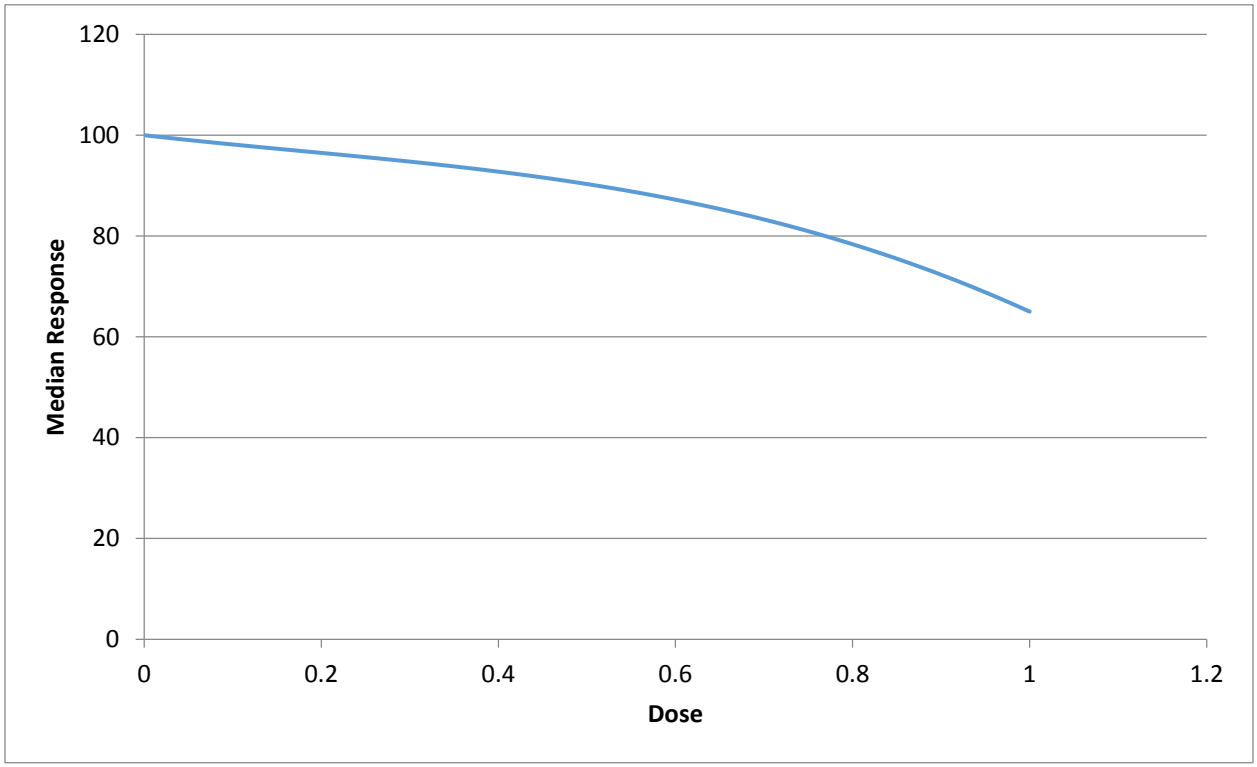


Figure 5: Dose-response for e1 Templates

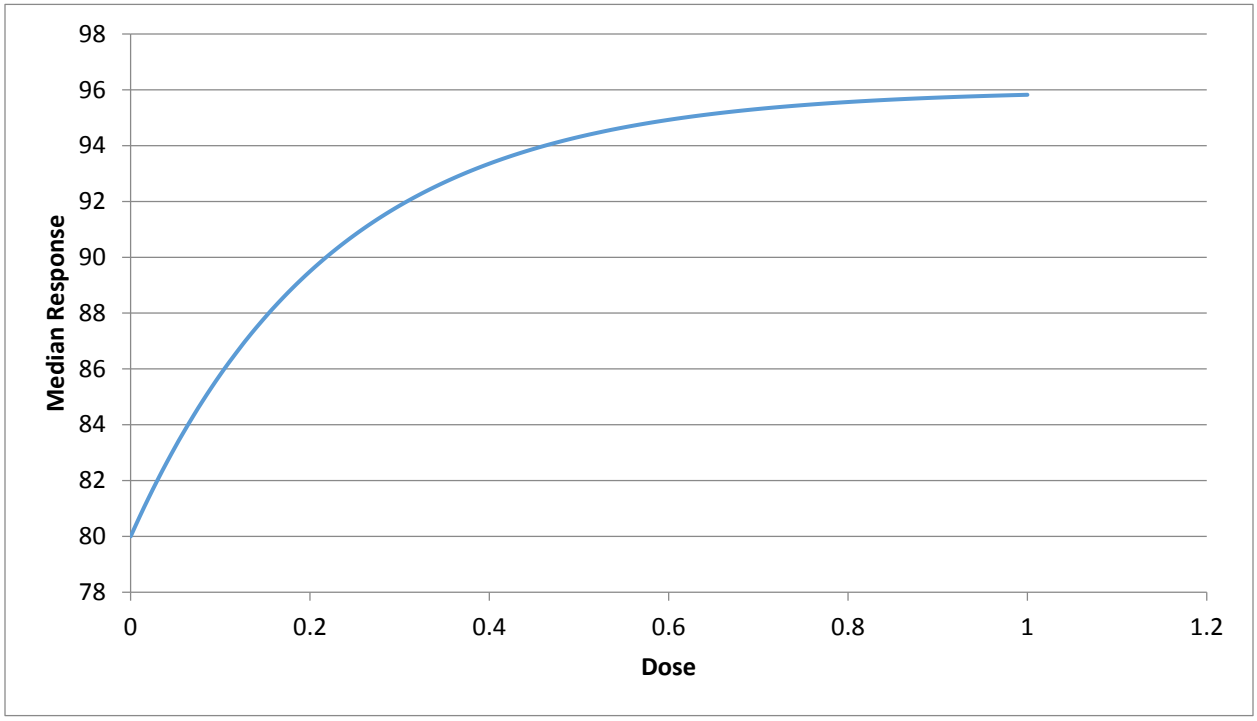


Figure 6: Dose-response for e2 Templates

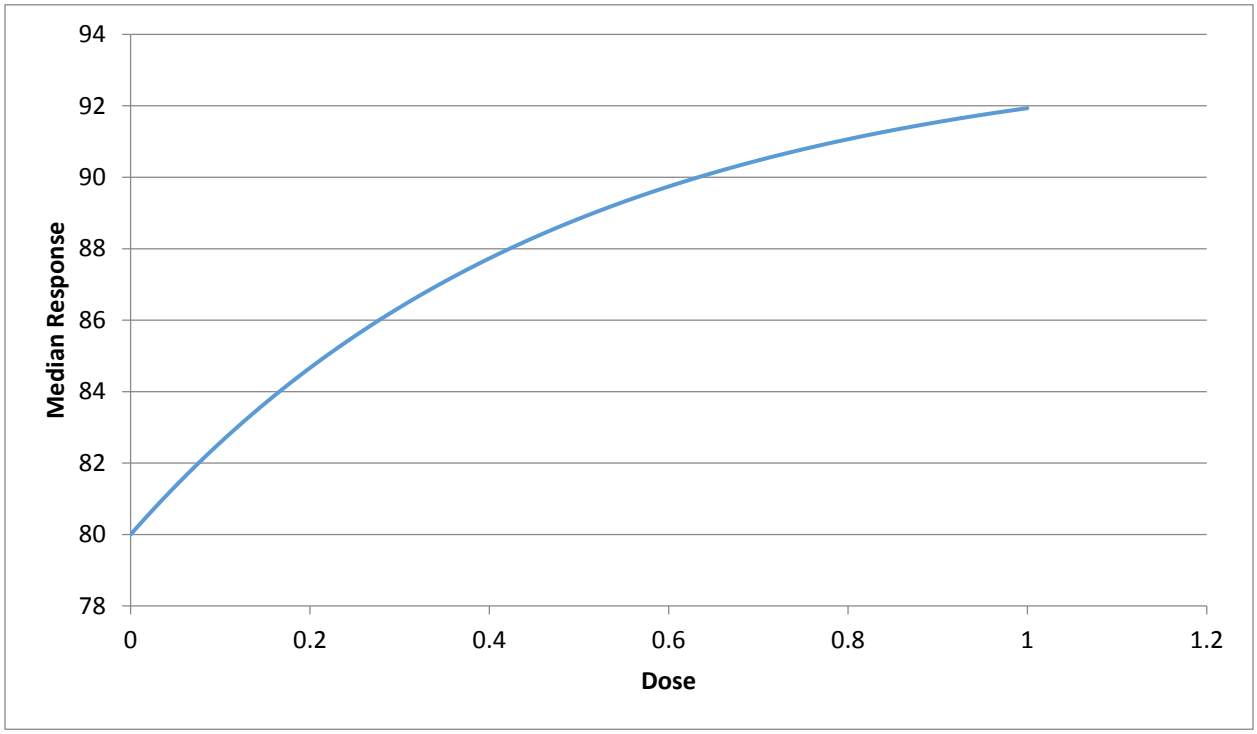


Figure 7: Dose-response for e3 Templates

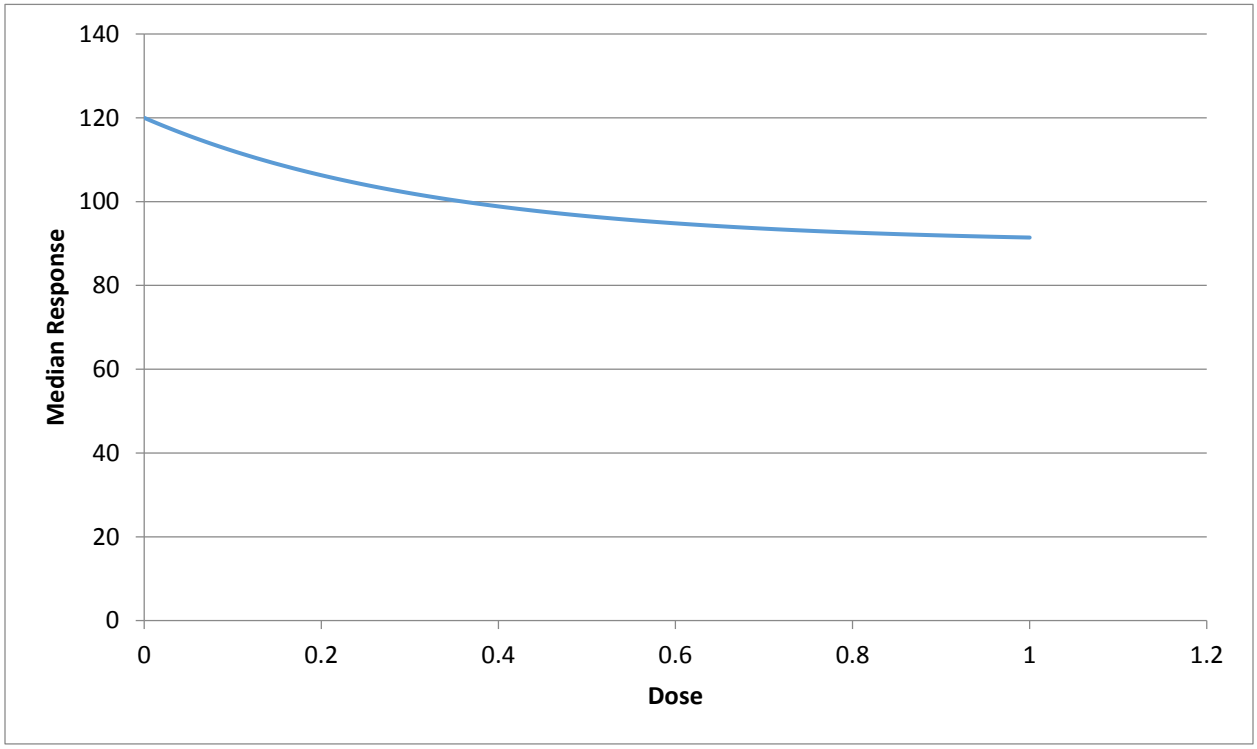


Figure 8: Dose-response for e4 Templates

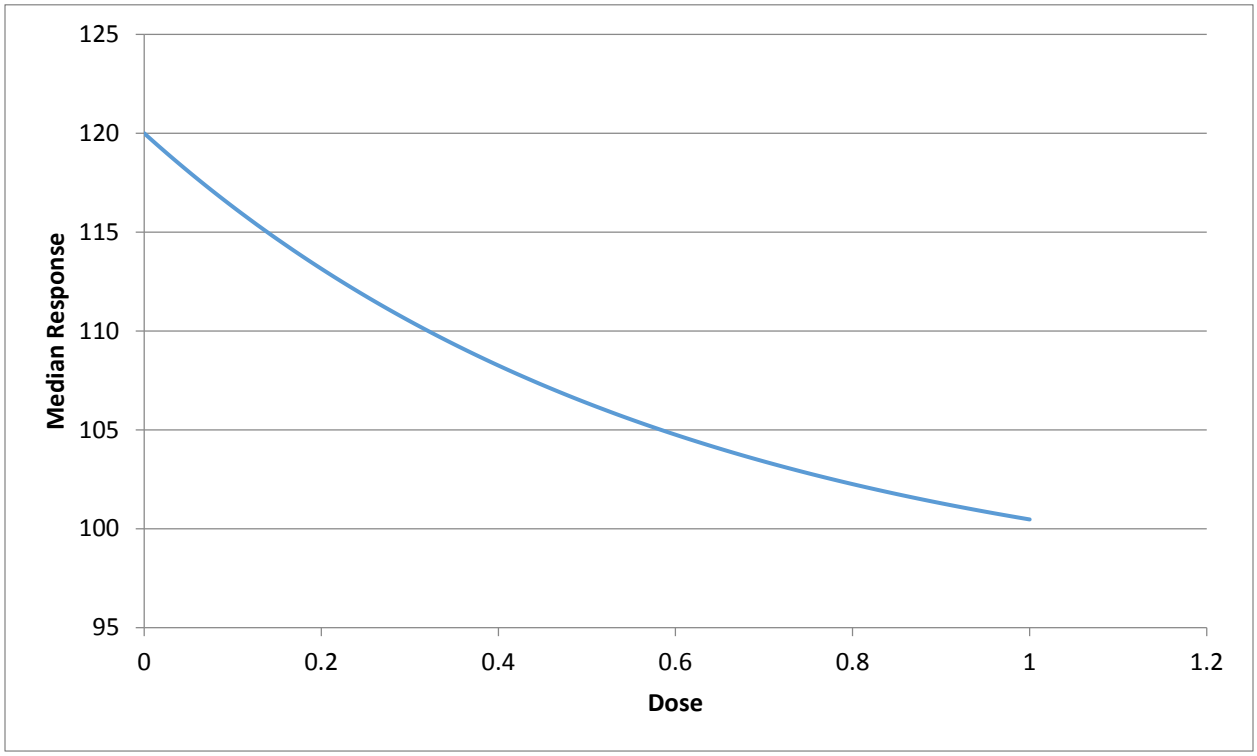


Figure 9: Dose-response for w1 Templates

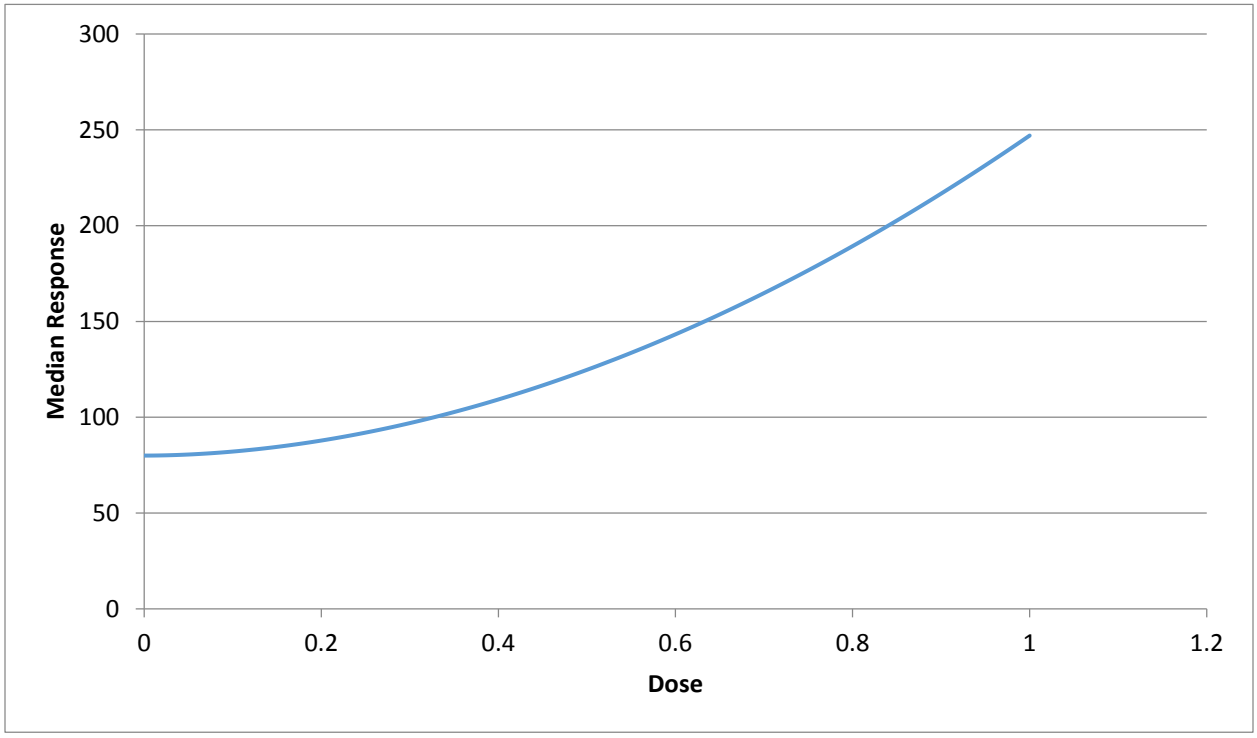


Figure 10: Dose-response for w2 Templates

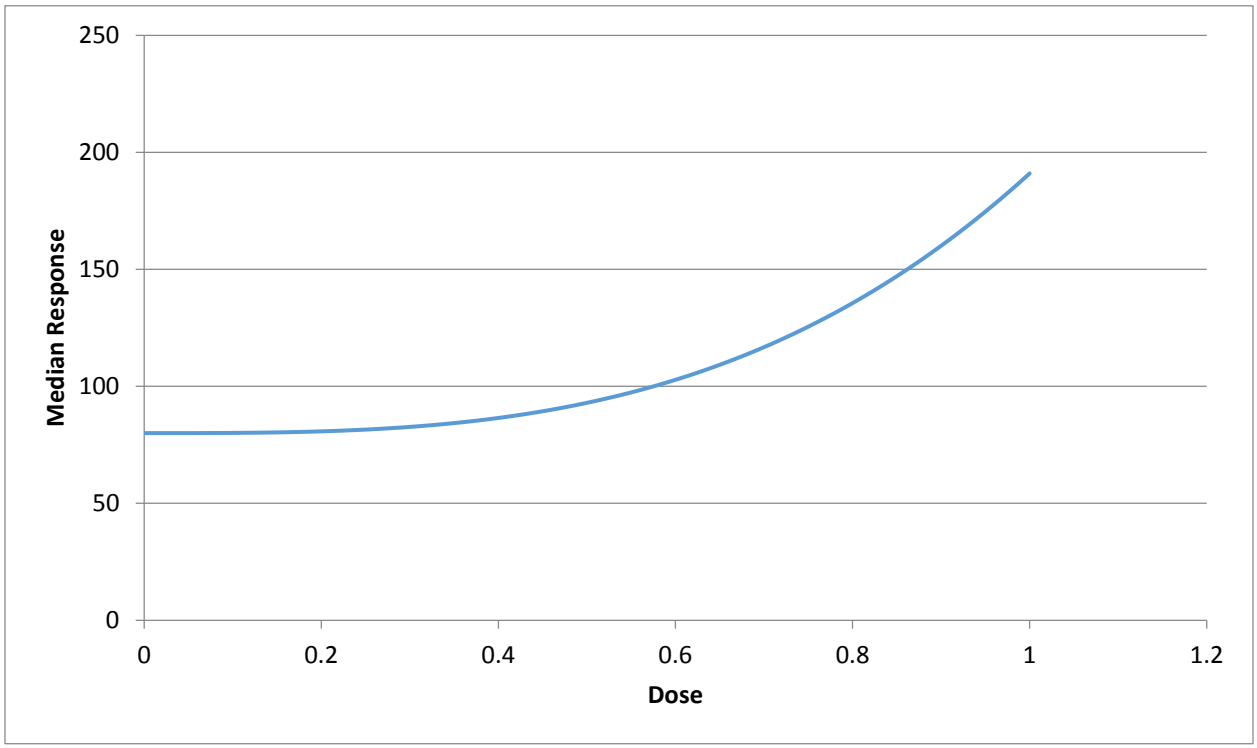


Figure 11: Dose-response for w3

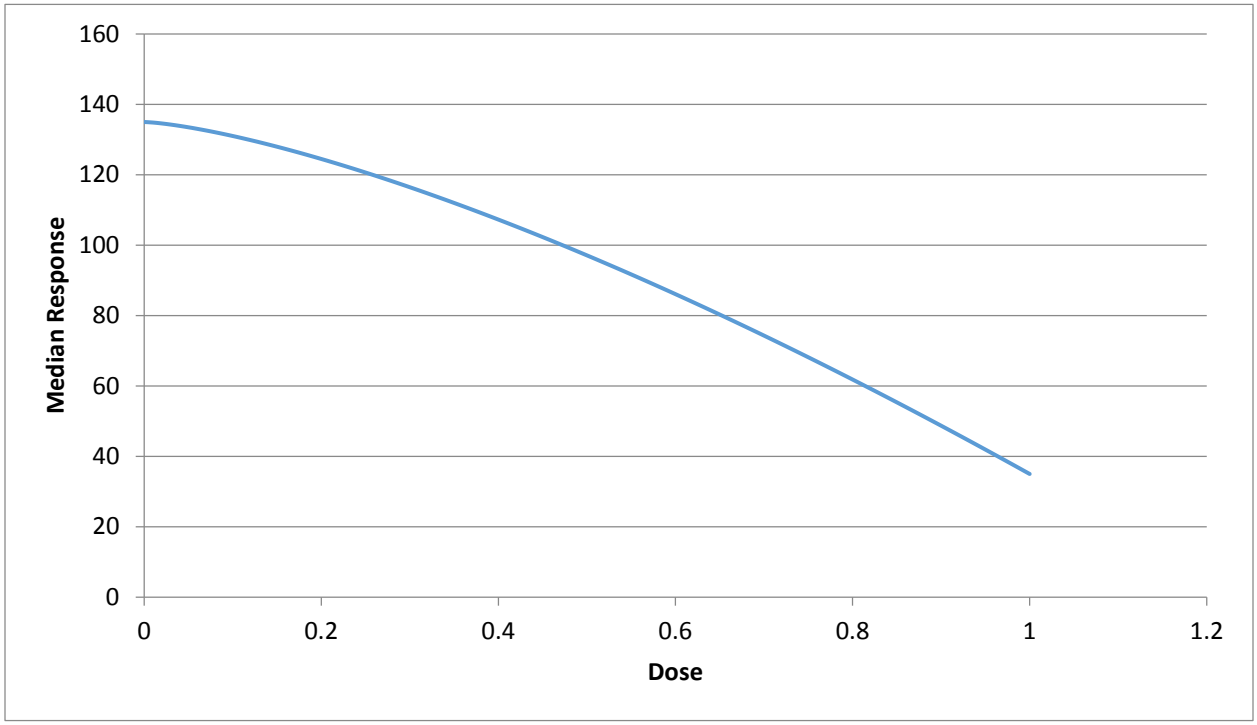


Figure 12: Dose-response for w4 Templates

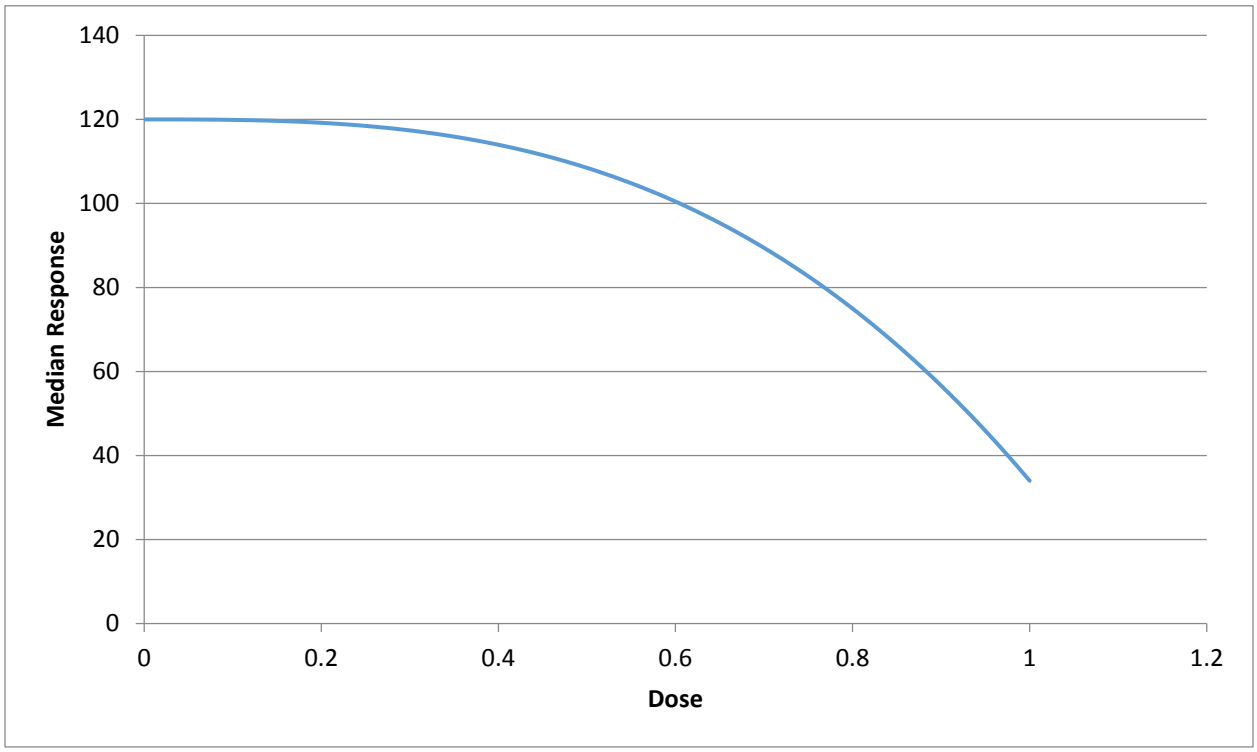


Figure 13: Dose-response for h1 Templates

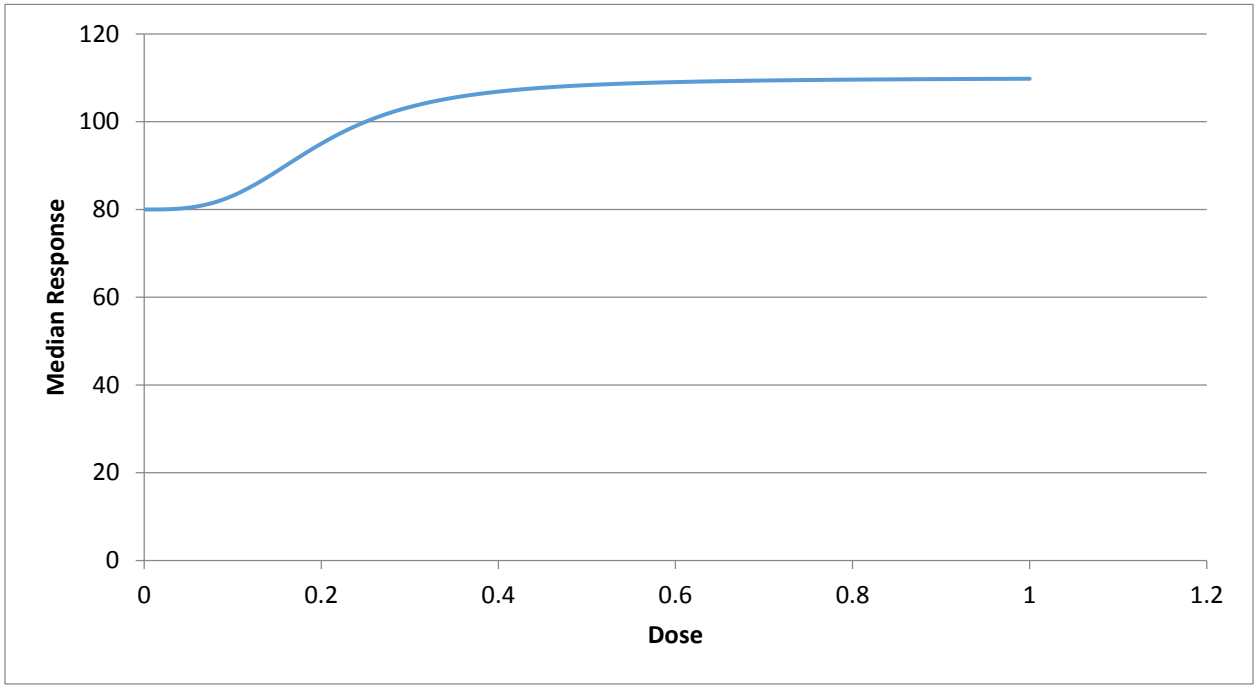


Figure 14: Dose-response for h2 Templates

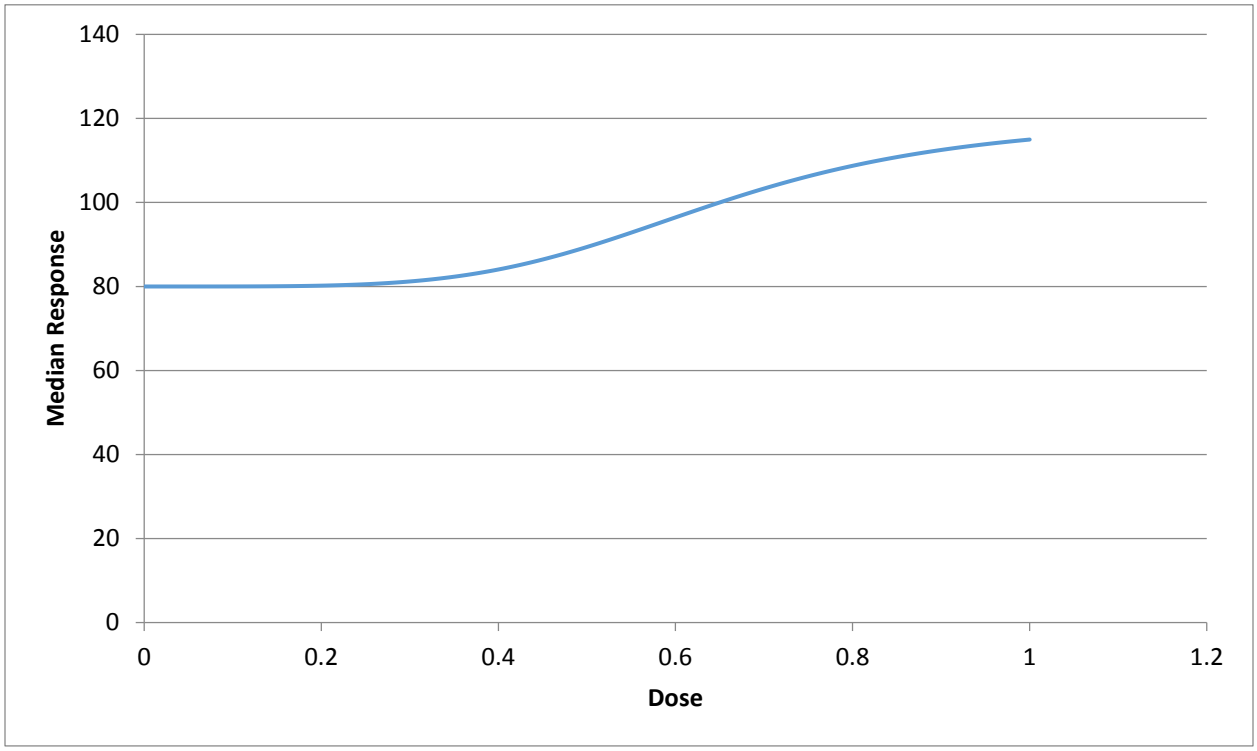


Figure 15: Dose-response for h3 Templates

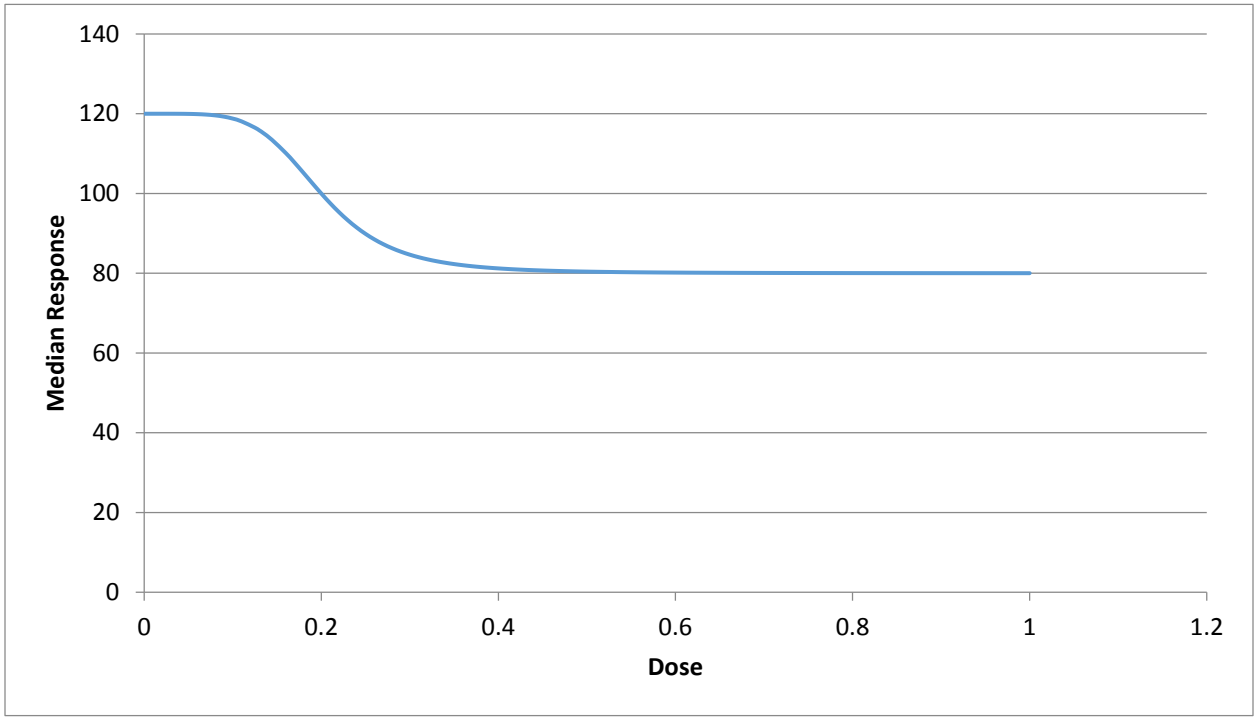


Figure 16: Dose-response for h4 Templates

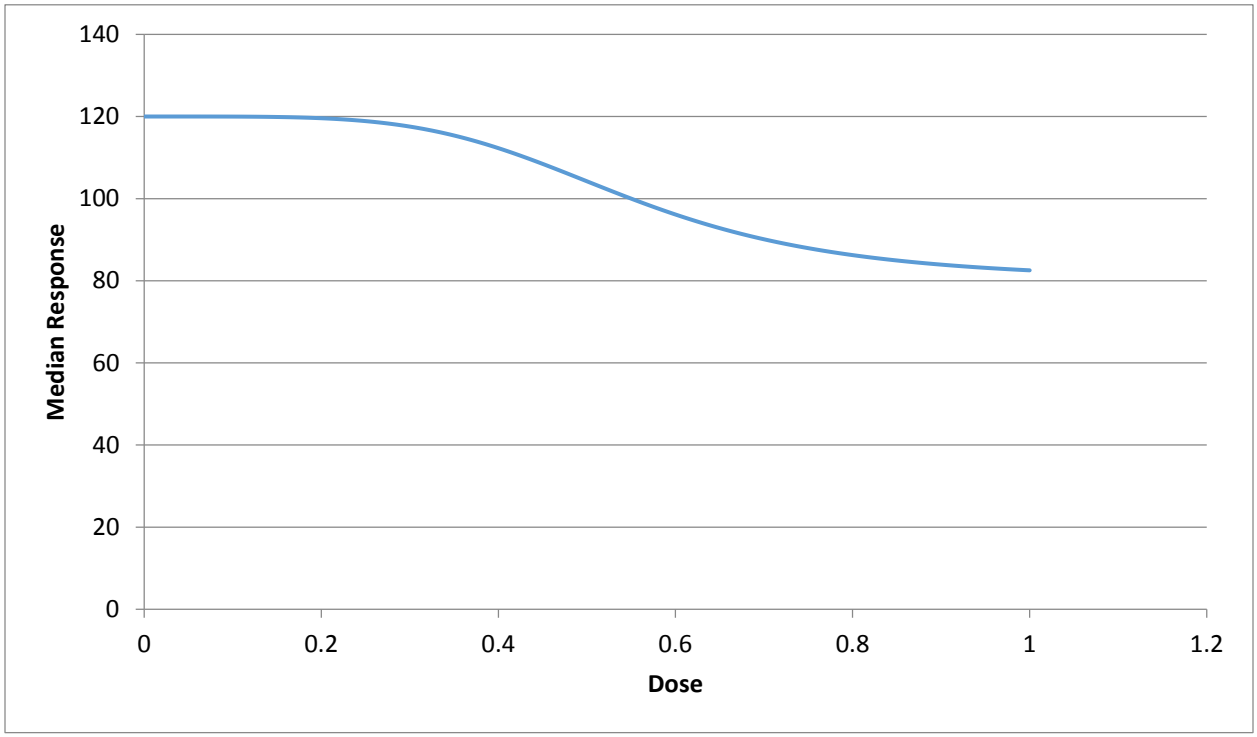
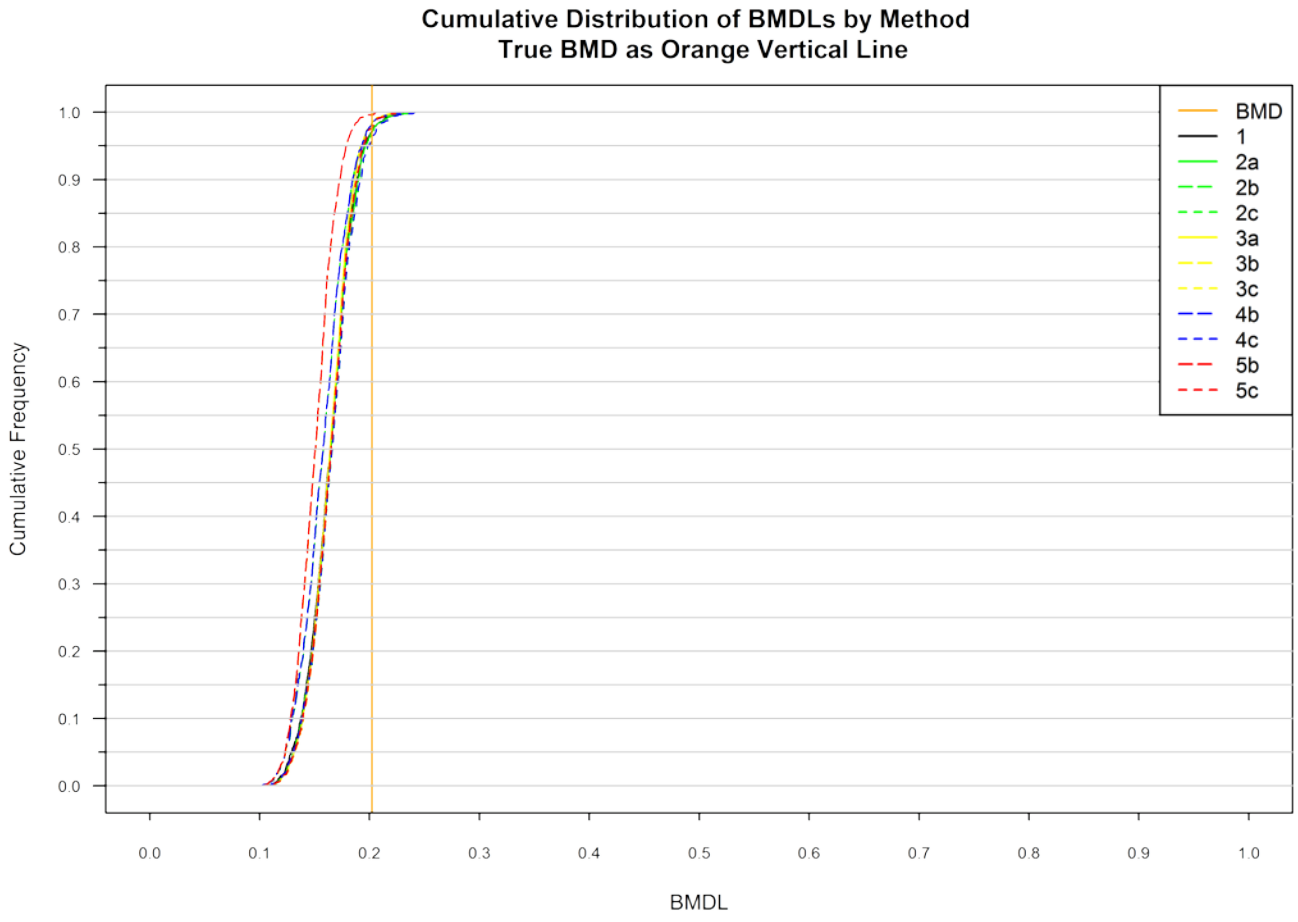


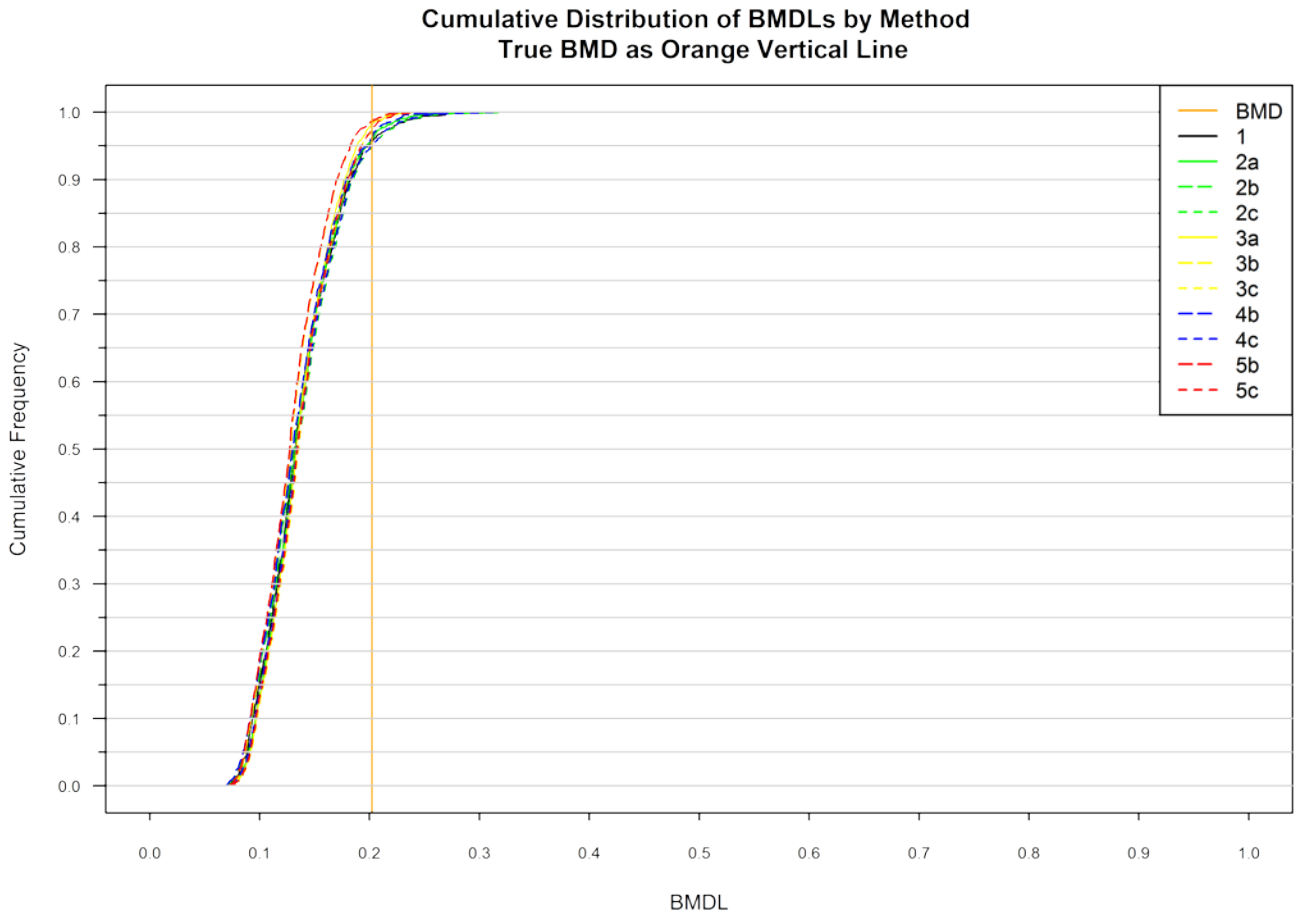
Figure 17: Template w1_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.176844	0.177393	
50	0.191976	0.192116	0.2021
75	0.205432	0.205638	
IQR	0.0285882	0.0282453	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.972		
Poly3	0.955	**2a**	0.969	**3a**	0.983
Power	0.937	**2b**	0.984	**3b**	0.996
Hill	0.899	**2c**	0.961	**3c**	0.976
Exp3	1	**4b**	0.984	**5b**	0.996
Exp5	0.912	**4c**	0.961	**5c**	0.976

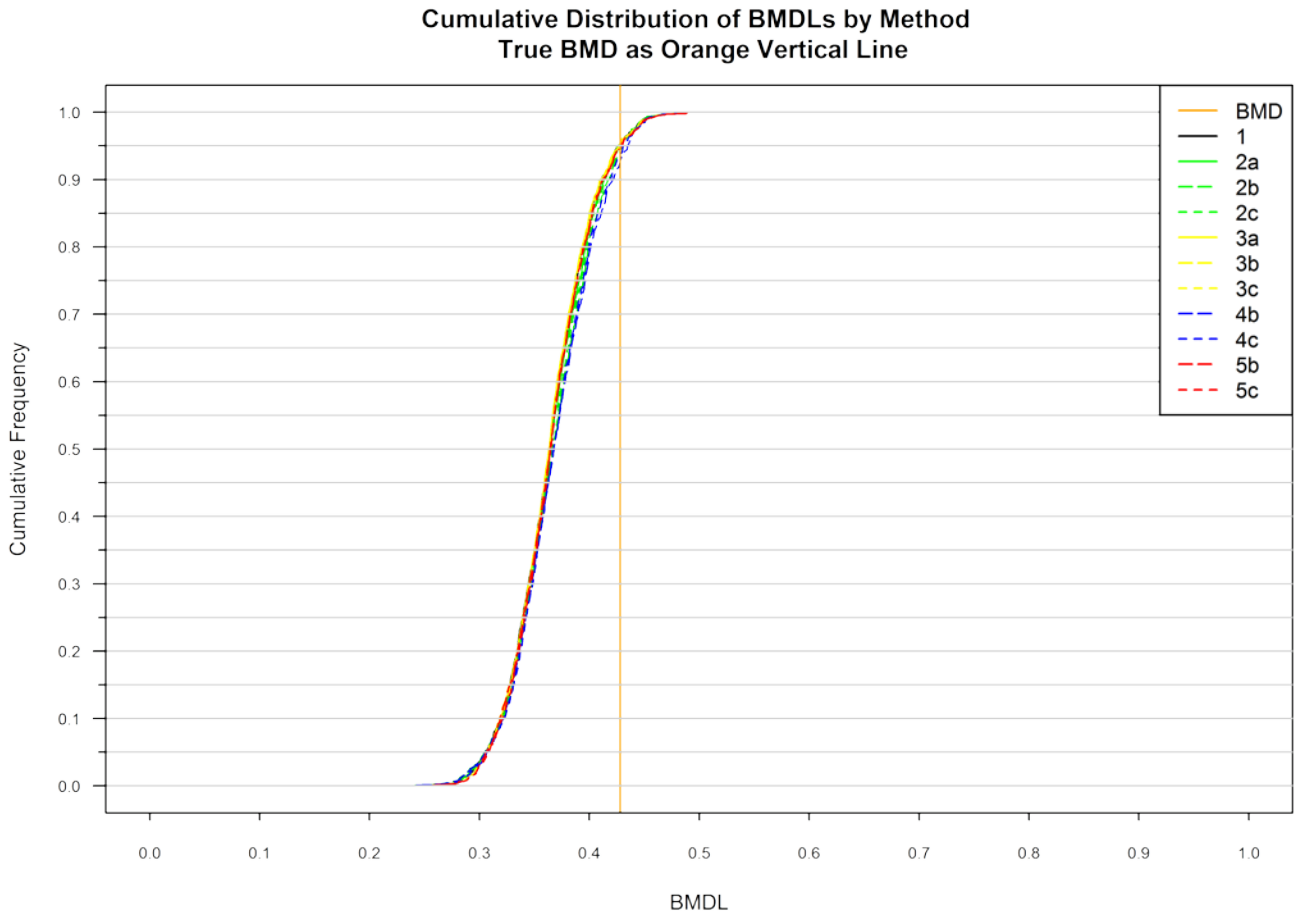
Figure 18: Template w1_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.160301	0.161422	
50	0.187973	0.188754	0.2021
75	0.221049	0.221267	
IQR	0.0607481	0.0598444	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.957		
Poly3	0.91	**2a**	0.959	**3a**	0.982
Power	0.936	**2b**	0.96	**3b**	0.987
Hill	0.873	**2c**	0.949	**3c**	0.972
Exp3	0.998	**4b**	0.961	**5b**	0.987
Exp5	0.905	**4c**	0.952	**5c**	0.975

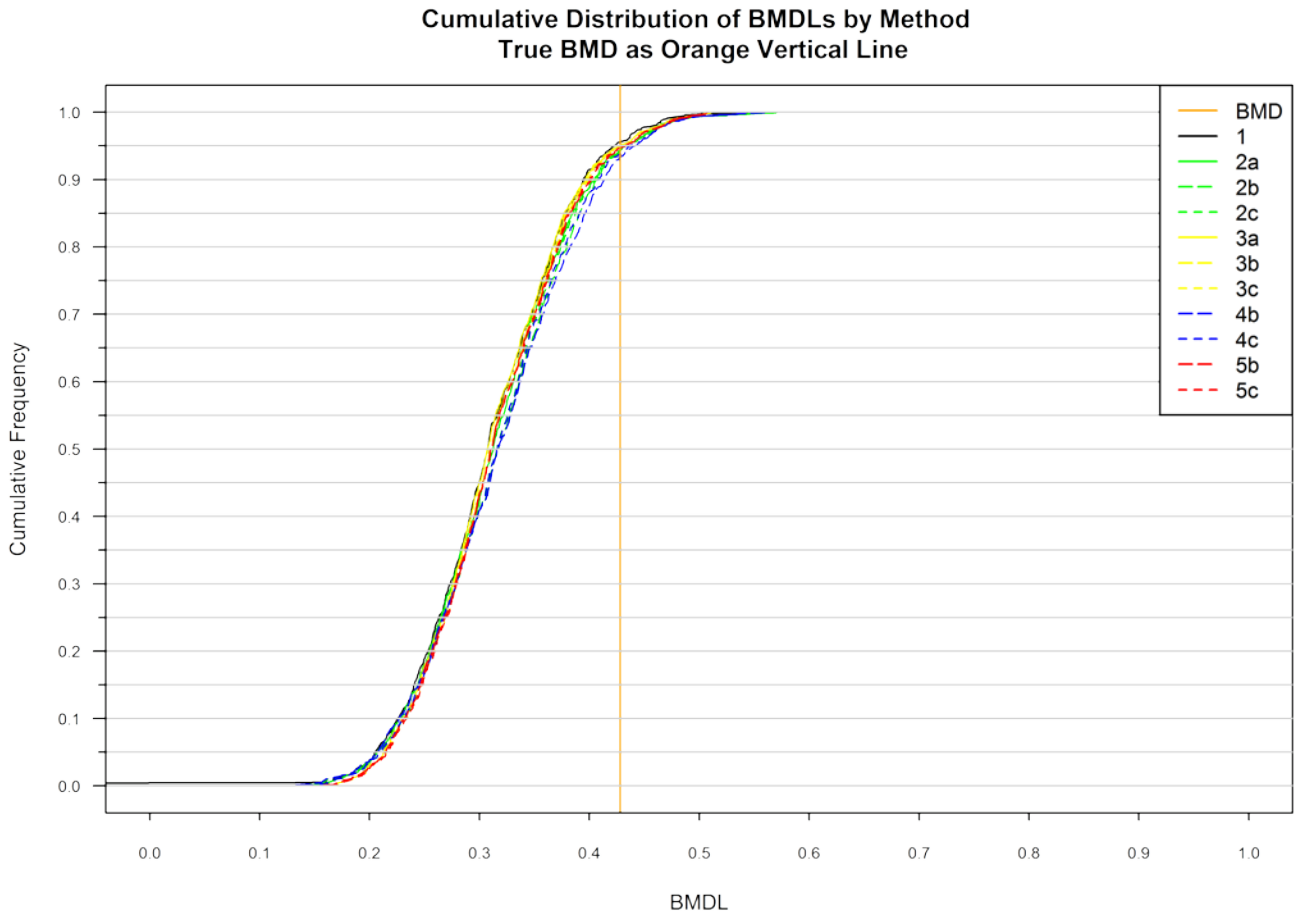
Figure 19: Template w2_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.400244	0.40391	
50	0.425902	0.429576	0.4281
75	0.451866	0.456503	
IQR	0.0516214	0.0525931	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.952		
Poly3	1	**2a**	0.952	**3a**	0.954
Power	0.938	**2b**	0.943	**3b**	0.949
Hill	0.922	**2c**	0.952	**3c**	0.951
Exp3	0.965	**4b**	0.926	**5b**	0.944
Exp5	0.947	**4c**	0.937	**5c**	0.947

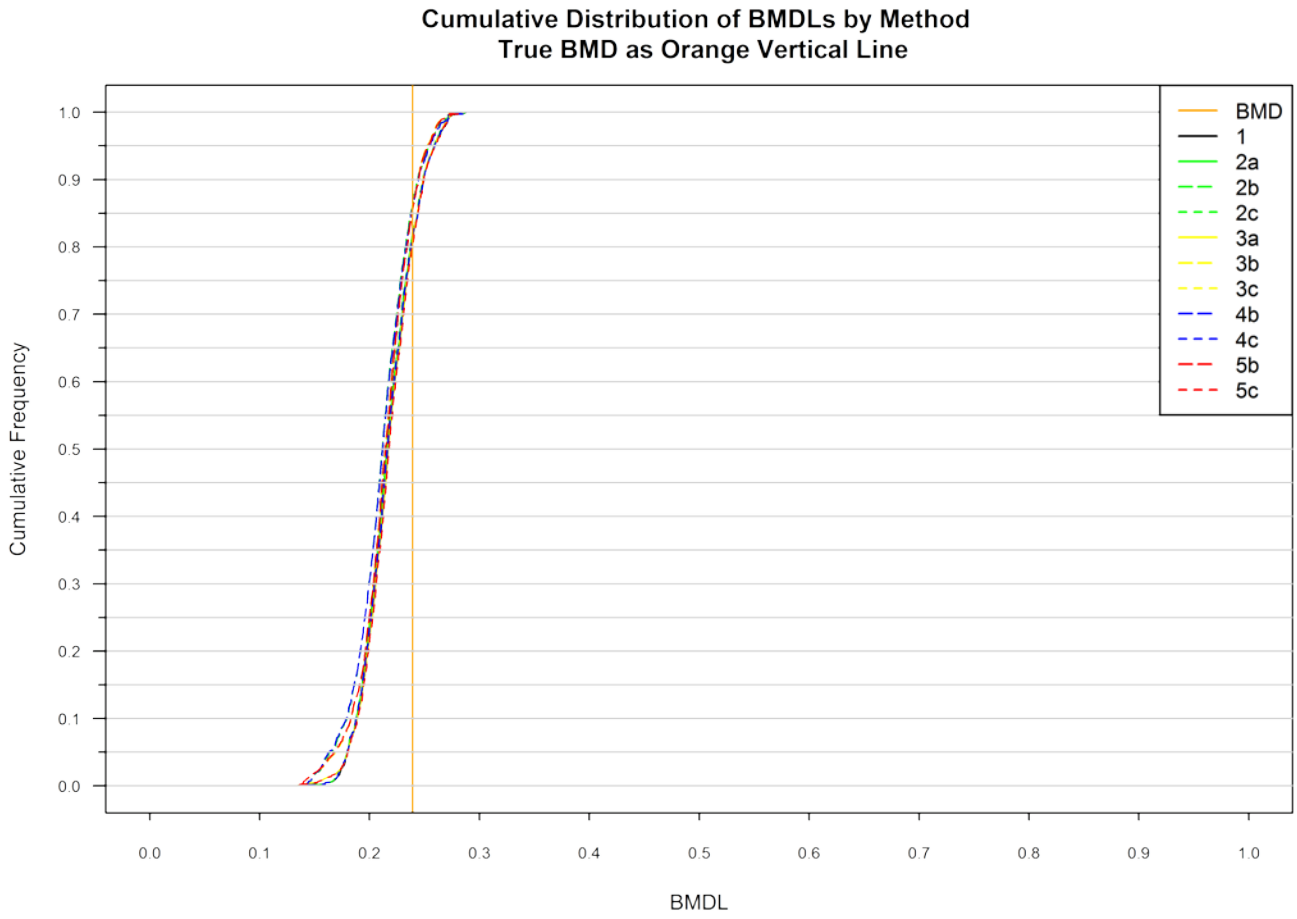
Figure 20: Template w2_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.378695	0.38203	
50	0.425097	0.432969	0.4281
75	0.481935	0.484141	
IQR	0.10324	0.102111	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.955		
Poly3	1	**2a**	0.944	**3a**	0.949
Power	0.944	**2b**	0.941	**3b**	0.948
Hill	0.944	**2c**	0.943	**3c**	0.948
Exp3	0.959	**4b**	0.934	**5b**	0.946
Exp5	0.957	**4c**	0.937	**5c**	0.947

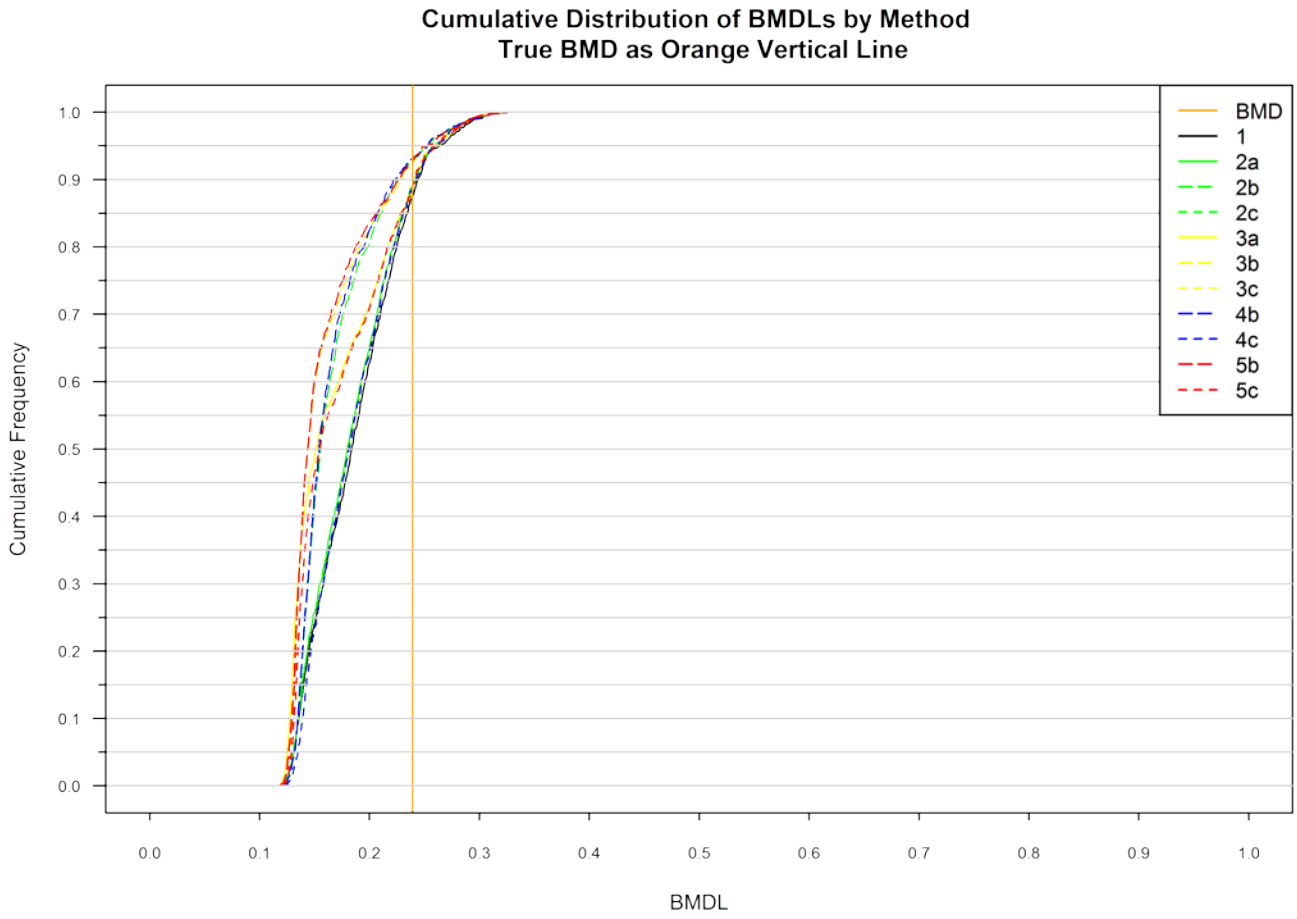
Figure 21: Template w3_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.233413	0.234039	
50	0.249751	0.250529	0.2392
75	0.267721	0.268477	
IQR	0.0343073	0.0344378	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.816		
Poly3	0.997	**2a**	0.819	**3a**	0.805
Power	0.936	**2b**	0.863	**3b**	0.861
Hill	0.887	**2c**	0.823	**3c**	0.805
Exp3	0.339	**4b**	0.858	**5b**	0.854
Exp5	0.338	**4c**	0.814	**5c**	0.798

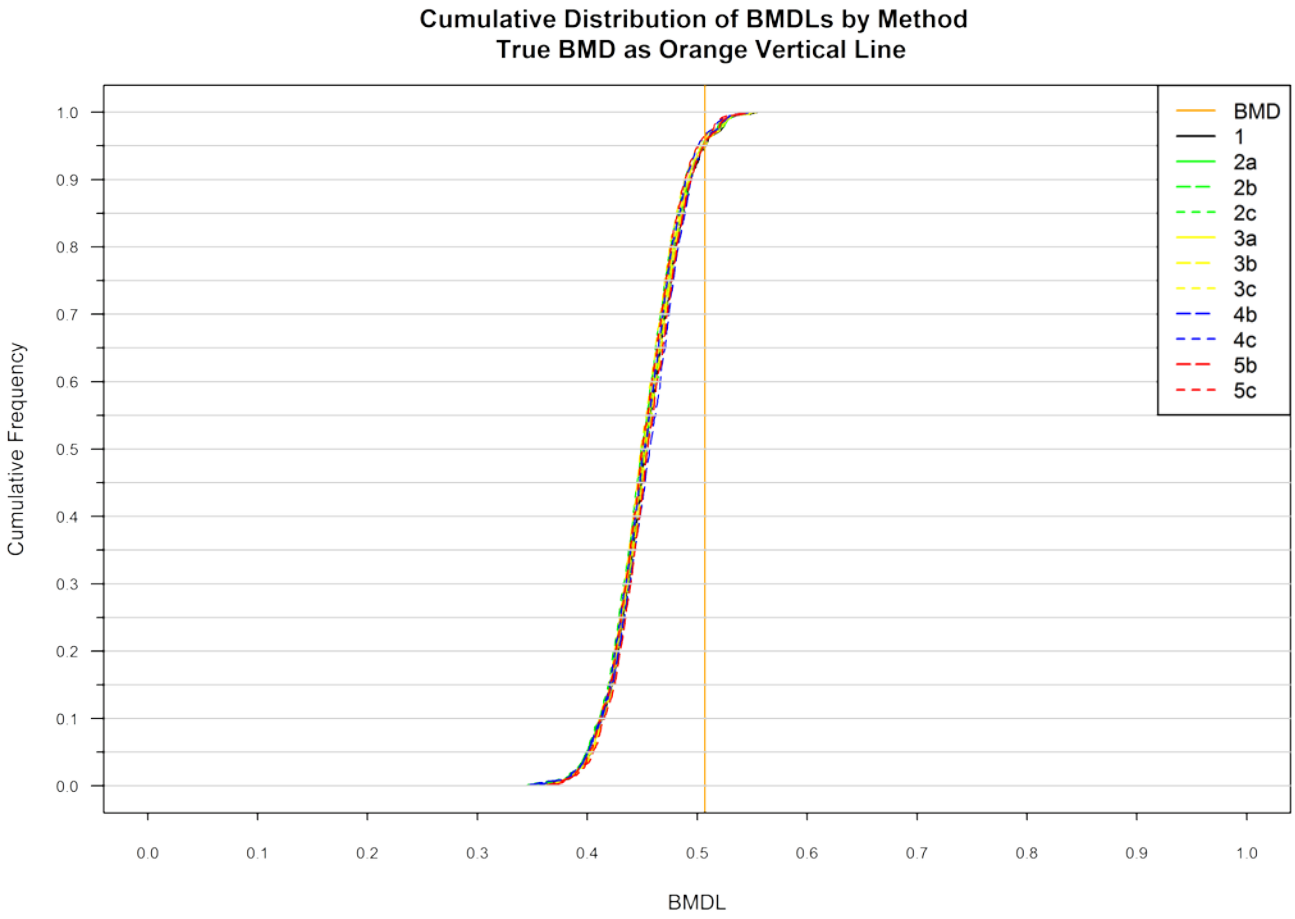
Figure 22: Template w3_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.196491	0.192879	
50	0.245456	0.24402	0.2392
75	0.288386	0.289079	
IQR	0.0918952	0.0961996	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.876		
Poly3	0.983	**2a**	0.893	**3a**	0.892
Power	0.947	**2b**	0.929	**3b**	0.922
Hill	0.905	**2c**	0.892	**3c**	0.895
Exp3	0.715	**4b**	0.932	**5b**	0.927
Exp5	0.694	**4c**	0.89	**5c**	0.886

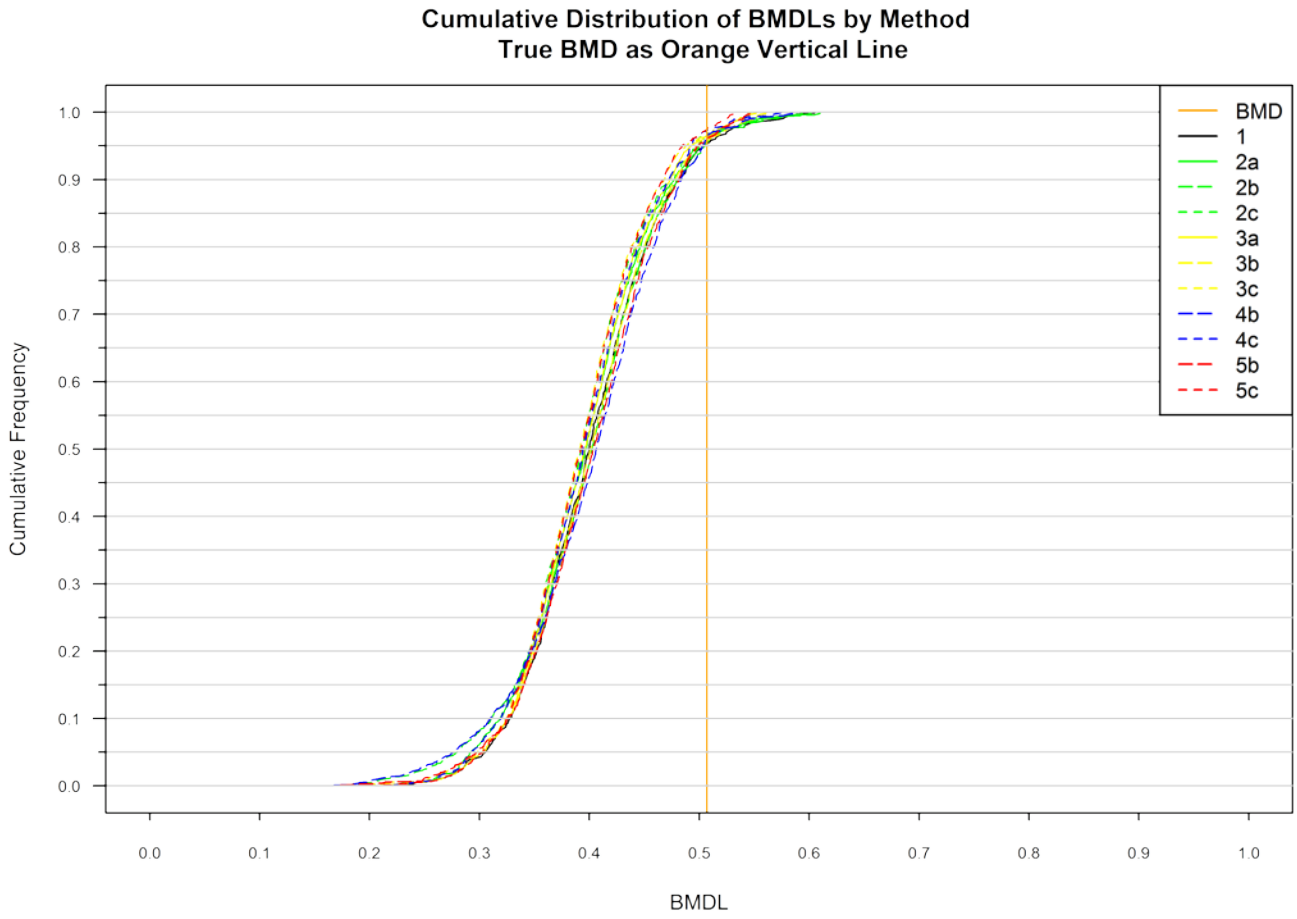
Figure 23: Template w4_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.485532	0.48711	
50	0.506508	0.506008	0.5071
75	0.531691	0.528259	
IQR	0.0461584	0.0411496	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.948		
Poly3	0.969	**2a**	0.958	**3a**	0.958
Power	0.949	**2b**	0.951	**3b**	0.952
Hill	0.967	**2c**	0.964	**3c**	0.964
Exp3	0.944	**4b**	0.954	**5b**	0.952
Exp5	0.959	**4c**	0.967	**5c**	0.965

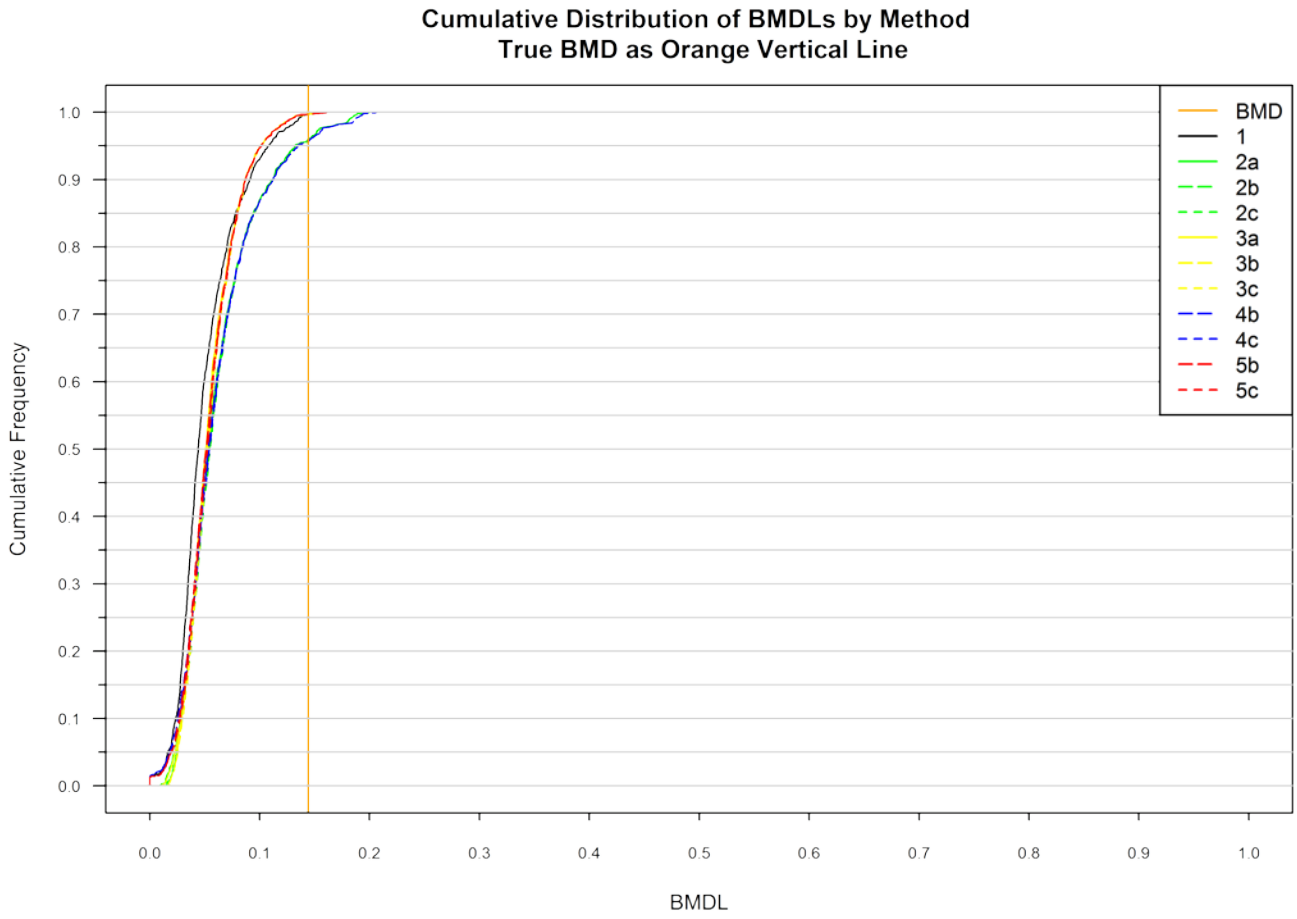
Figure 24: Template w4_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.460557	0.465478	
50	0.504909	0.503688	0.5071
75	0.551827	0.544815	
IQR	0.09127	0.0793372	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.954		
Poly3	0.977	**2a**	0.965	**3a**	0.963
Power	0.951	**2b**	0.955	**3b**	0.957
Hill	0.966	**2c**	0.968	**3c**	0.965
Exp3	0.954	**4b**	0.961	**5b**	0.96
Exp5	0.955	**4c**	0.971	**5c**	0.974

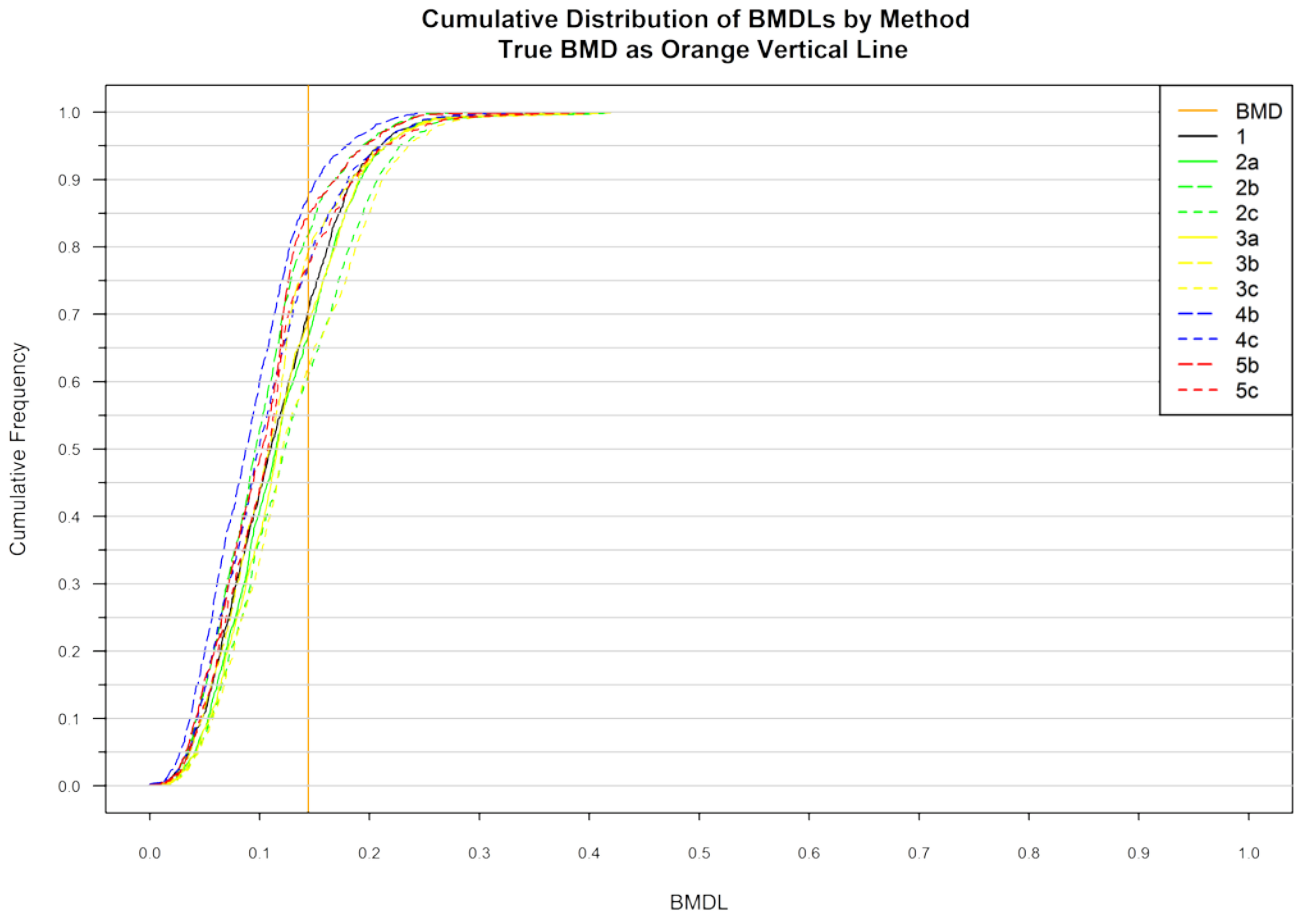
Figure 25: Template h1_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.0863961	0.0866751	
50	0.133543	0.135916	0.1443
75	0.19488	0.201558	
IQR	0.108484	0.114883	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.997		
Poly3	0	**2a**	0.96	**3a**	0.997
Power	0	**2b**	0.961	**3b**	0.997
Hill	0.979	**2c**	0.961	**3c**	0.996
Exp3	0	**4b**	0.956	**5b**	0.996
Exp5	0.999	**4c**	0.957	**5c**	0.996

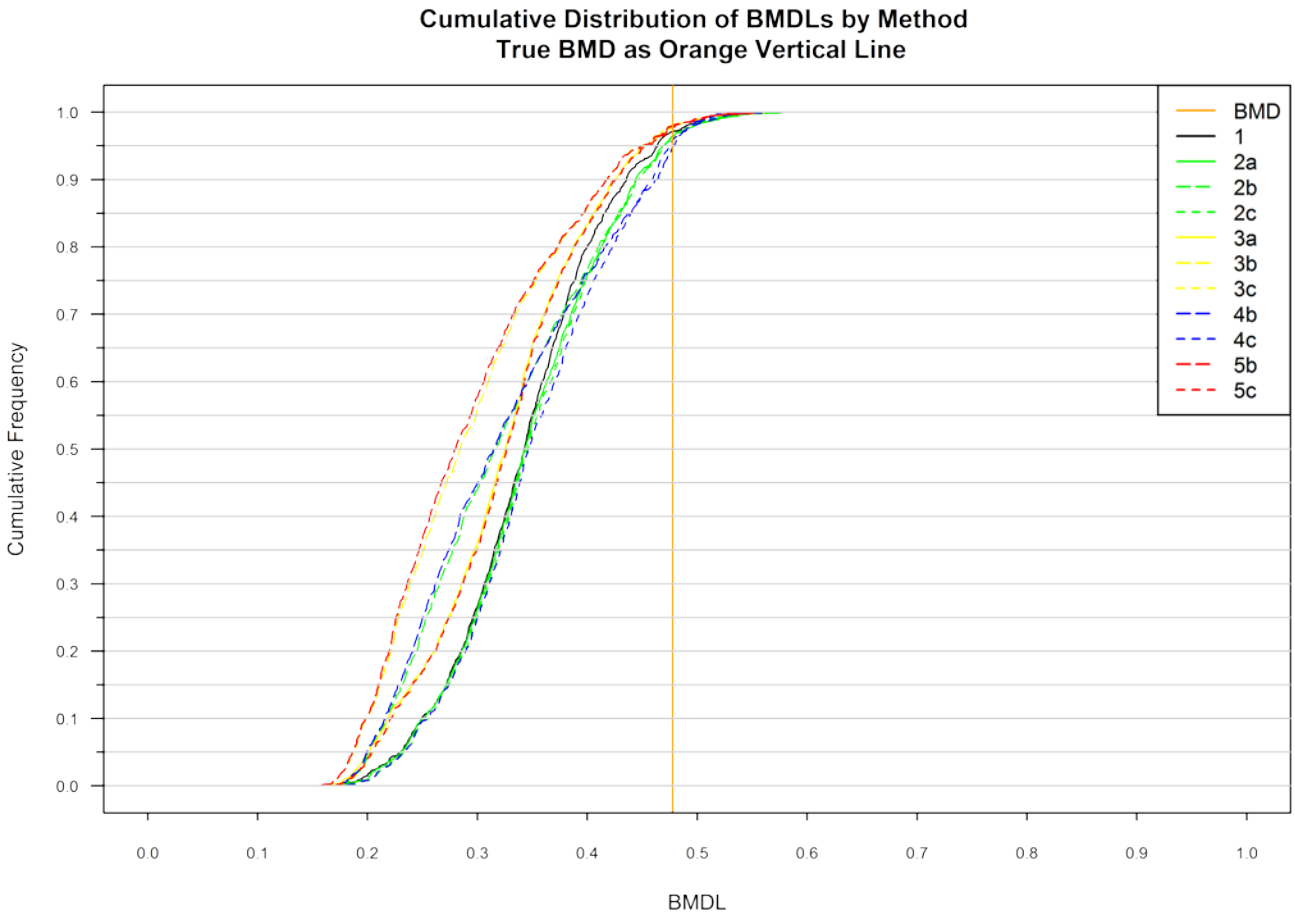
Figure 26: Template h1_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.141553	0.126939	
50	0.199694	0.172439	0.1443
75	0.23972	0.227303	
IQR	0.0981671	0.100364	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.008	**1**	0.707		
Poly3	0.008	**2a**	0.668	**3a**	0.691
Power	0.008	**2b**	0.818	**3b**	0.793
Hill	0.946	**2c**	0.61	**3c**	0.622
Exp3	0	**4b**	0.876	**5b**	0.85
Exp5	0.957	**4c**	0.775	**5c**	0.775

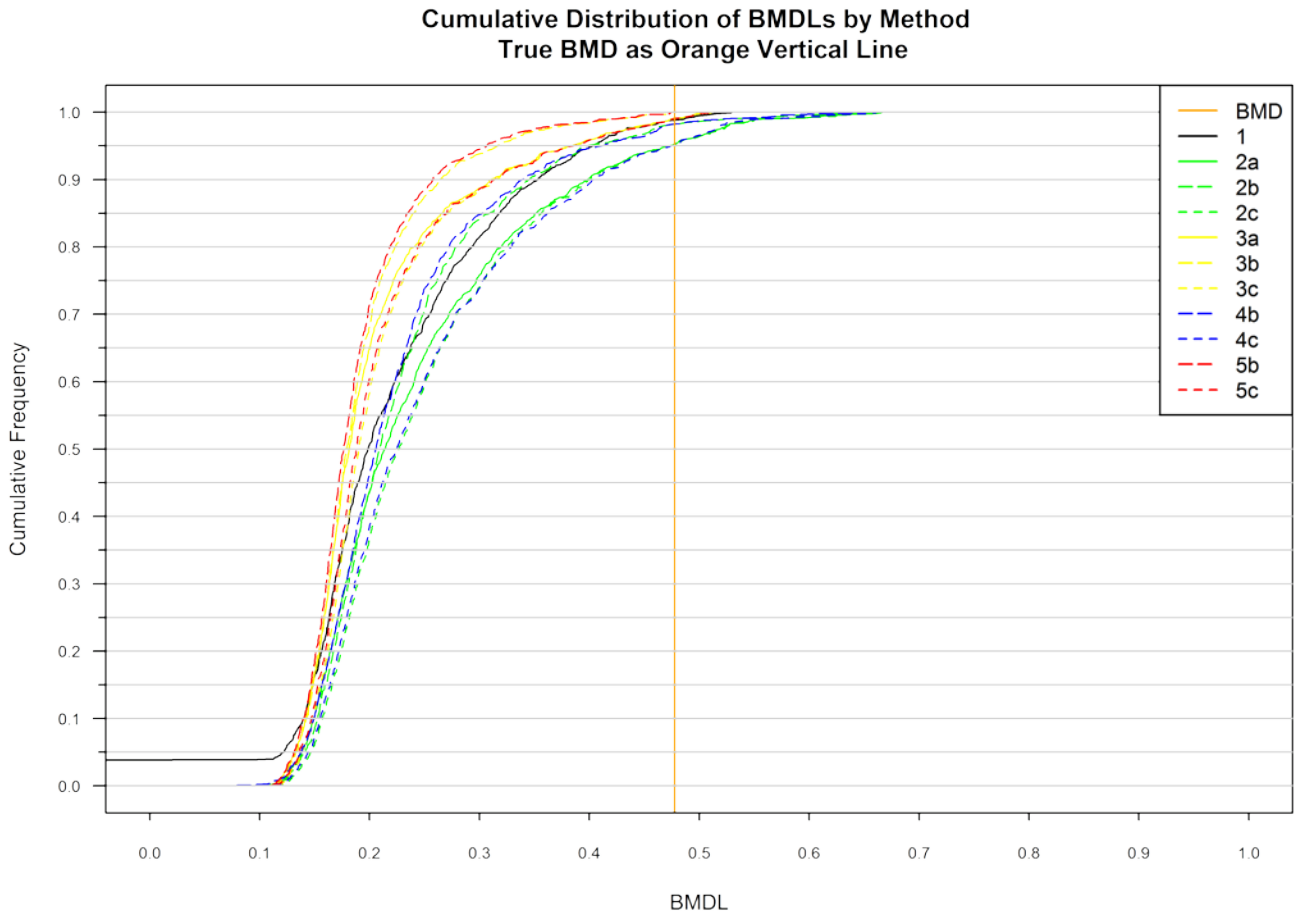
Figure 27: Template h2_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.415388	0.419346	
50	0.463034	0.473736	0.4777
75	0.509649	0.505302	
IQR	0.0942606	0.0859558	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.971		
Poly3	0.968	**2a**	0.963	**3a**	0.976
Power	0.971	**2b**	0.964	**3b**	0.98
Hill	0.967	**2c**	0.963	**3c**	0.976
Exp3	0.971	**4b**	0.947	**5b**	0.979
Exp5	0.967	**4c**	0.959	**5c**	0.976

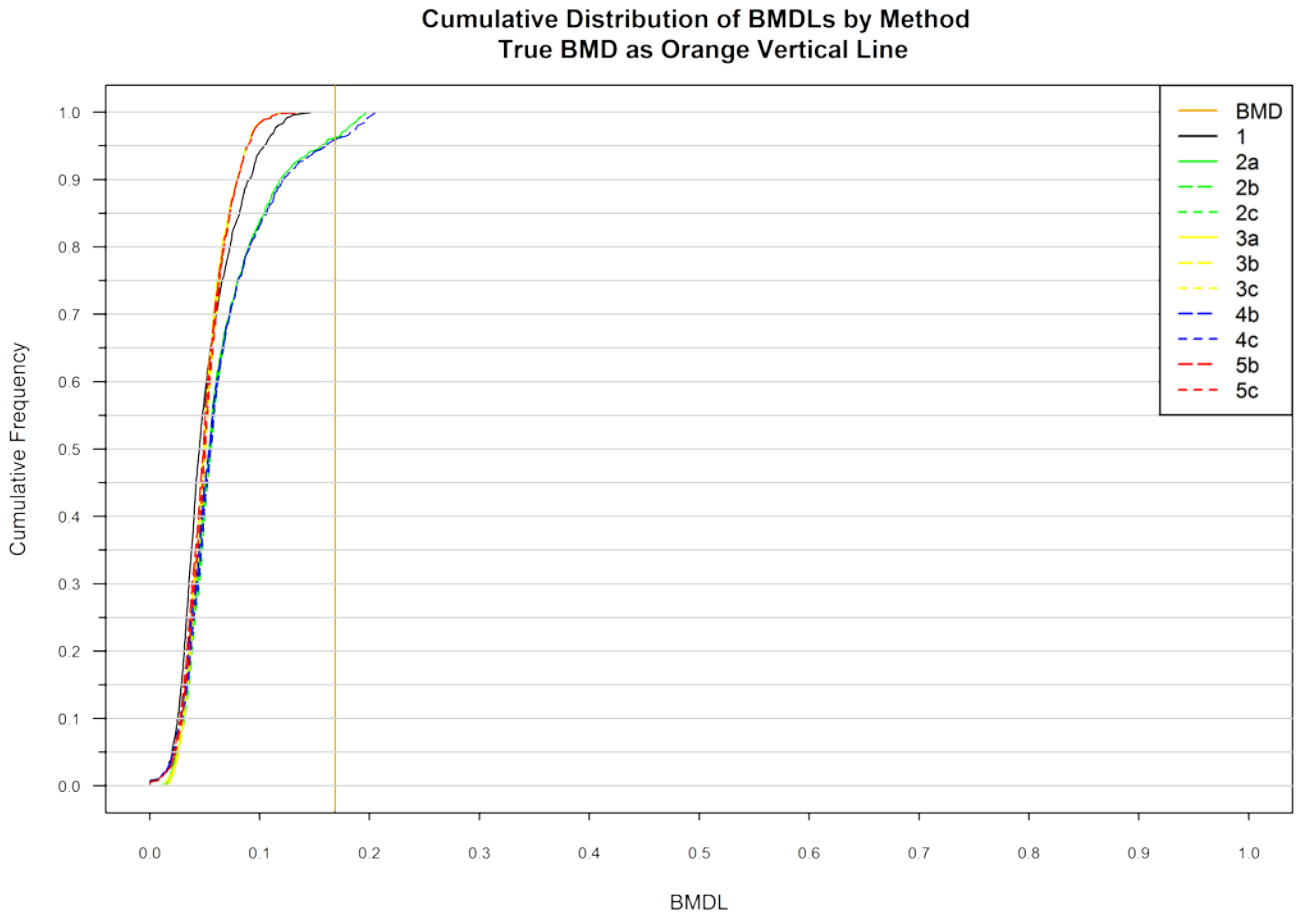
Figure 28: Template h2_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.271126	0.264344	
50	0.37863	0.381363	0.4777
75	0.507871	0.506784	
IQR	0.236745	0.24244	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.988		
Poly3	0.952	**2a**	0.952	**3a**	0.991
Power	0.952	**2b**	0.982	**3b**	0.999
Hill	0.956	**2c**	0.952	**3c**	0.99
Exp3	0.986	**4b**	0.982	**5b**	0.999
Exp5	0.986	**4c**	0.951	**5c**	0.99

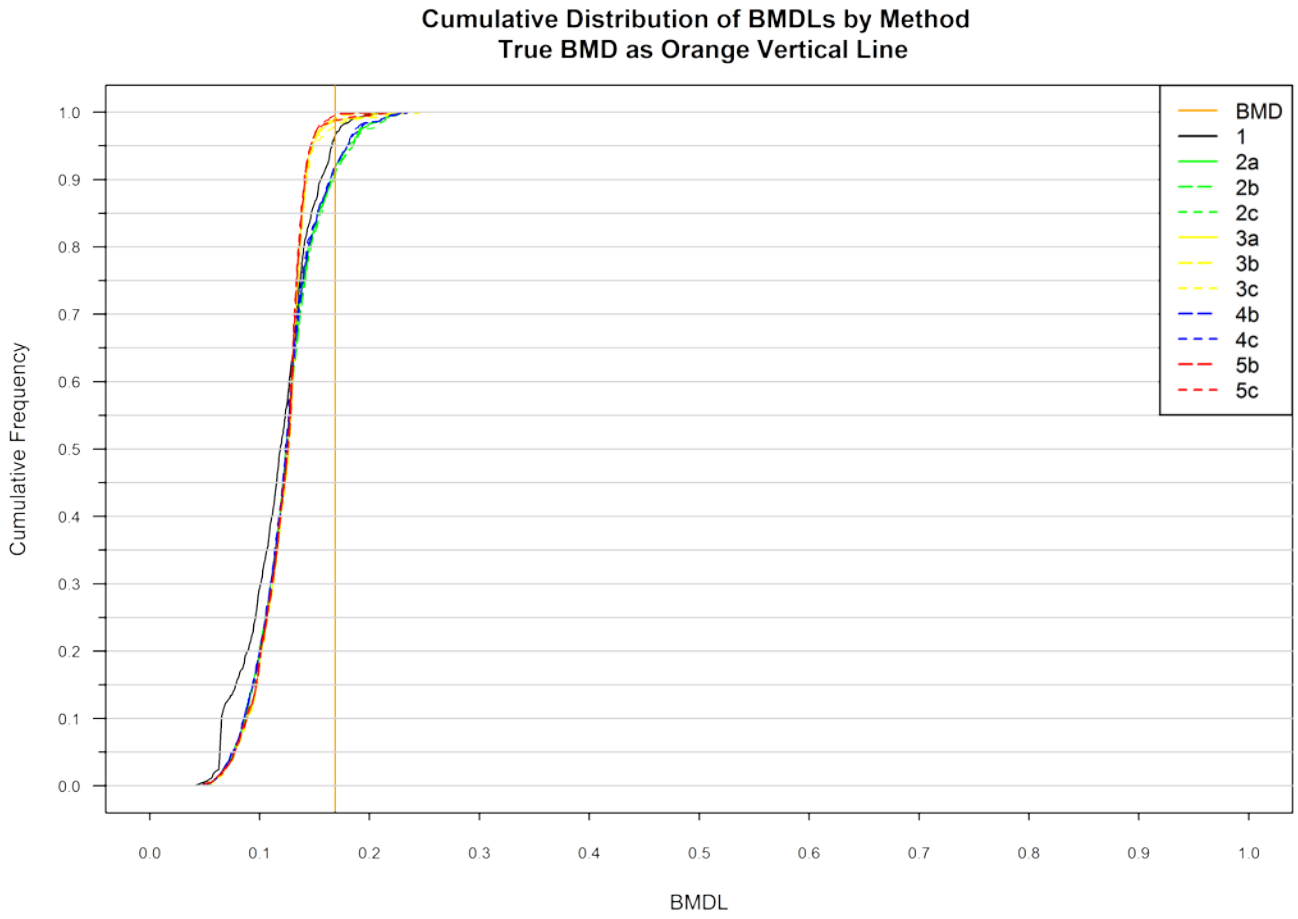
Figure 29: Template h3_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.097332	0.0991703	
50	0.151619	0.156612	0.1688
75	0.198402	0.206705	
IQR	0.10107	0.107535	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	1		
Poly3	0	**2a**	0.962	**3a**	1
Power	0	**2b**	0.961	**3b**	1
Hill	1	**2c**	0.961	**3c**	1
Exp3	0.001	**4b**	0.959	**5b**	1
Exp5	1	**4c**	0.959	**5c**	1

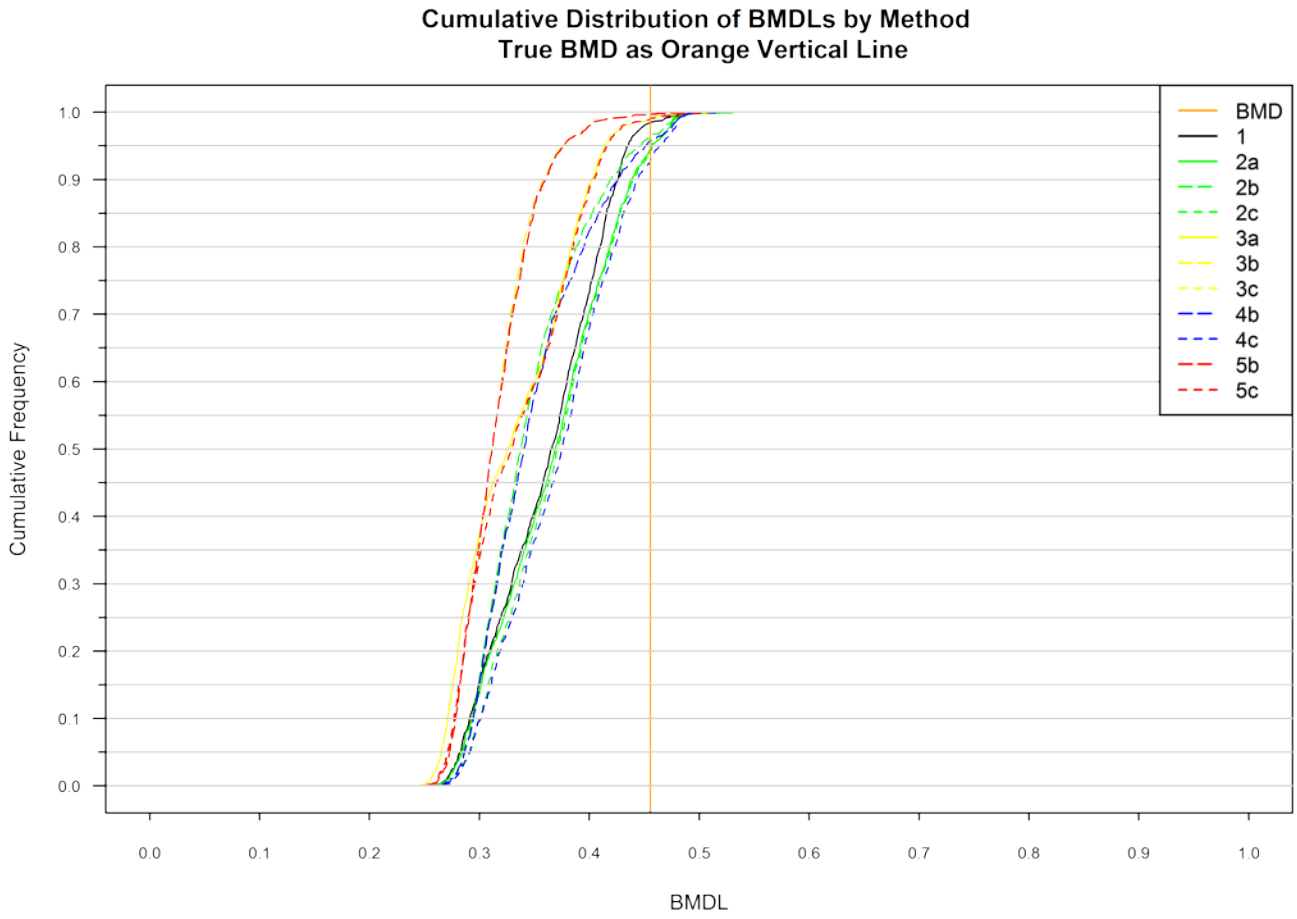
Figure 30: Template h3_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.149477	0.148269	
50	0.179932	0.177922	0.1688
75	0.22087	0.221705	
IQR	0.0713935	0.073436	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.966		
Poly3	0	**2a**	0.92	**3a**	0.987
Power	0	**2b**	0.907	**3b**	0.99
Hill	0.982	**2c**	0.911	**3c**	0.98
Exp3	0.488	**4b**	0.918	**5b**	0.995
Exp5	0.971	**4c**	0.92	**5c**	0.988

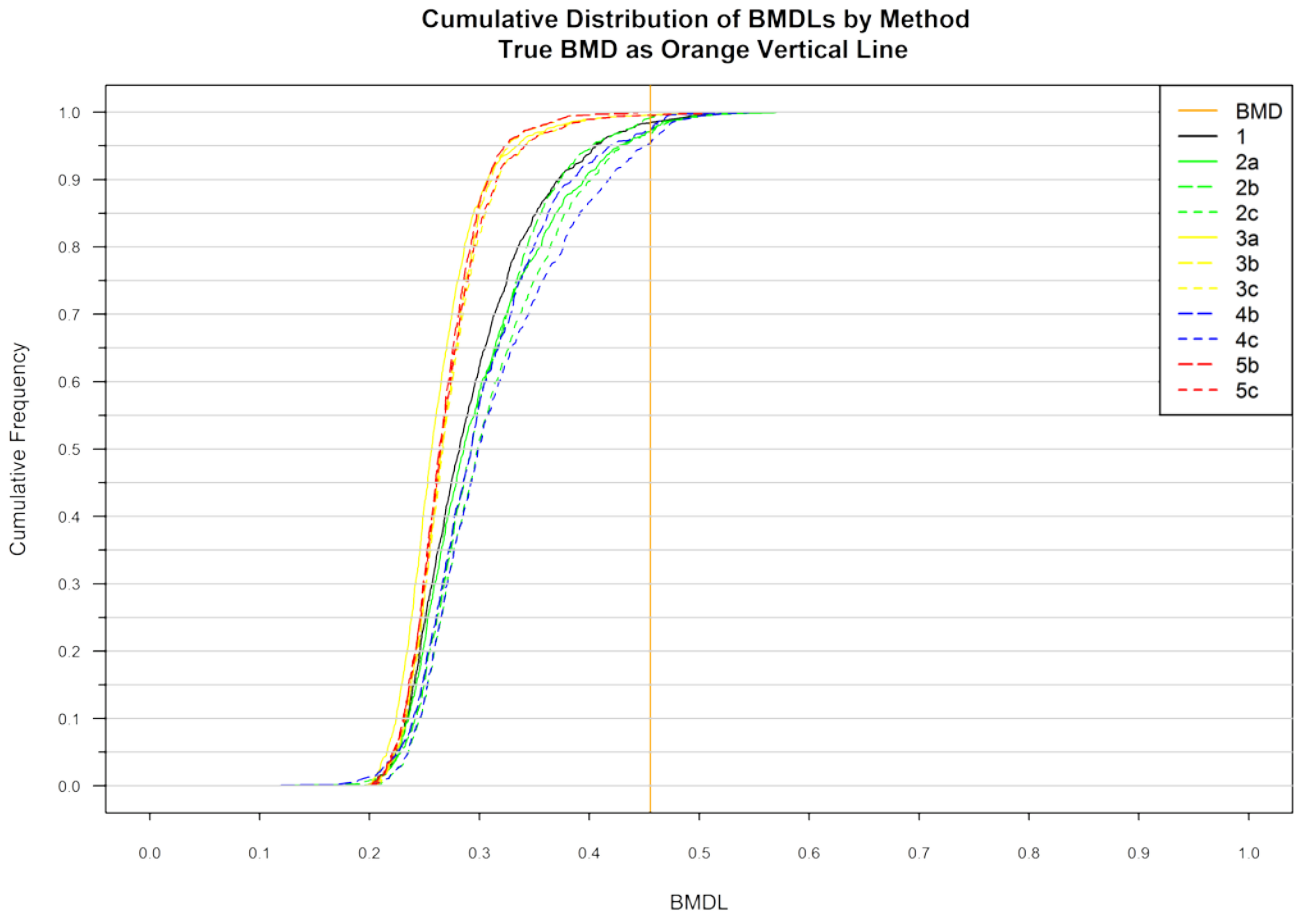
Figure 31: Template h4_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.38464	0.389975	
50	0.450407	0.456269	0.4556
75	0.481698	0.487047	
IQR	0.0970574	0.0970718	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.985		
Poly3	0.973	**2a**	0.948	**3a**	0.99
Power	0.985	**2b**	0.964	**3b**	0.996
Hill	0.979	**2c**	0.941	**3c**	0.99
Exp3	0.984	**4b**	0.958	**5b**	0.996
Exp5	0.974	**4c**	0.929	**5c**	0.989

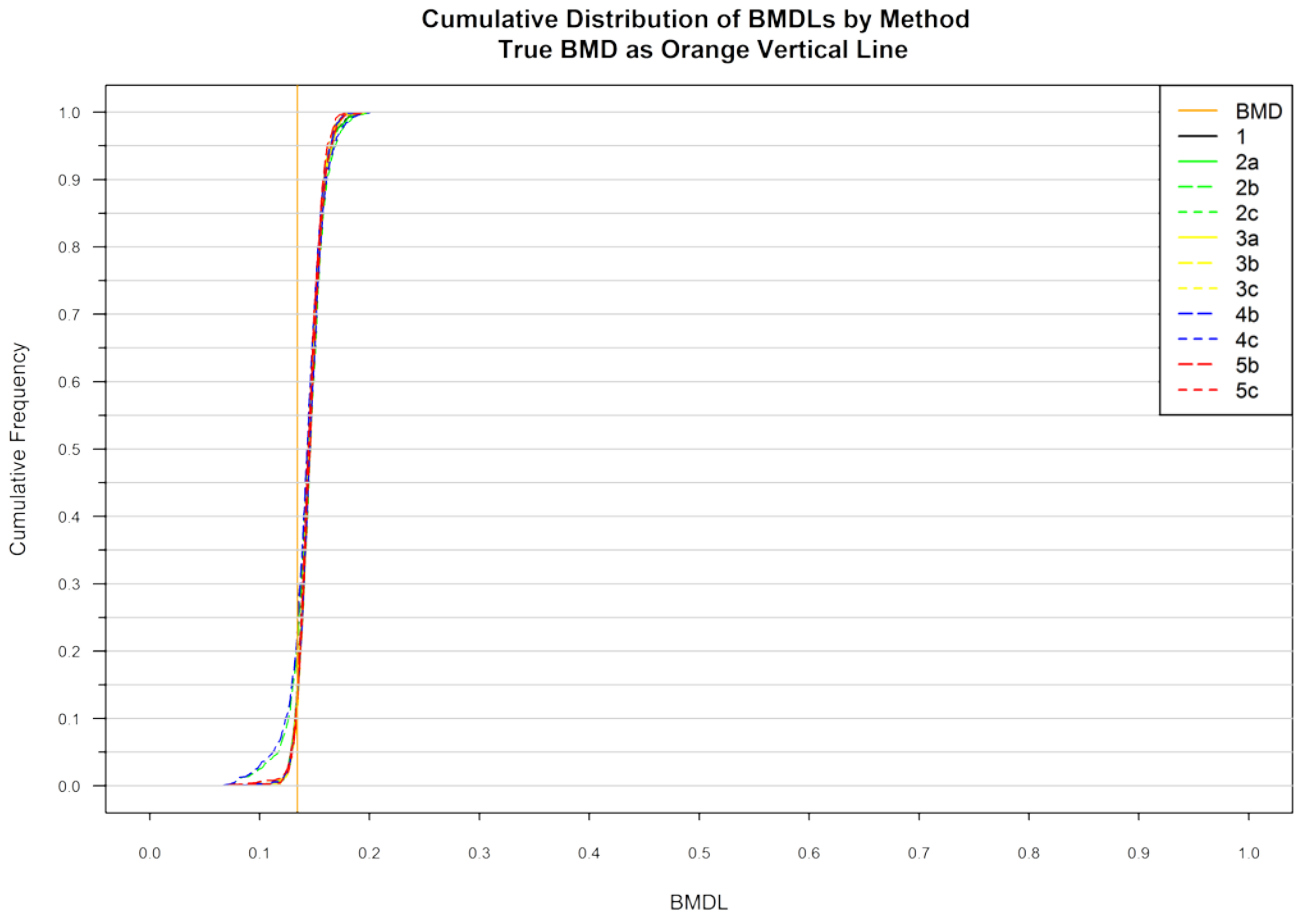
Figure 32: Template h4_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.333805	0.339512	
50	0.389566	0.406619	0.4556
75	0.455084	0.473136	
IQR	0.12128	0.133624	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.985		
Poly3	0.901	**2a**	0.972	**3a**	0.995
Power	0.98	**2b**	0.991	**3b**	1
Hill	0.975	**2c**	0.967	**3c**	0.995
Exp3	0.979	**4b**	0.975	**5b**	0.999
Exp5	0.975	**4c**	0.954	**5c**	0.995

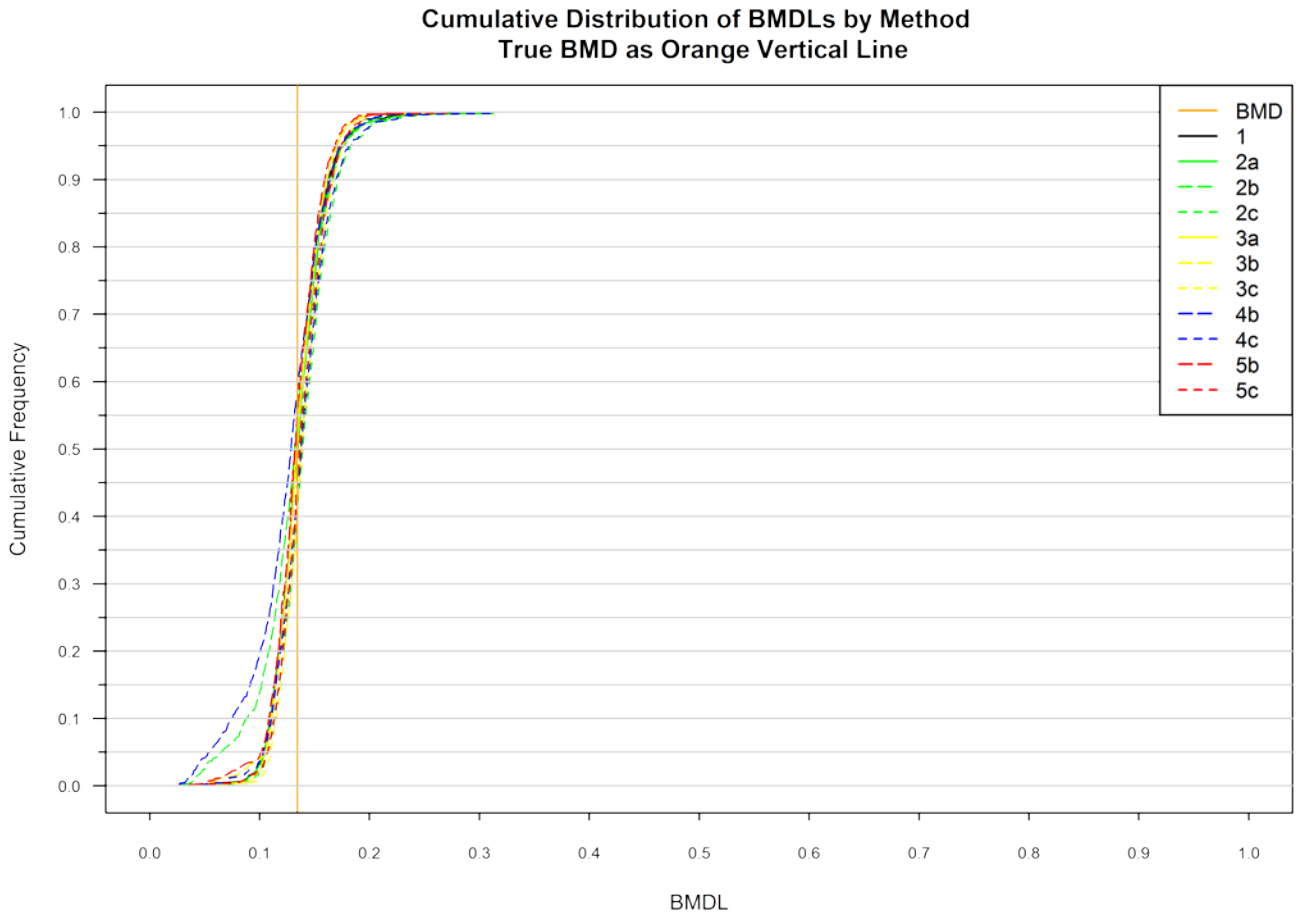
Figure 33: Template p1_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.154708	0.154453	
50	0.162718	0.162461	0.1345
75	0.172119	0.171801	
IQR	0.0174106	0.0173476	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.186	**1**	0.159		
Poly3	0.151	**2a**	0.155	**3a**	0.142
Power	0.156	**2b**	0.211	**3b**	0.14
Hill	0.57	**2c**	0.127	**3c**	0.113
Exp3	0	**4b**	0.239	**5b**	0.149
Exp5	0.684	**4c**	0.135	**5c**	0.122

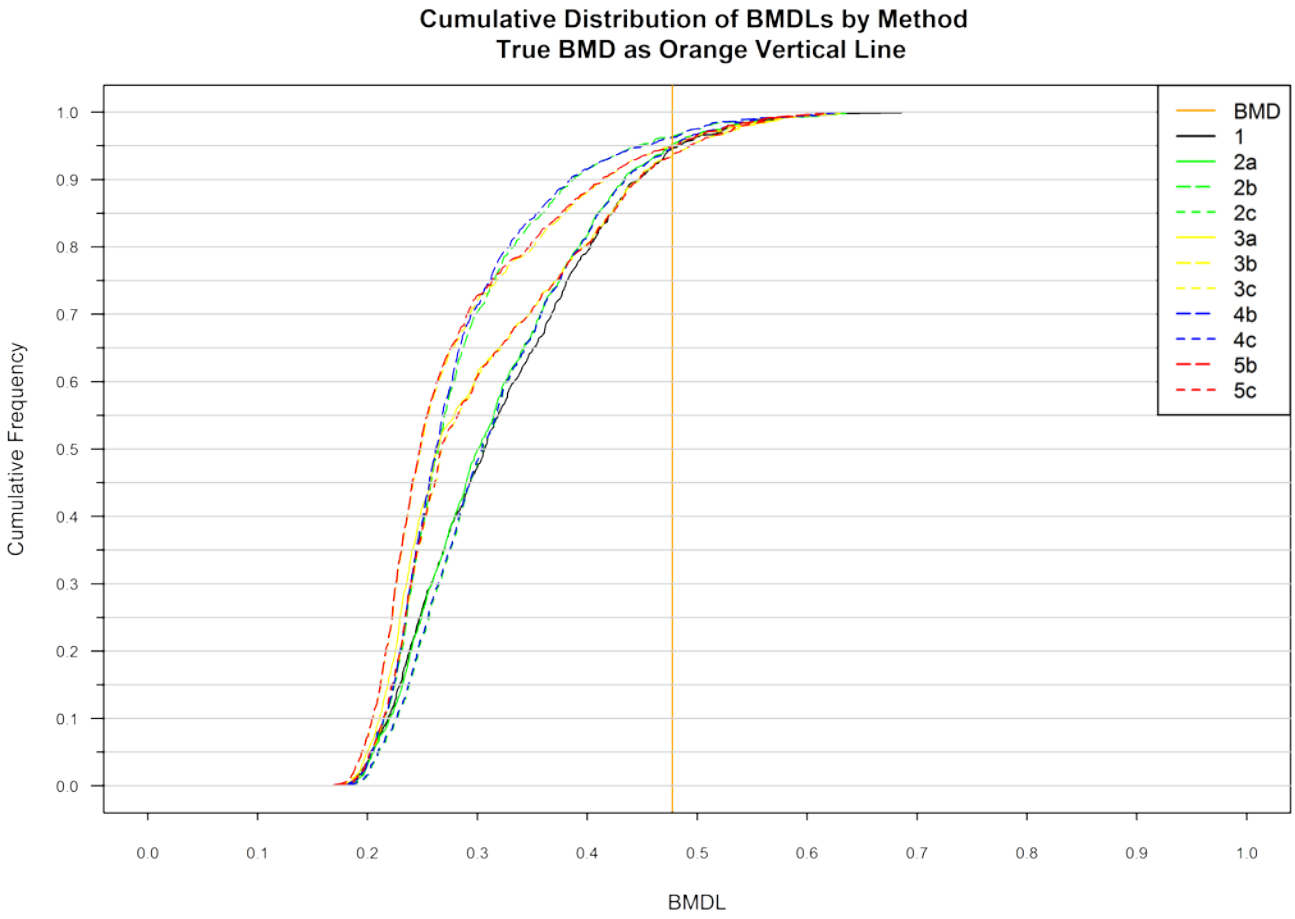
Figure 34: Template p1_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.154113	0.152289	
50	0.172775	0.171382	0.1345
75	0.200738	0.197847	
IQR	0.046625	0.045557	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.597	**1**	0.519		
Poly3	0.531	**2a**	0.493	**3a**	0.518
Power	0.538	**2b**	0.548	**3b**	0.506
Hill	0.714	**2c**	0.411	**3c**	0.418
Exp3	0.002	**4b**	0.589	**5b**	0.562
Exp5	0.655	**4c**	0.444	**5c**	0.454

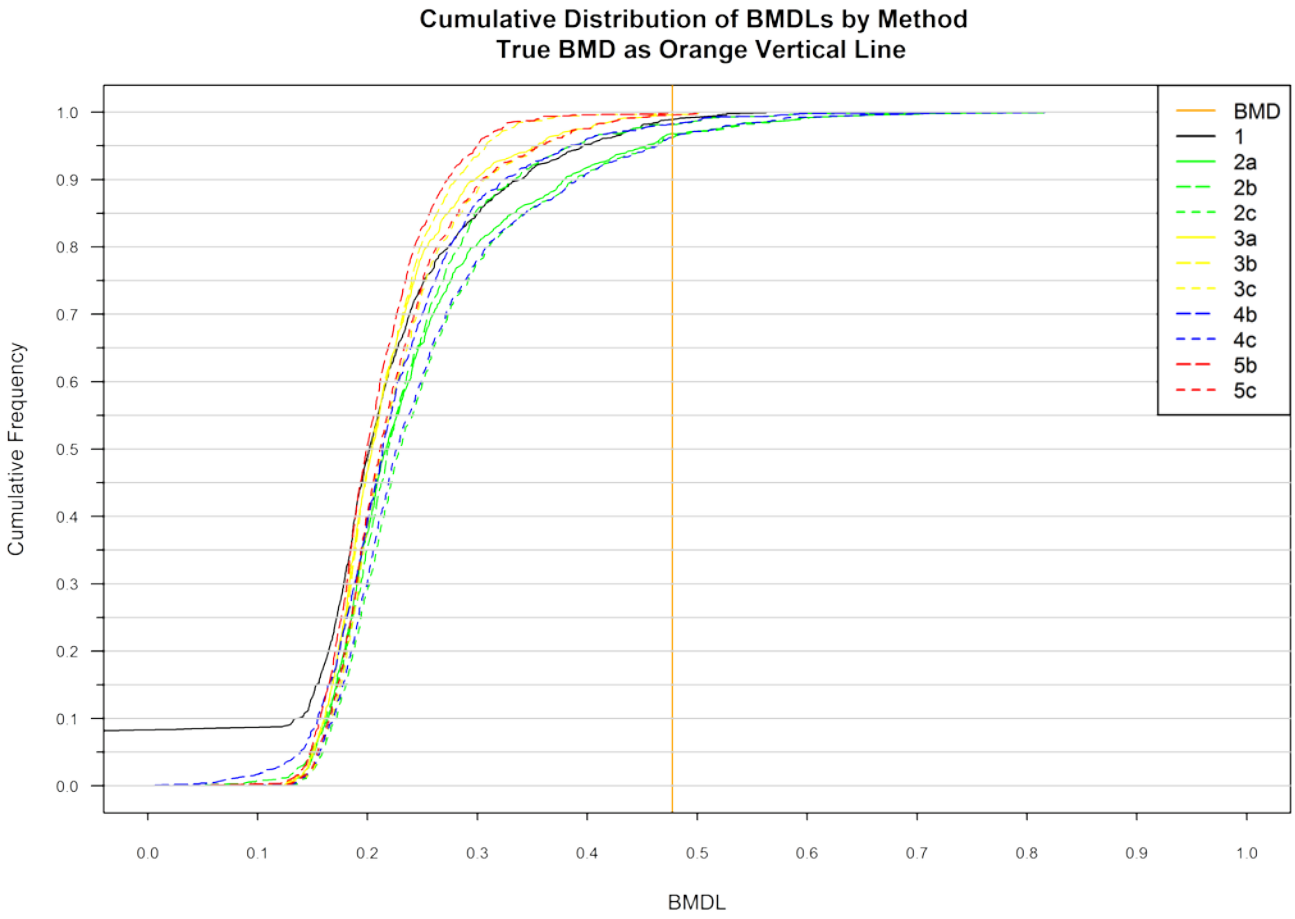
Figure 35: Template p2_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.35065	0.345455	
50	0.445874	0.444634	0.4775
75	0.542072	0.539971	
IQR	0.191422	0.194516	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.945		
Poly3	0.935	**2a**	0.953	**3a**	0.937
Power	0.935	**2b**	0.963	**3b**	0.949
Hill	0.933	**2c**	0.952	**3c**	0.936
Exp3	0.944	**4b**	0.962	**5b**	0.948
Exp5	0.943	**4c**	0.951	**5c**	0.935

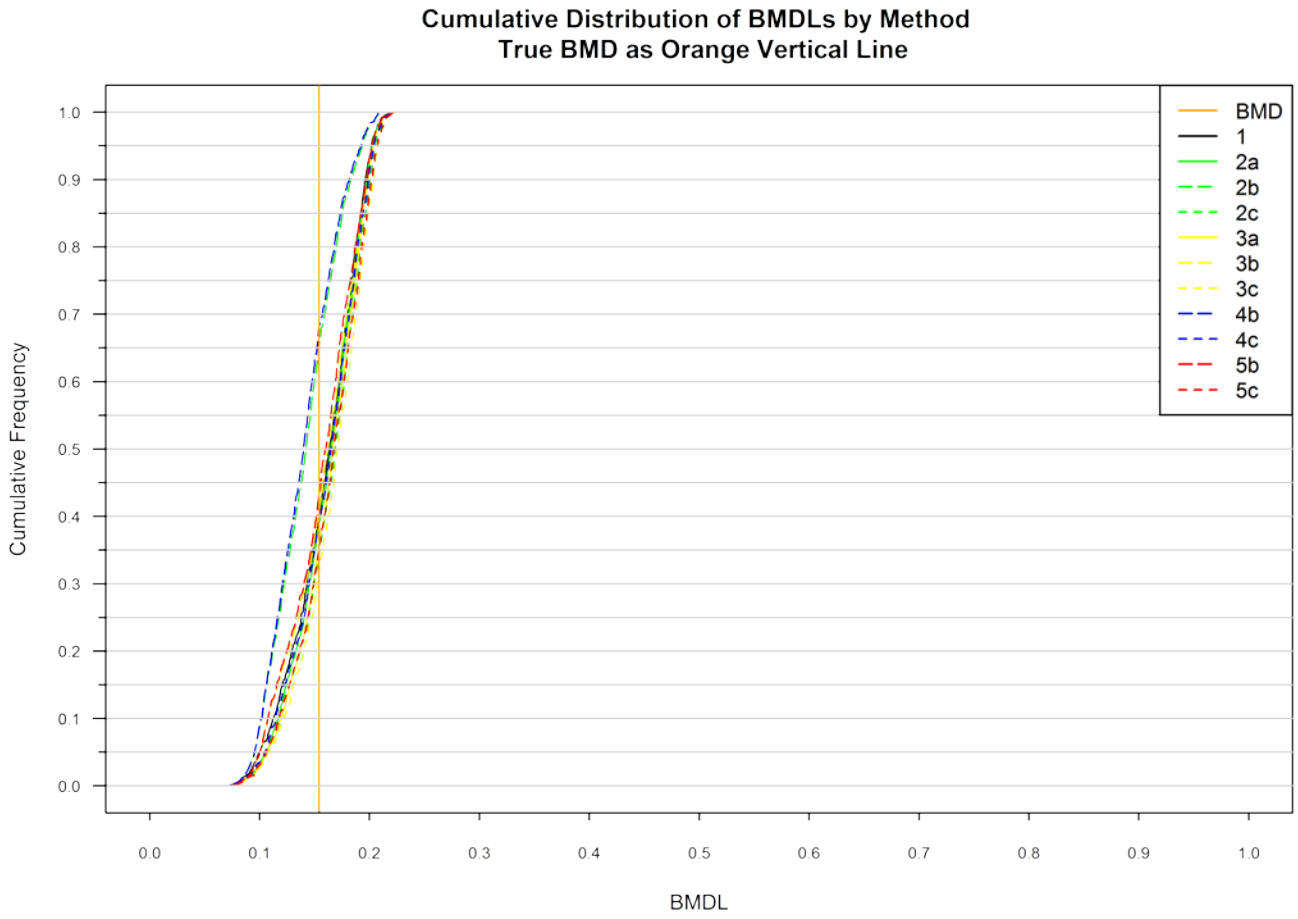
Figure 36: Template p2_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.28451	0.280504	
50	0.374354	0.369563	0.4775
75	0.537319	0.534884	
IQR	0.252809	0.25438	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.989		
Poly3	0.937	**2a**	0.968	**3a**	0.996
Power	0.928	**2b**	0.981	**3b**	0.999
Hill	0.932	**2c**	0.962	**3c**	0.996
Exp3	0.978	**4b**	0.983	**5b**	0.999
Exp5	0.981	**4c**	0.962	**5c**	0.996

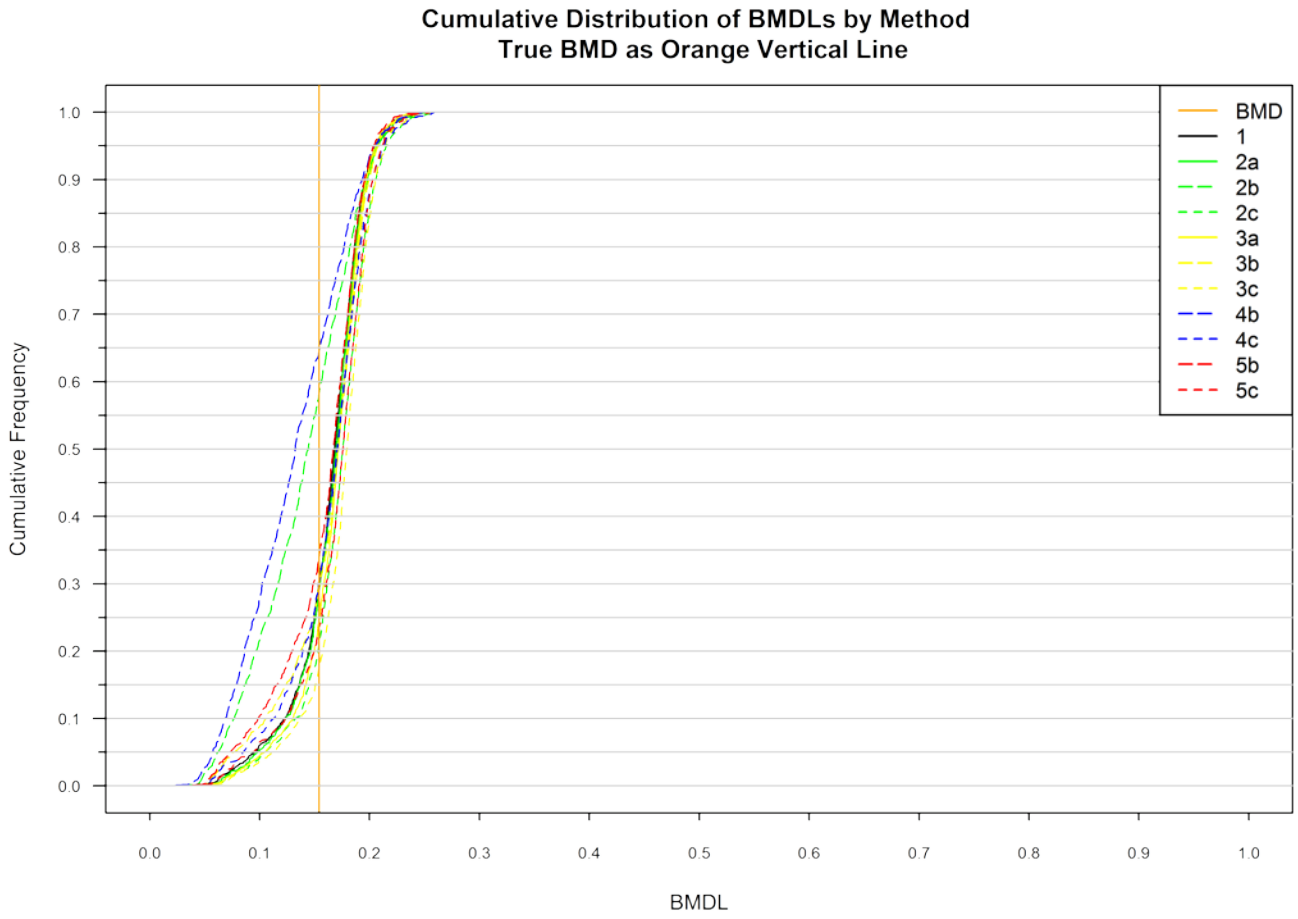
Figure 37: Template p3_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.168565	0.166723	
50	0.190933	0.188814	0.1541
75	0.206943	0.205893	
IQR	0.0383781	0.0391697	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.405		
Poly3	0	**2a**	0.39	**3a**	0.356
Power	0	**2b**	0.65	**3b**	0.41
Hill	0.879	**2c**	0.344	**3c**	0.318
Exp3	0.455	**4b**	0.682	**5b**	0.432
Exp5	0.874	**4c**	0.393	**5c**	0.355

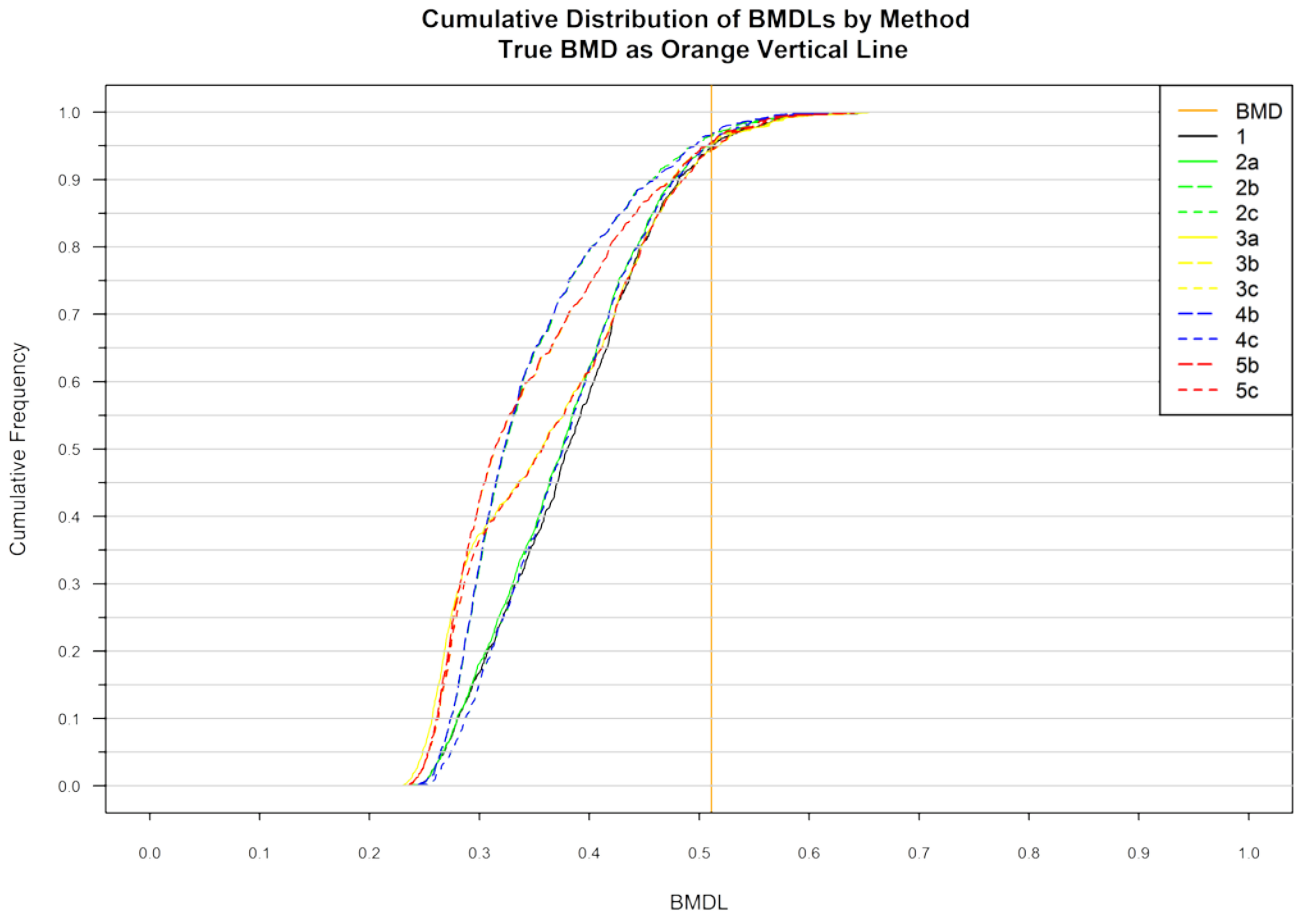
Figure 38: Template p3_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.186569	0.180091	
50	0.207125	0.20448	0.1541
75	0.229969	0.227928	
IQR	0.0433999	0.0478363	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.007	**1**	0.305		
Poly3	0.007	**2a**	0.291	**3a**	0.244
Power	0.007	**2b**	0.582	**3b**	0.296
Hill	0.85	**2c**	0.214	**3c**	0.176
Exp3	0.739	**4b**	0.646	**5b**	0.341
Exp5	0.888	**4c**	0.304	**5c**	0.234

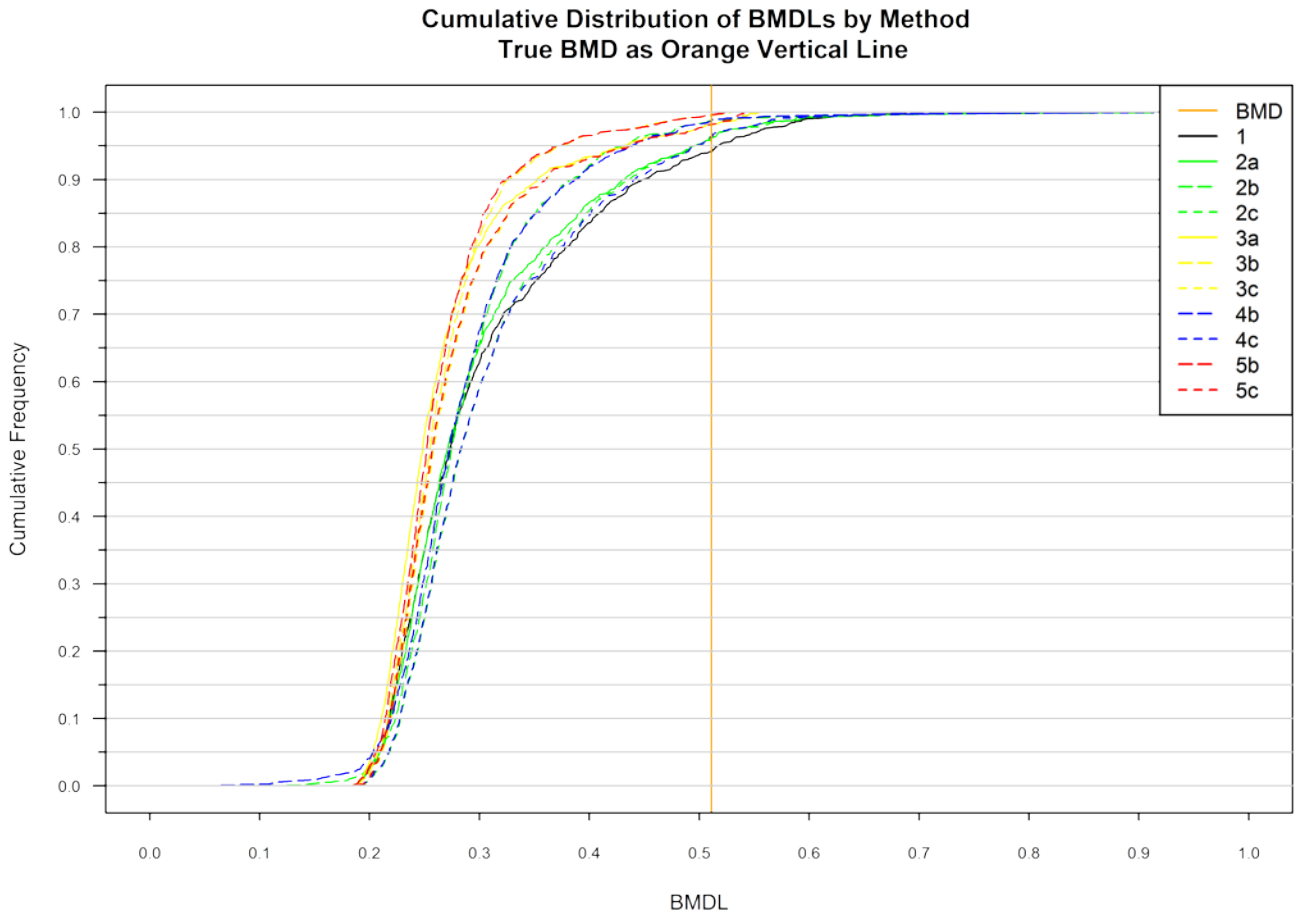
Figure 39: Template p4_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.41406	0.413191	
50	0.493123	0.493853	0.5112
75	0.557354	0.555246	
IQR	0.143294	0.142055	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.946		
Poly3	0.937	**2a**	0.951	**3a**	0.944
Power	0.946	**2b**	0.964	**3b**	0.952
Hill	0.957	**2c**	0.949	**3c**	0.944
Exp3	0.945	**4b**	0.968	**5b**	0.955
Exp5	0.956	**4c**	0.956	**5c**	0.944

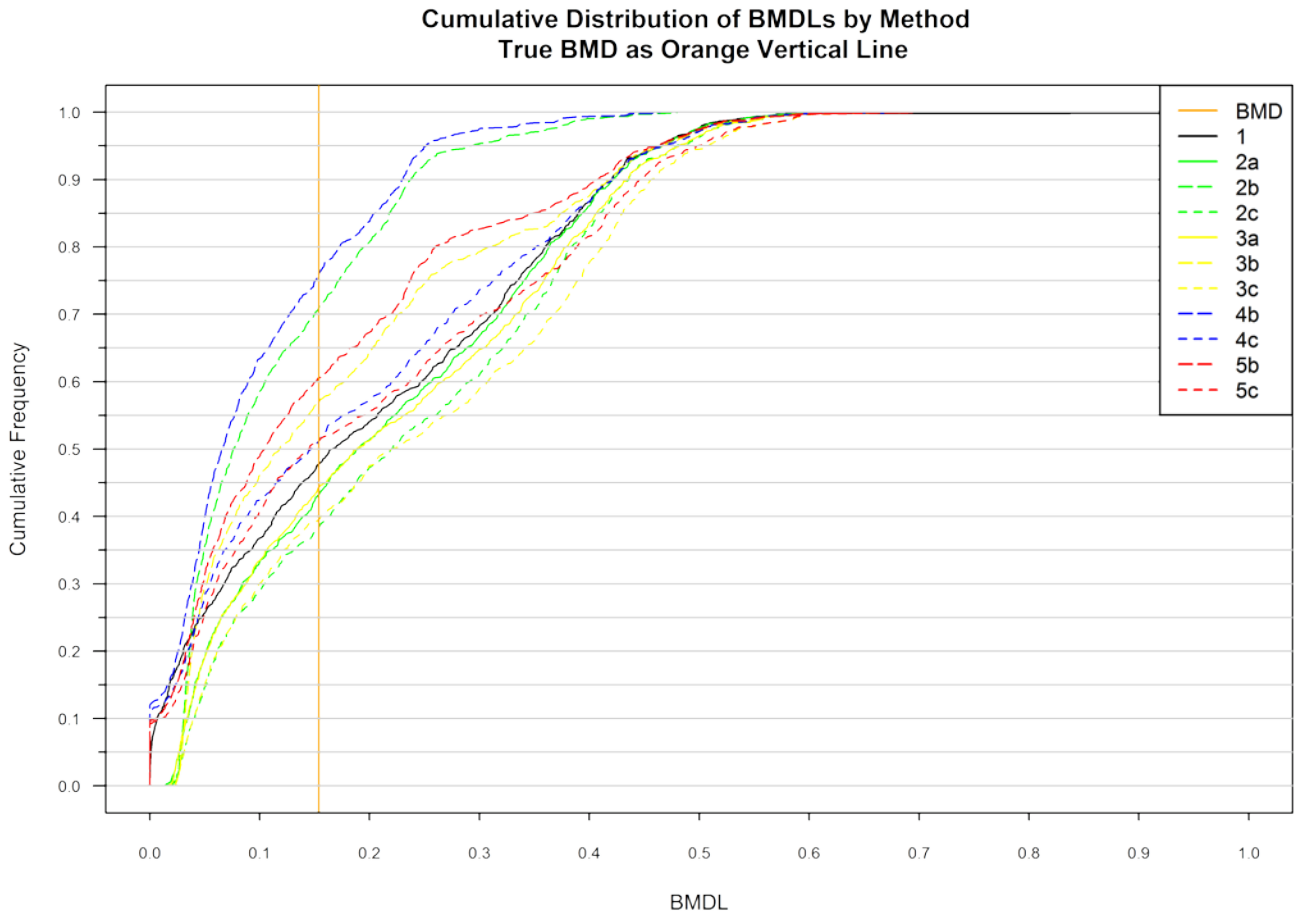
Figure 40: Template p4_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.33768	0.337994	
50	0.42041	0.424875	0.5112
75	0.556208	0.54714	
IQR	0.218528	0.209145	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.941		
Poly3	0.937	**2a**	0.963	**3a**	0.981
Power	0.933	**2b**	0.987	**3b**	0.995
Hill	0.941	**2c**	0.961	**3c**	0.981
Exp3	0.931	**4b**	0.988	**5b**	0.996
Exp5	0.936	**4c**	0.964	**5c**	0.982

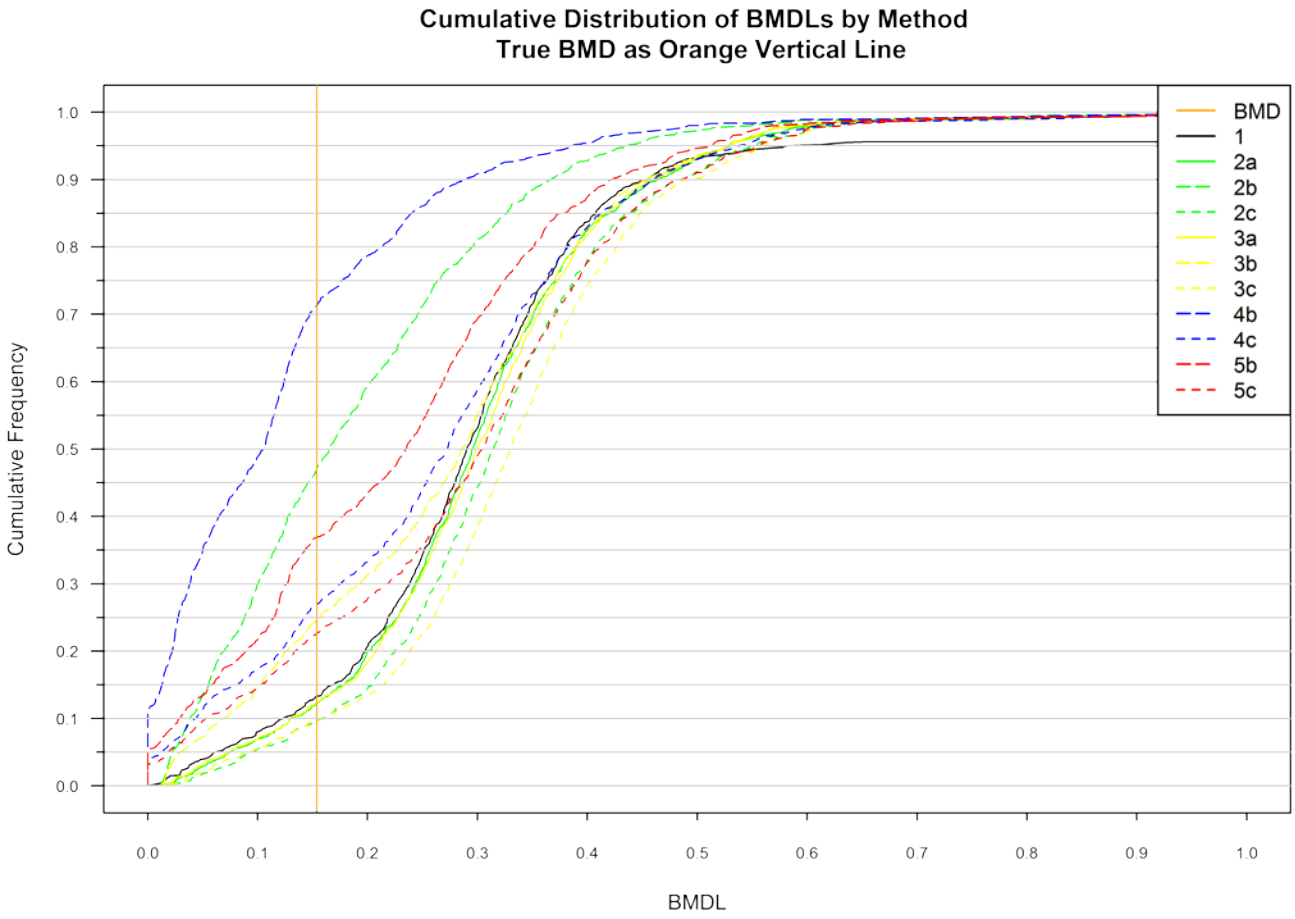
Figure 41: Template e1_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.209167	0.197894	
50	0.297258	0.246475	0.154
75	0.480379	0.454505	
IQR	0.271212	0.256611	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.477		
Poly3	0	**2a**	0.432	**3a**	0.444
Power	0	**2b**	0.71	**3b**	0.57
Hill	0.917	**2c**	0.385	**3c**	0.396
Exp3	0	**4b**	0.76	**5b**	0.606
Exp5	0.905	**4c**	0.513	**5c**	0.512

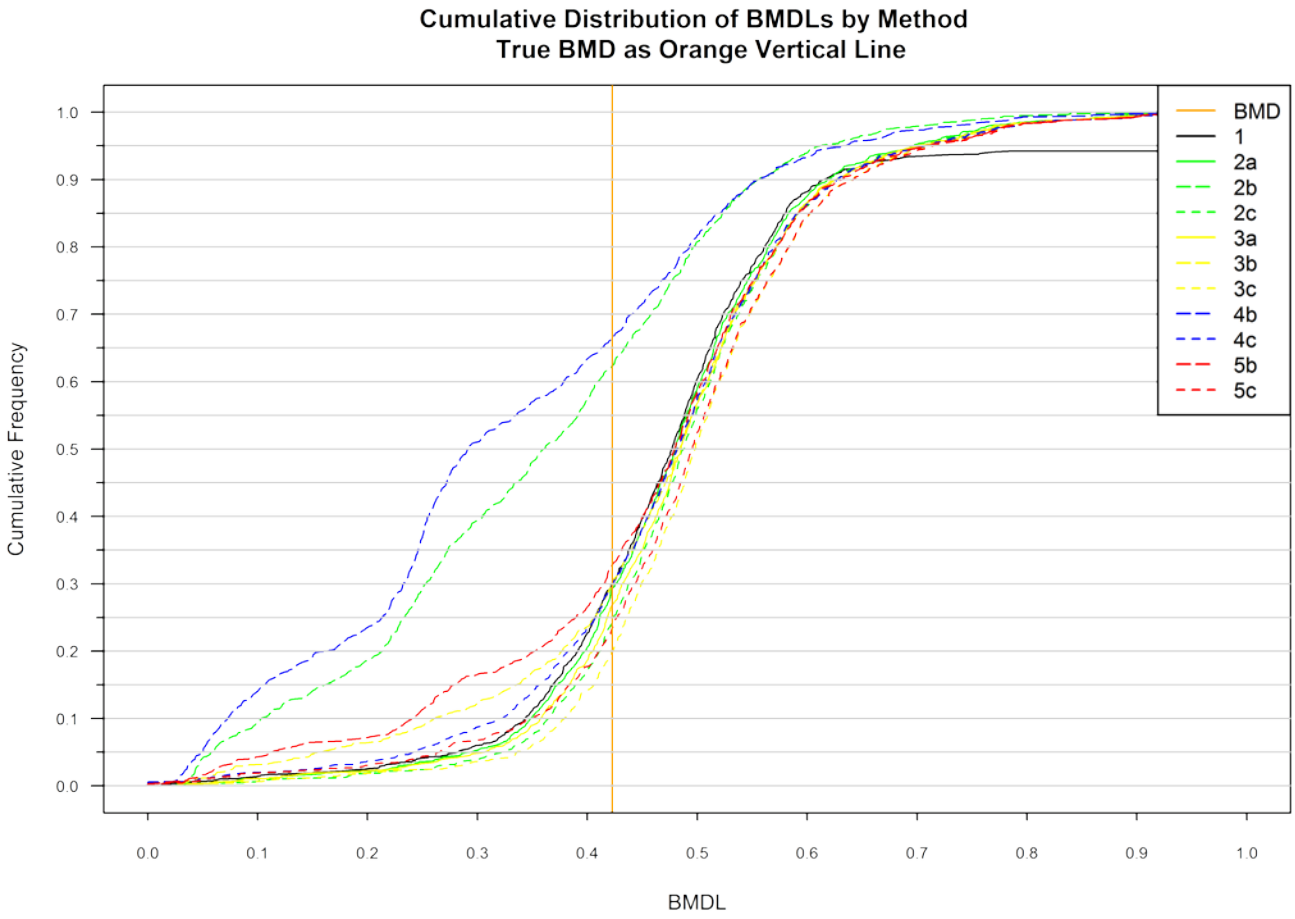
Figure 42: Template e1_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.366967	0.270363	
50	0.511295	0.474799	0.154
75	0.707826	0.708172	
IQR	0.340859	0.437809	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.134		
Poly3	0	**2a**	0.122	**3a**	0.123
Power	0.002	**2b**	0.472	**3b**	0.247
Hill	0.857	**2c**	0.096	**3c**	0.097
Exp3	0.009	**4b**	0.713	**5b**	0.369
Exp5	0.912	**4c**	0.27	**5c**	0.227

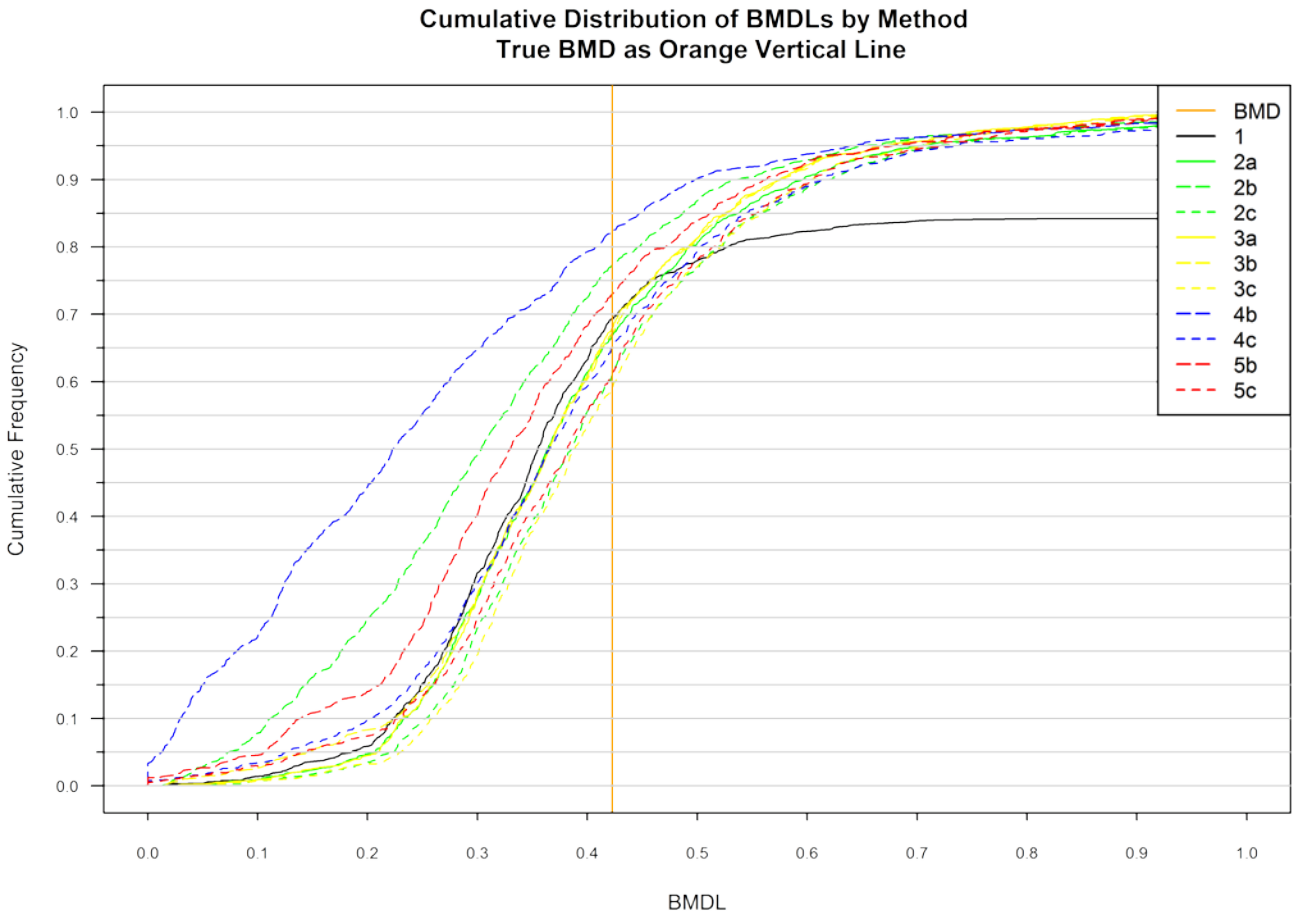
Figure 43: Template e2_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.569655	0.567714	
50	0.683129	0.689622	0.4225
75	0.813988	0.835391	
IQR	0.244333	0.267678	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.146	**1**	0.301		
Poly3	0.145	**2a**	0.294	**3a**	0.268
Power	0.146	**2b**	0.623	**3b**	0.302
Hill	0.891	**2c**	0.243	**3c**	0.201
Exp3	0.065	**4b**	0.665	**5b**	0.328
Exp5	0.947	**4c**	0.297	**5c**	0.234

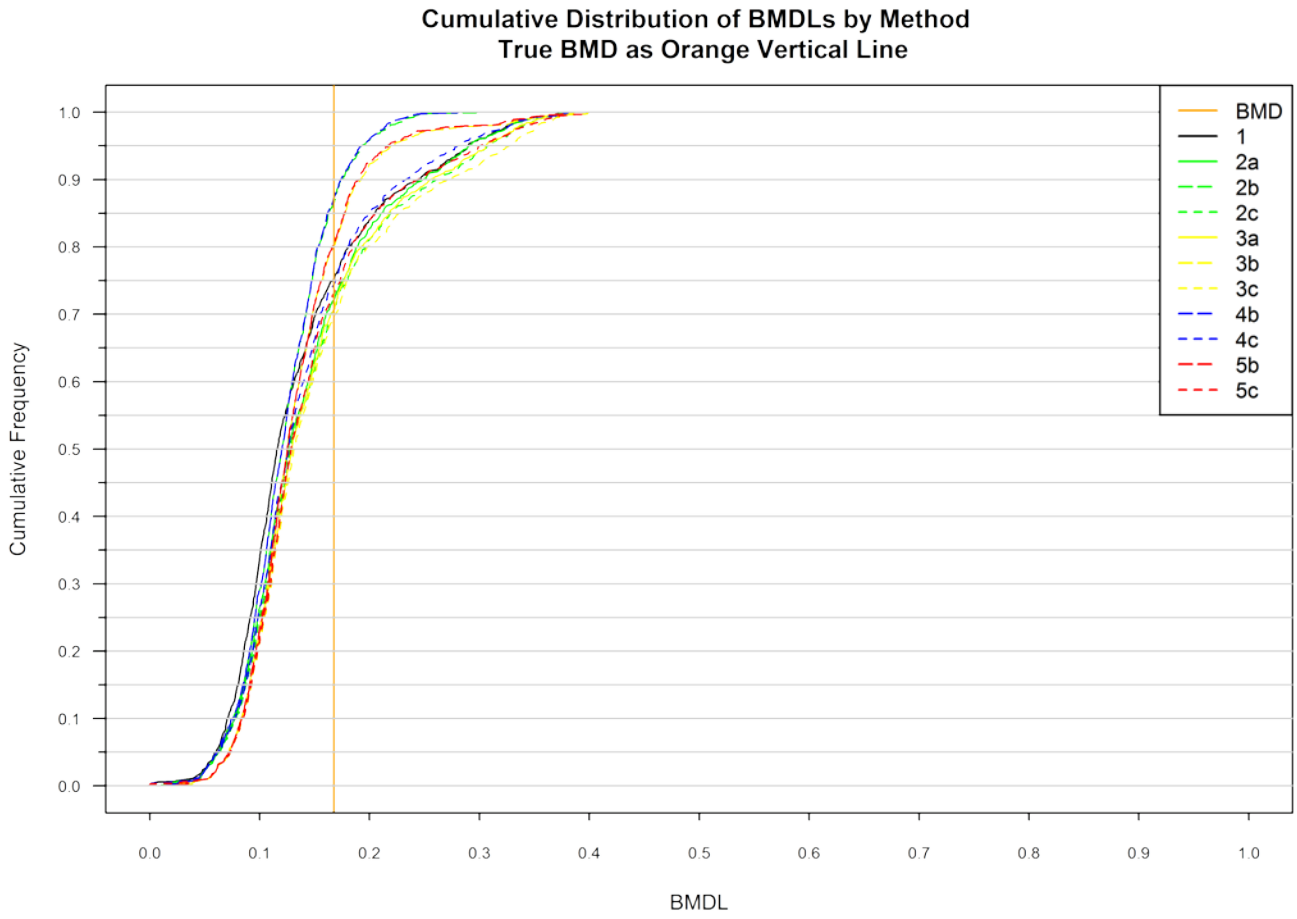
Figure 44: Template e2_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.486051	0.468908	
50	0.693794	0.704292	0.4225
75	1.10935	1.06368	
IQR	0.623302	0.594777	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.635	**1**	0.693		
Poly3	0.586	**2a**	0.669	**3a**	0.68
Power	0.619	**2b**	0.775	**3b**	0.672
Hill	0.78	**2c**	0.611	**3c**	0.591
Exp3	0.548	**4b**	0.823	**5b**	0.729
Exp5	0.931	**4c**	0.65	**5c**	0.613

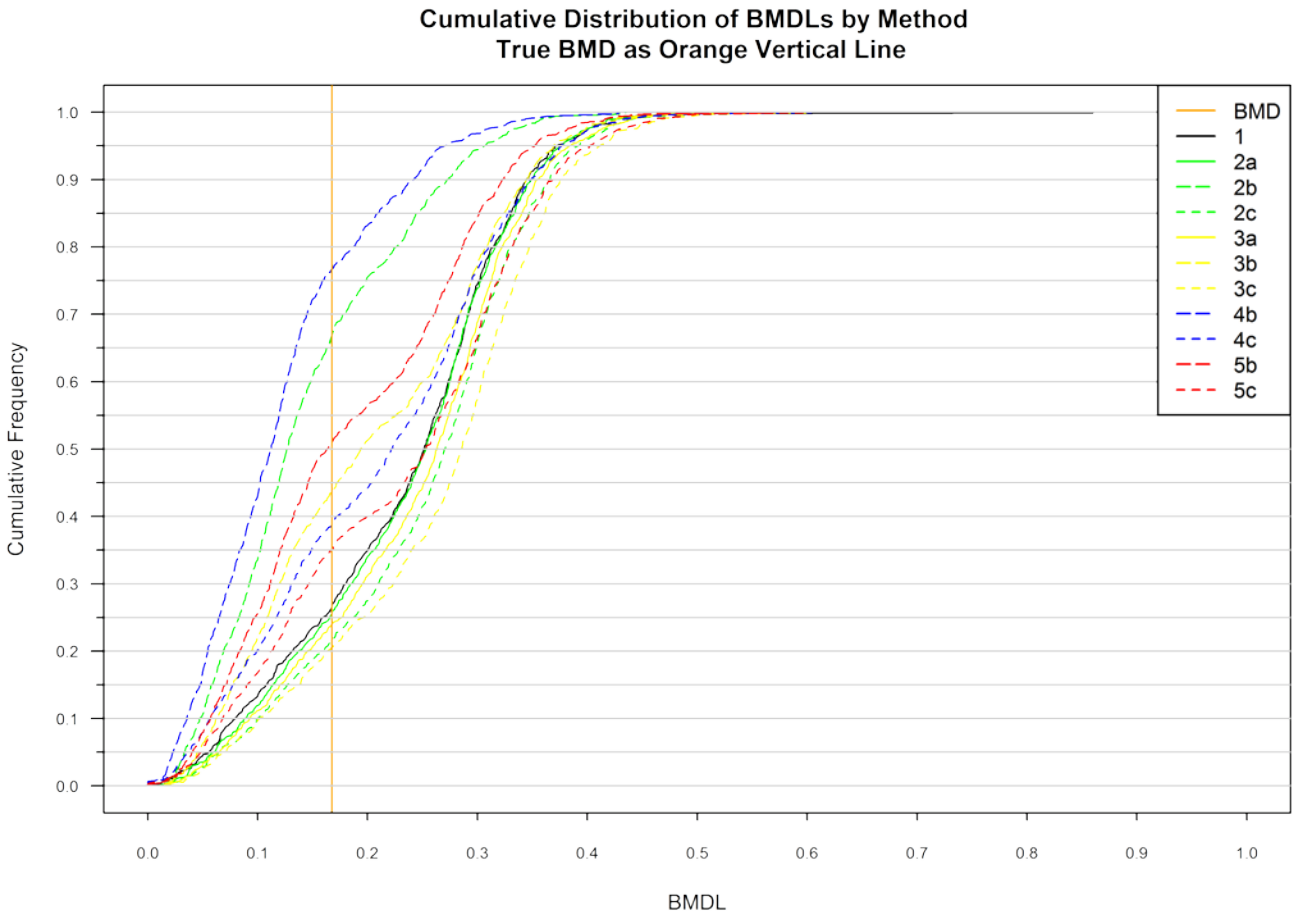
Figure 45: Template e3_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.156669	0.155976	
50	0.193499	0.191916	0.1675
75	0.238655	0.233416	
IQR	0.081986	0.0774399	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.752		
Poly3	0	**2a**	0.722	**3a**	0.707
Power	0	**2b**	0.865	**3b**	0.802
Hill	0.925	**2c**	0.709	**3c**	0.696
Exp3	0	**4b**	0.873	**5b**	0.804
Exp5	0.926	**4c**	0.744	**5c**	0.731

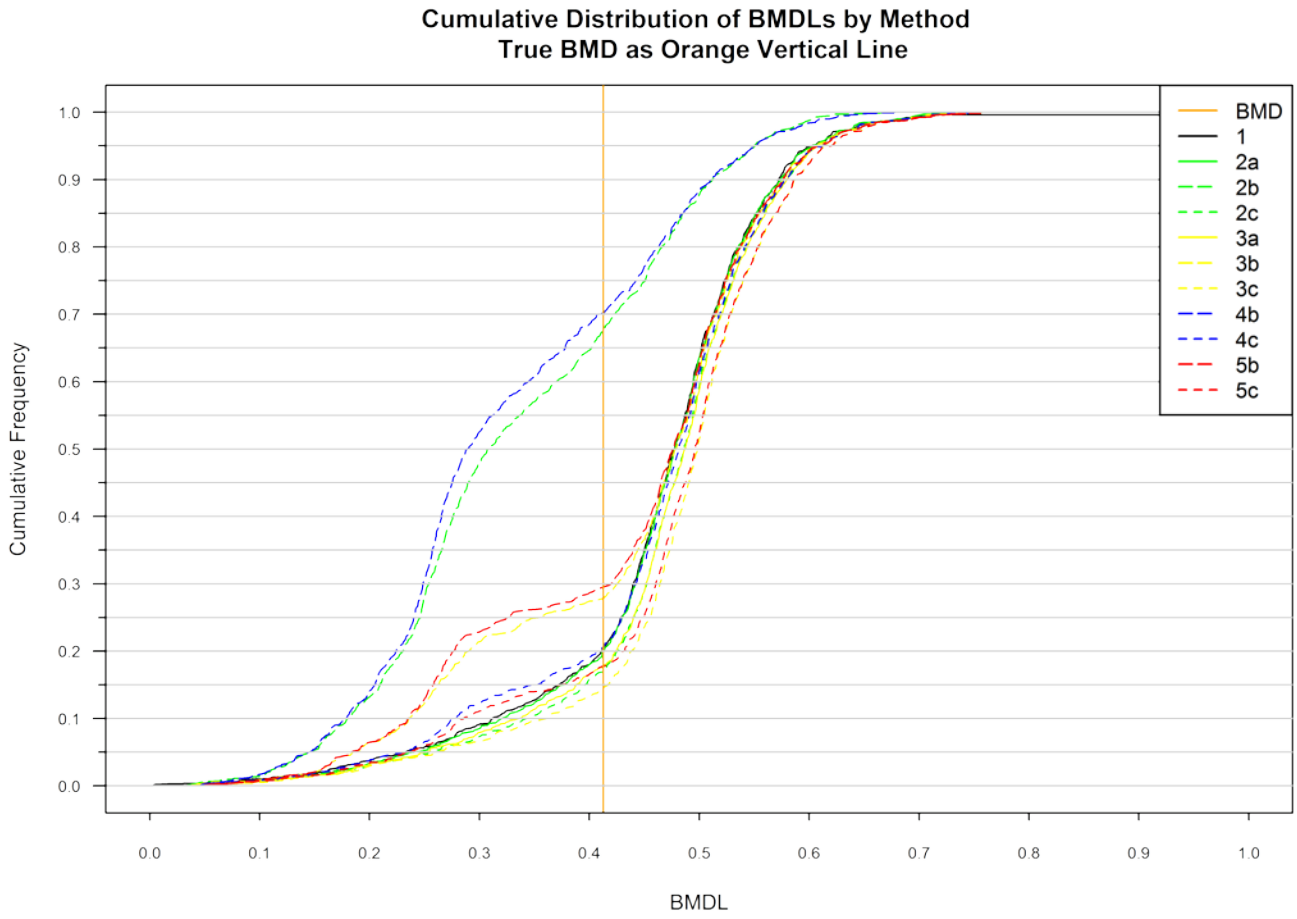
Figure 46: Template e3_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.248589	0.193359	
50	0.343584	0.309303	0.1675
75	0.419891	0.402938	
IQR	0.171302	0.20958	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.267		
Poly3	0	**2a**	0.257	**3a**	0.241
Power	0	**2b**	0.668	**3b**	0.438
Hill	0.899	**2c**	0.219	**3c**	0.206
Exp3	0	**4b**	0.766	**5b**	0.511
Exp5	0.889	**4c**	0.388	**5c**	0.351

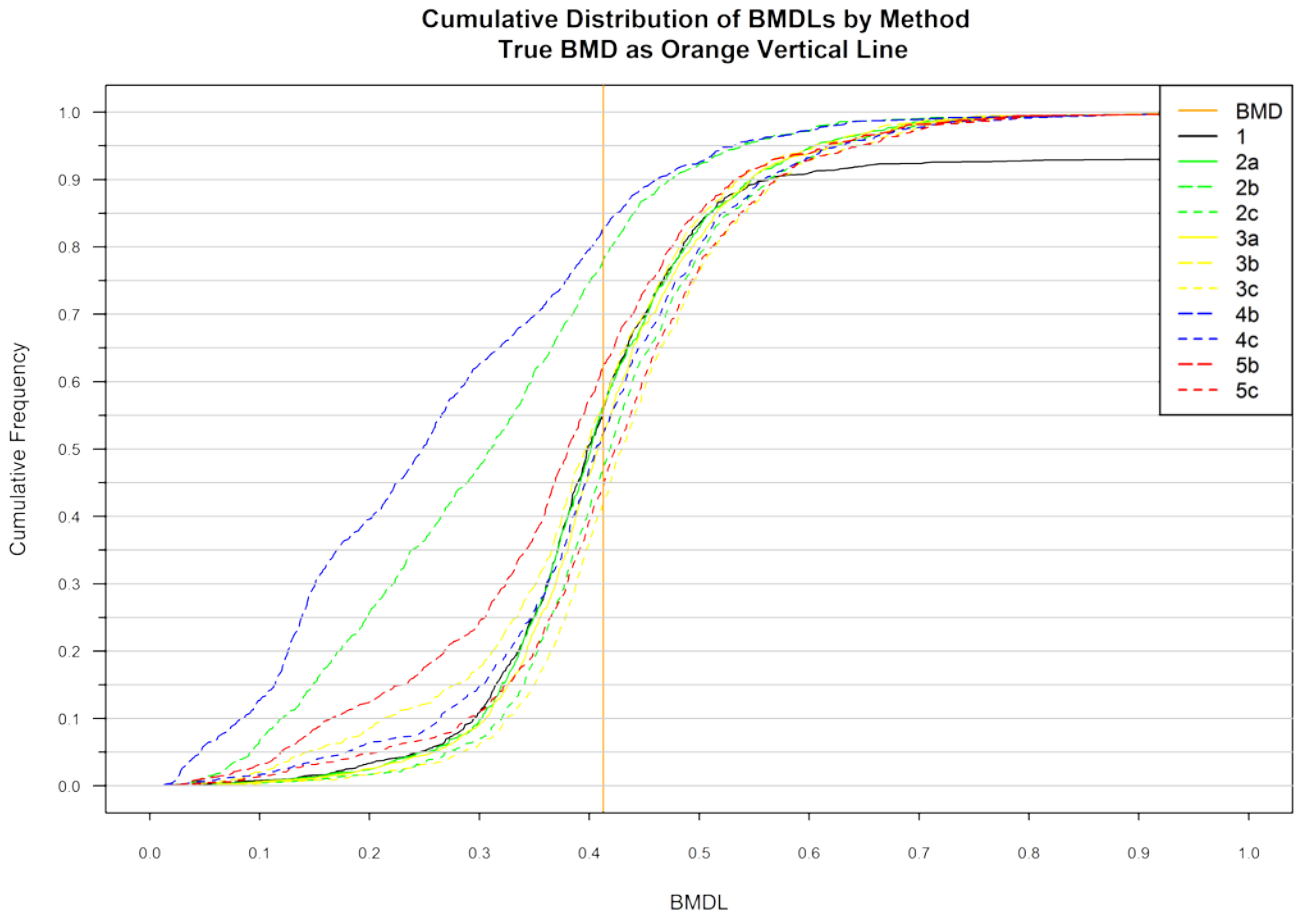
Figure 47: Template e4_normal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.522717	0.517259	
50	0.585215	0.585866	0.4126
75	0.65128	0.654846	
IQR	0.128563	0.137587	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.004	**1**	0.203		
Poly3	0.004	**2a**	0.201	**3a**	0.176
Power	0.004	**2b**	0.677	**3b**	0.279
Hill	0.949	**2c**	0.17	**3c**	0.144
Exp3	0.08	**4b**	0.702	**5b**	0.296
Exp5	0.943	**4c**	0.206	**5c**	0.179

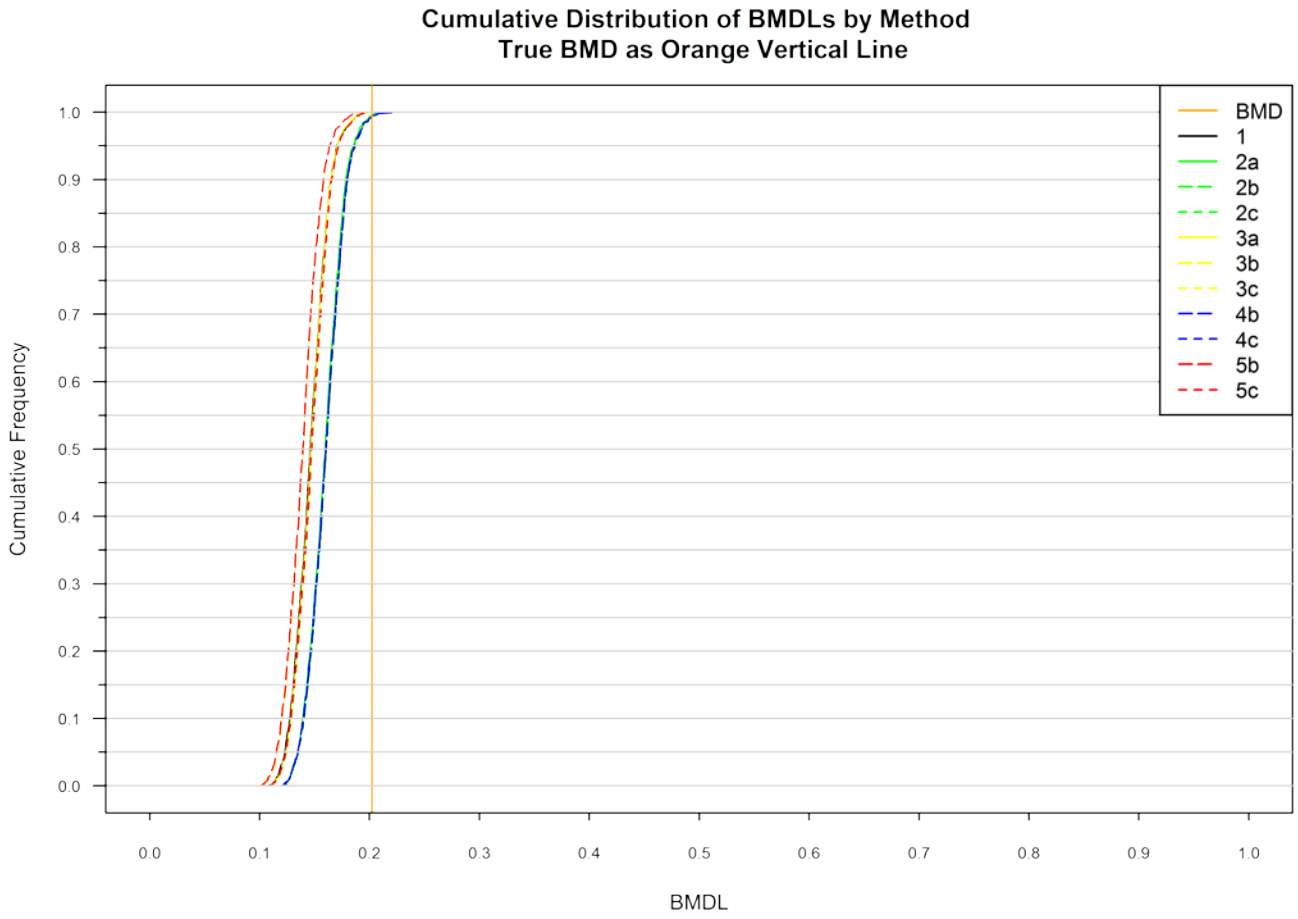
Figure 48: Template e4_normal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.496564	0.489506	
50	0.596	0.599969	0.4126
75	0.74221	0.766374	
IQR	0.245647	0.276868	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.39	**1**	0.555		
Poly3	0.372	**2a**	0.564	**3a**	0.521
Power	0.372	**2b**	0.781	**3b**	0.559
Hill	0.856	**2c**	0.47	**3c**	0.419
Exp3	0.521	**4b**	0.827	**5b**	0.624
Exp5	0.935	**4c**	0.523	**5c**	0.445

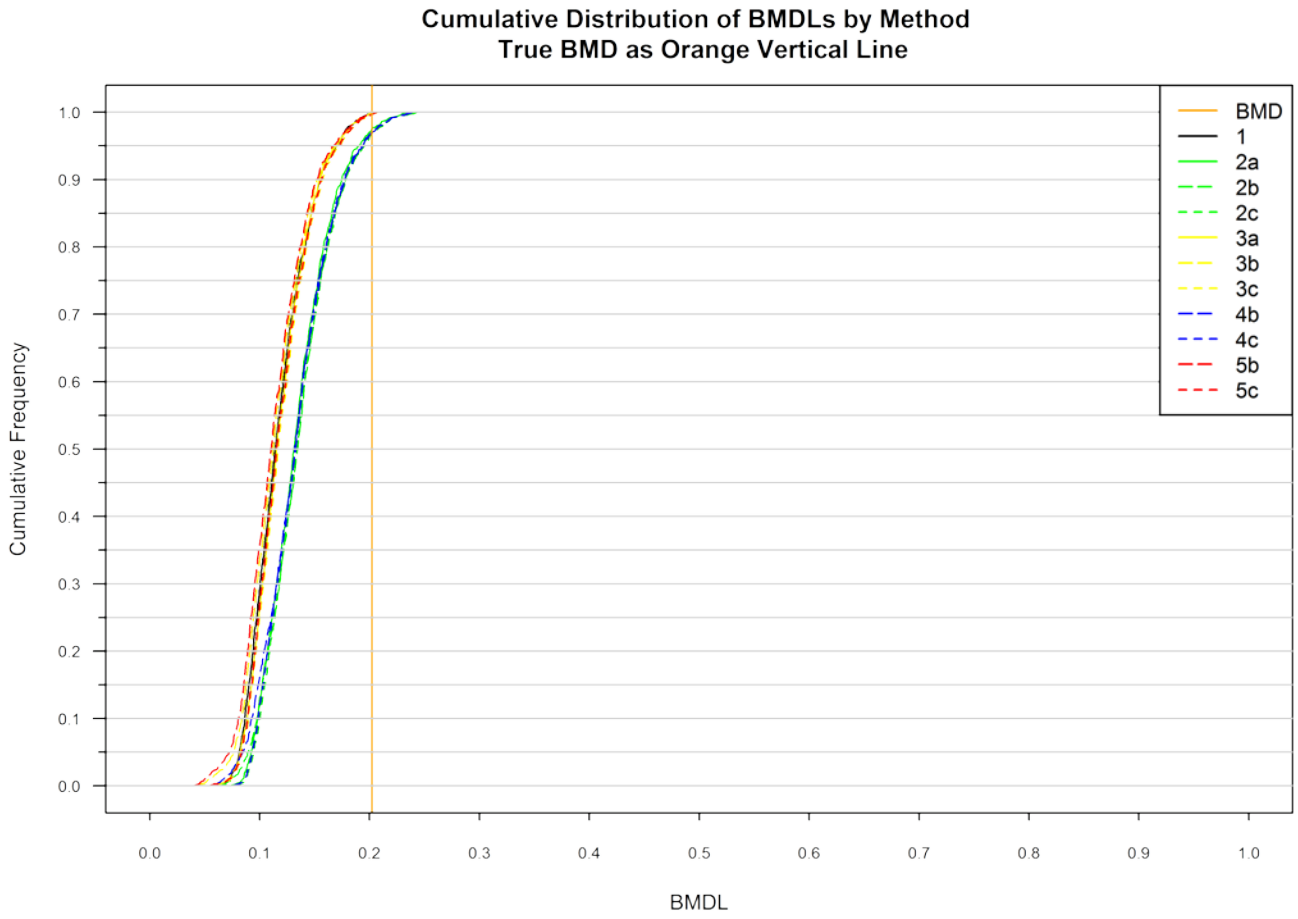
Figure 49: Template w1_lognormal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.174468	0.17472	
50	0.186791	0.186988	0.2021
75	0.19833	0.198651	
IQR	0.0238619	0.0239311	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.999		
Poly3	0.982	**2a**	0.995	**3a**	1
Power	0.988	**2b**	0.992	**3b**	1
Hill	0.985	**2c**	0.993	**3c**	0.999
Exp3	1	**4b**	0.992	**5b**	1
Exp5	0.989	**4c**	0.993	**5c**	0.999

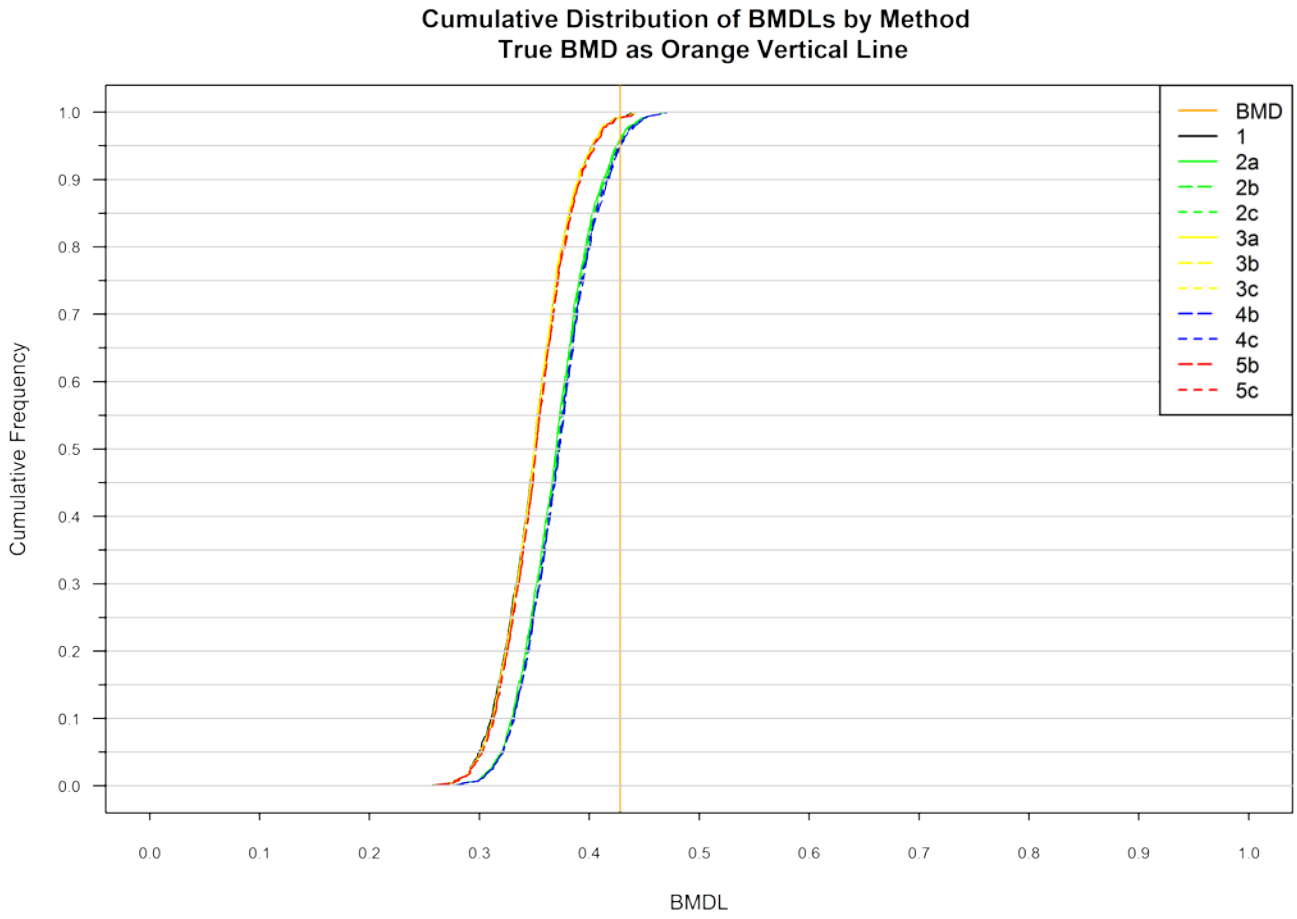
Figure 50: Template w1_lognormal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.161364	0.16167	
50	0.186344	0.187965	0.2021
75	0.212934	0.213913	
IQR	0.0515702	0.0522433	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.998		
Poly3	0.953	**2a**	0.975	**3a**	0.998
Power	0.971	**2b**	0.969	**3b**	0.997
Hill	0.962	**2c**	0.968	**3c**	0.996
Exp3	1	**4b**	0.969	**5b**	0.997
Exp5	0.979	**4c**	0.971	**5c**	0.998

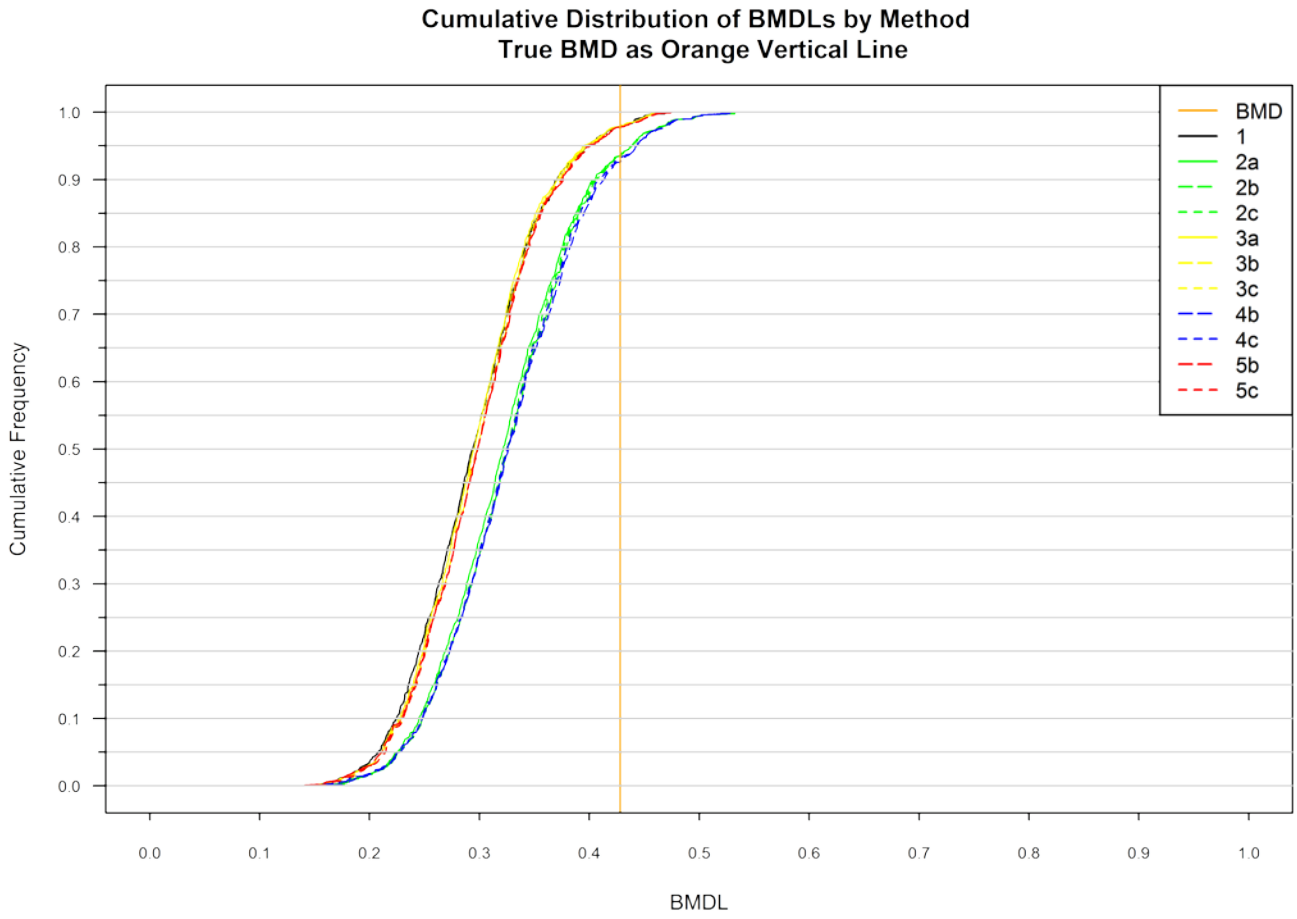
Figure 51: Template w2_lognormal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.399056	0.401075	
50	0.423594	0.42567	0.4281
75	0.445144	0.448073	
IQR	0.0460885	0.0469979	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.993		
Poly3	1	**2a**	0.96	**3a**	0.992
Power	0.989	**2b**	0.958	**3b**	0.991
Hill	0.986	**2c**	0.959	**3c**	0.992
Exp3	0.996	**4b**	0.95	**5b**	0.991
Exp5	0.994	**4c**	0.954	**5c**	0.991

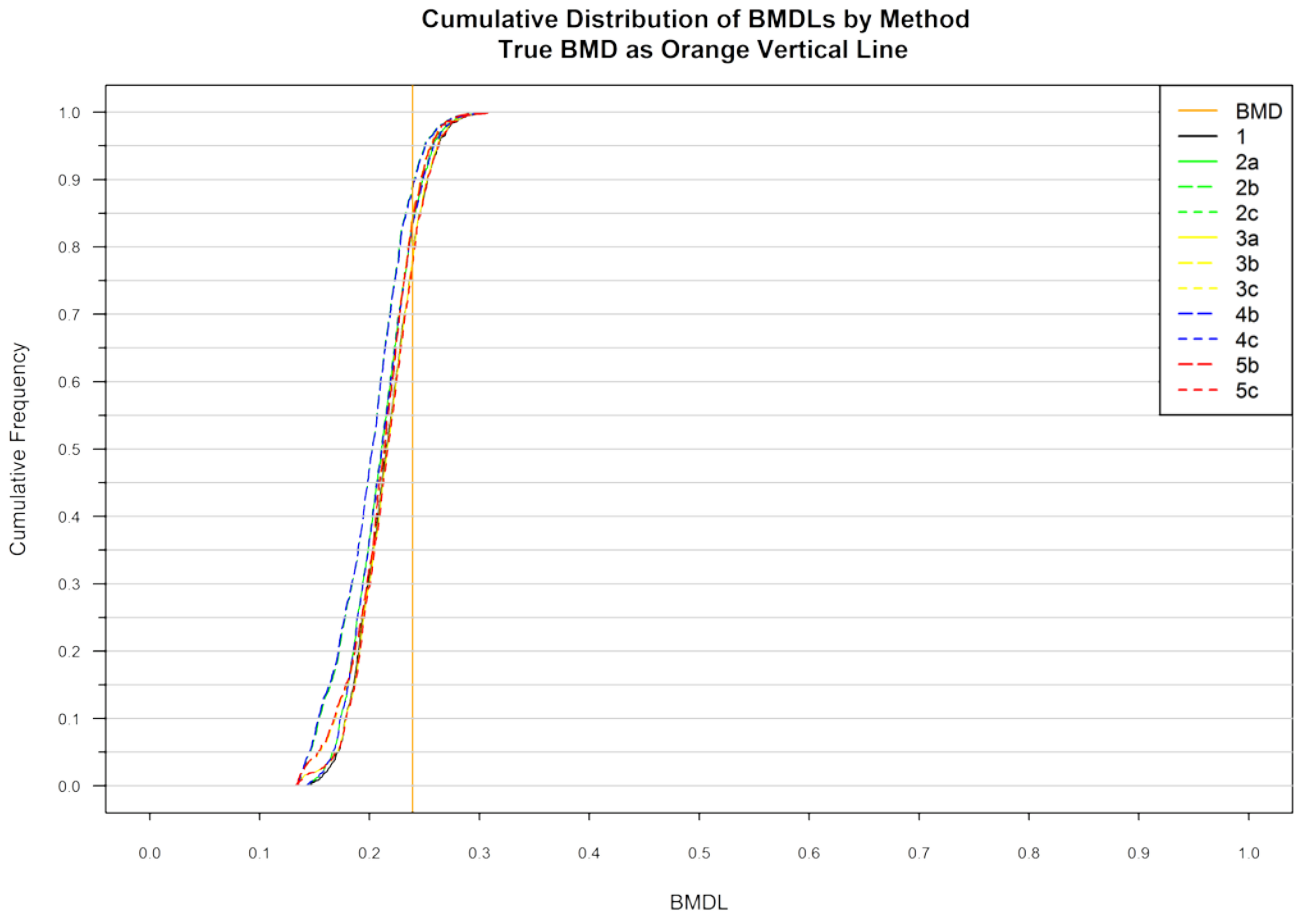
Figure 52: Template w2_lognormal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.380818	0.384356	
50	0.424401	0.427869	0.4281
75	0.470464	0.473598	
IQR	0.0896464	0.0892417	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.979		
Poly3	0.998	**2a**	0.935	**3a**	0.978
Power	0.976	**2b**	0.935	**3b**	0.978
Hill	0.975	**2c**	0.934	**3c**	0.978
Exp3	0.98	**4b**	0.926	**5b**	0.978
Exp5	0.979	**4c**	0.931	**5c**	0.978

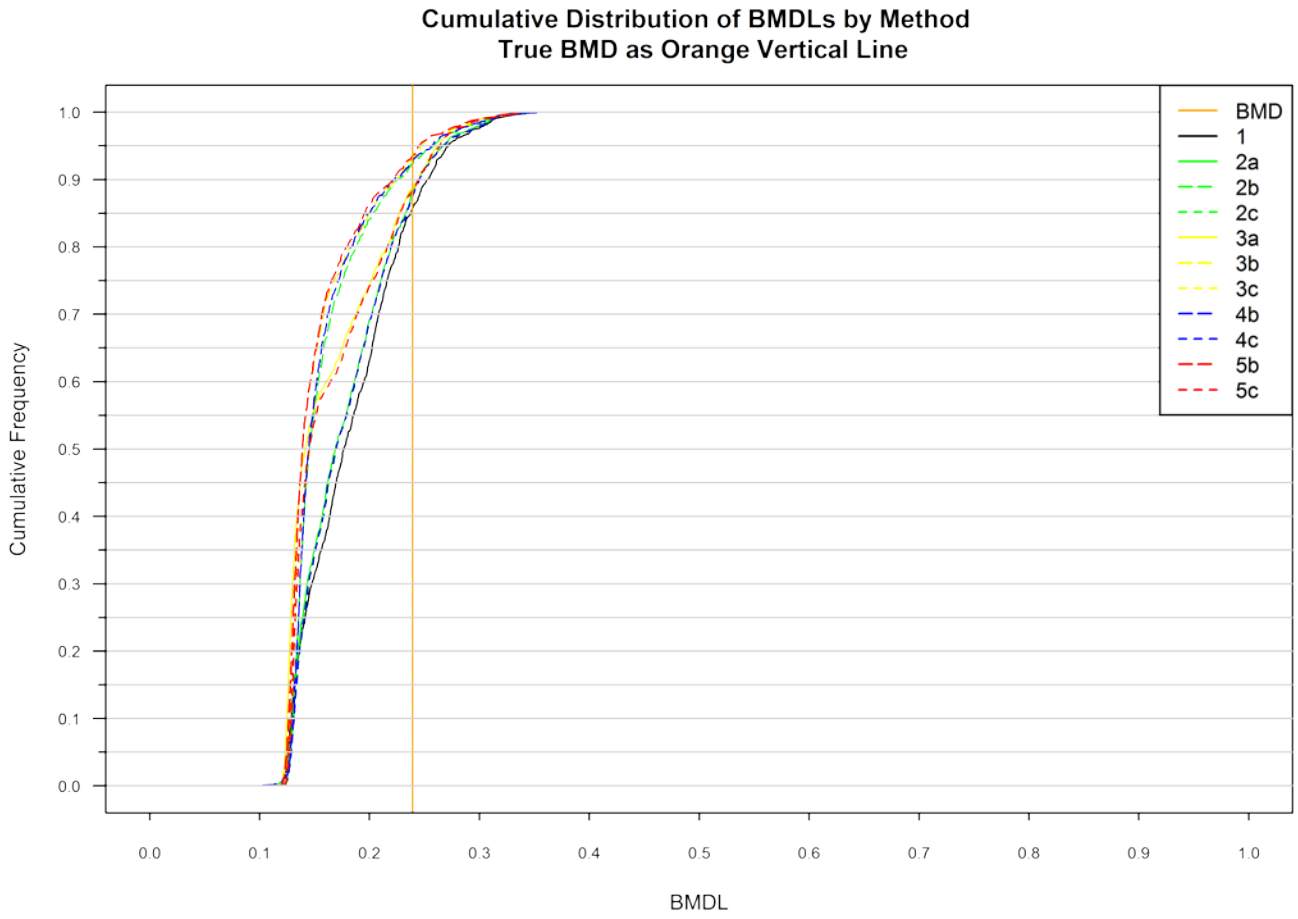
Figure 53: Template w3_lognormal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.226424	0.226995	
50	0.249991	0.250801	0.2392
75	0.271589	0.272134	
IQR	0.0451654	0.0451389	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.781		
Poly3	0.988	**2a**	0.833	**3a**	0.78
Power	0.931	**2b**	0.892	**3b**	0.846
Hill	0.867	**2c**	0.839	**3c**	0.785
Exp3	0.386	**4b**	0.888	**5b**	0.841
Exp5	0.386	**4c**	0.832	**5c**	0.772

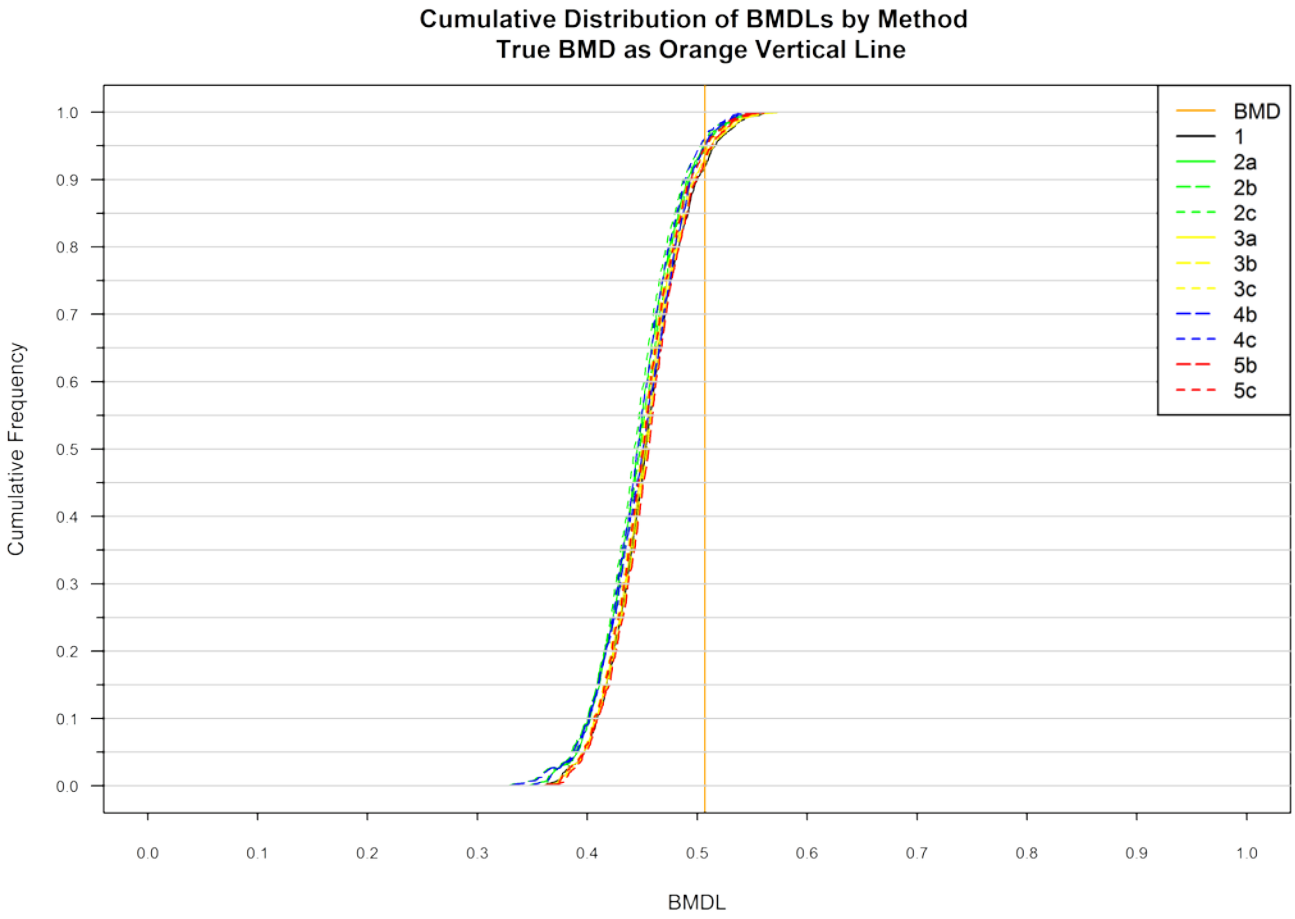
Figure 54: Template w3_lognormal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.182798	0.178882	
50	0.243011	0.239842	0.2392
75	0.293527	0.294127	
IQR	0.110729	0.115245	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.858		
Poly3	0.975	**2a**	0.88	**3a**	0.885
Power	0.937	**2b**	0.924	**3b**	0.93
Hill	0.859	**2c**	0.877	**3c**	0.885
Exp3	0.736	**4b**	0.927	**5b**	0.937
Exp5	0.705	**4c**	0.874	**5c**	0.885

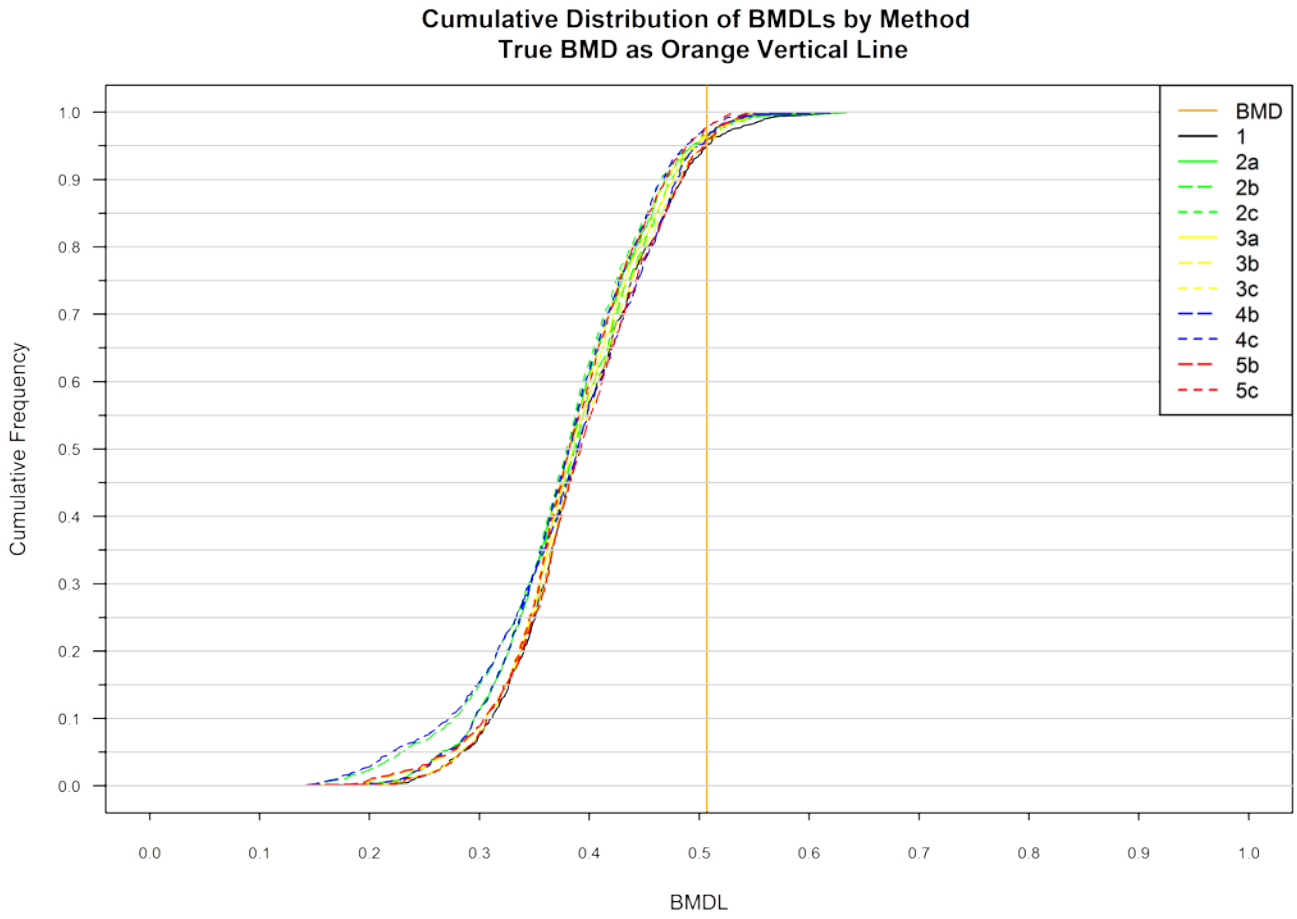
Figure 55: Template w4_lognormal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.483978	0.485105	
50	0.508746	0.508204	0.5071
75	0.532366	0.528799	
IQR	0.0483877	0.0436942	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.919		
Poly3	0.971	**2a**	0.951	**3a**	0.936
Power	0.922	**2b**	0.947	**3b**	0.924
Hill	0.951	**2c**	0.958	**3c**	0.945
Exp3	0.913	**4b**	0.949	**5b**	0.928
Exp5	0.942	**4c**	0.96	**5c**	0.945

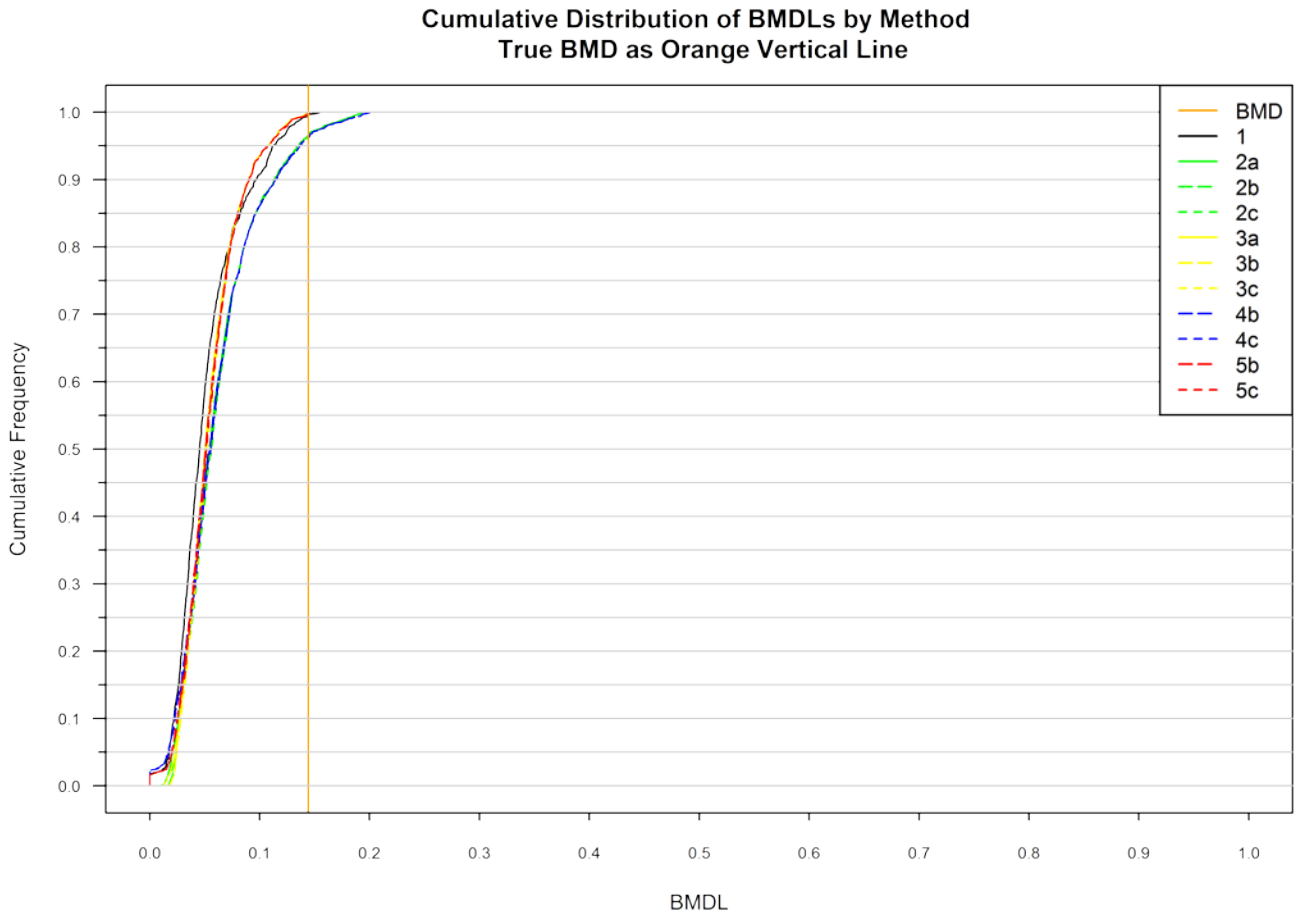
Figure 56: Template w4_lognormal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.457513	0.461786	
50	0.499192	0.499209	0.5071
75	0.555677	0.547561	
IQR	0.0981637	0.0857753	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.95		
Poly3	0.996	**2a**	0.964	**3a**	0.964
Power	0.948	**2b**	0.957	**3b**	0.957
Hill	0.964	**2c**	0.967	**3c**	0.965
Exp3	0.946	**4b**	0.965	**5b**	0.959
Exp5	0.947	**4c**	0.979	**5c**	0.977

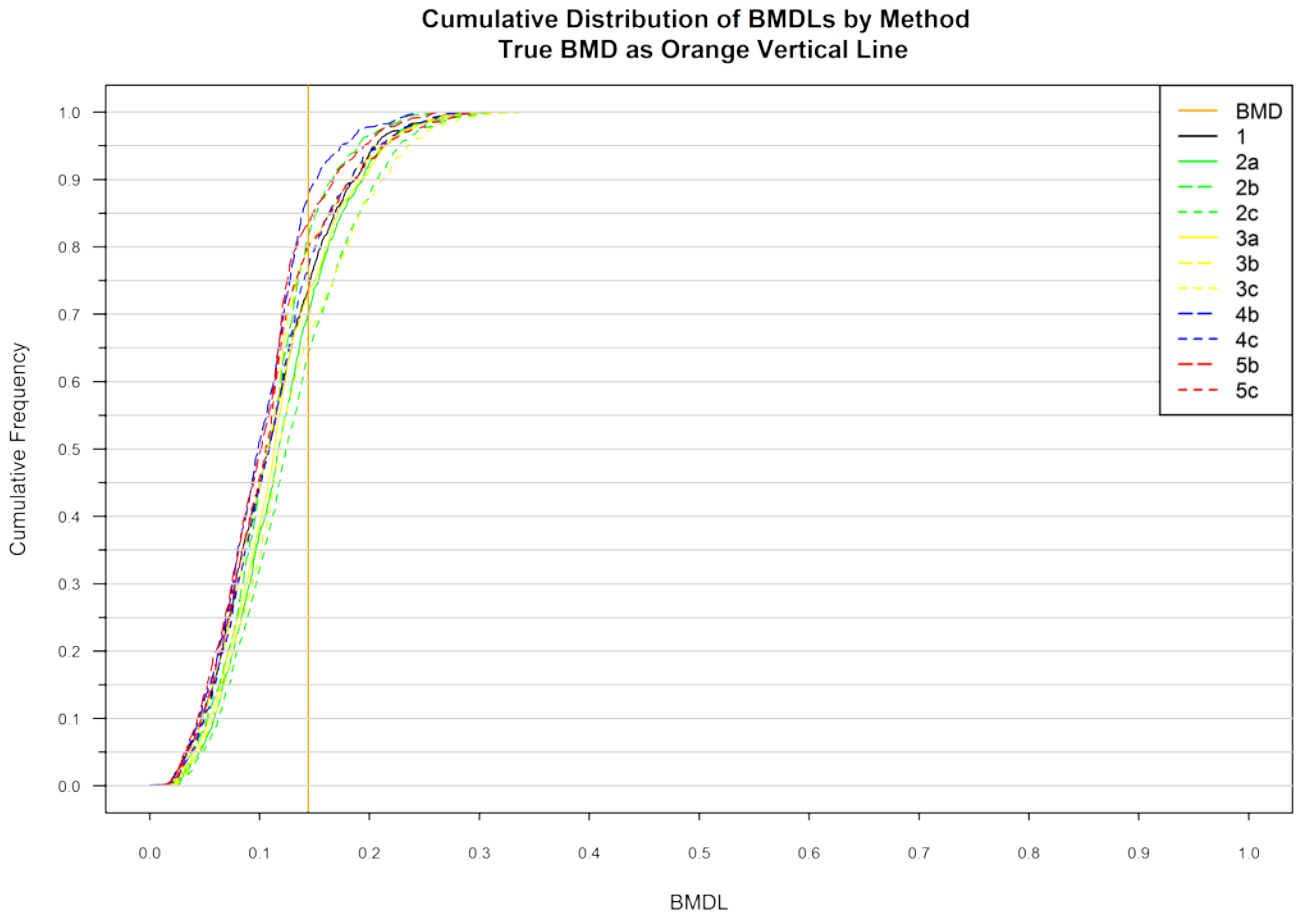
Figure 57: Template h1_lognormal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.08576	0.0857399	
50	0.135251	0.137	0.1443
75	0.198188	0.205155	
IQR	0.112428	0.119415	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.996		
Poly3	0	**2a**	0.966	**3a**	0.999
Power	0	**2b**	0.965	**3b**	0.998
Hill	0.974	**2c**	0.968	**3c**	0.999
Exp3	0	**4b**	0.962	**5b**	0.996
Exp5	0.999	**4c**	0.962	**5c**	0.998

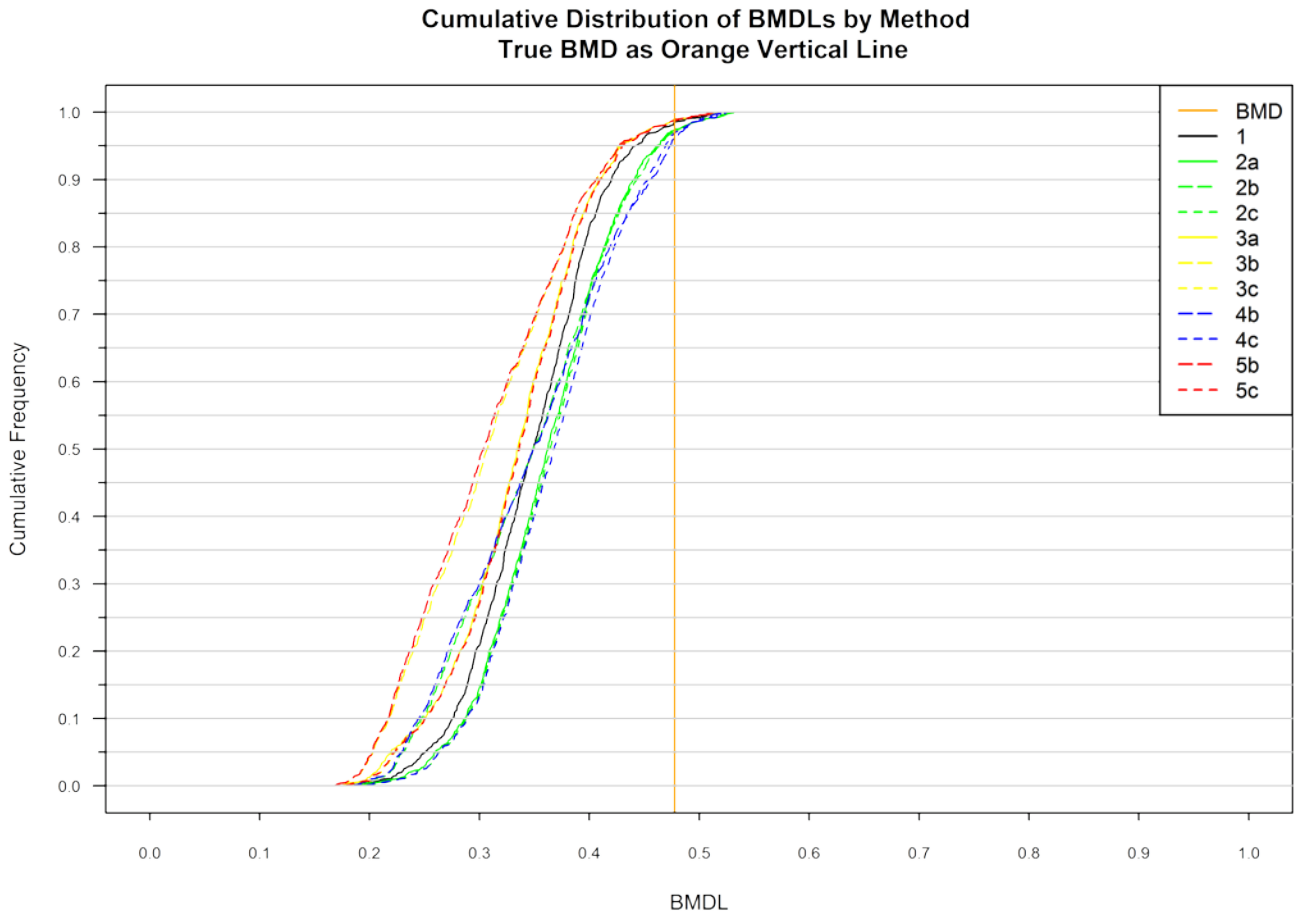
Figure 58: Template h1_lognormal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.139391	0.128069	
50	0.187716	0.168573	0.1443
75	0.236928	0.22583	
IQR	0.0975368	0.0977607	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.005	**1**	0.739		
Poly3	0.005	**2a**	0.7	**3a**	0.731
Power	0.005	**2b**	0.816	**3b**	0.796
Hill	0.955	**2c**	0.645	**3c**	0.667
Exp3	0	**4b**	0.879	**5b**	0.838
Exp5	0.973	**4c**	0.77	**5c**	0.8

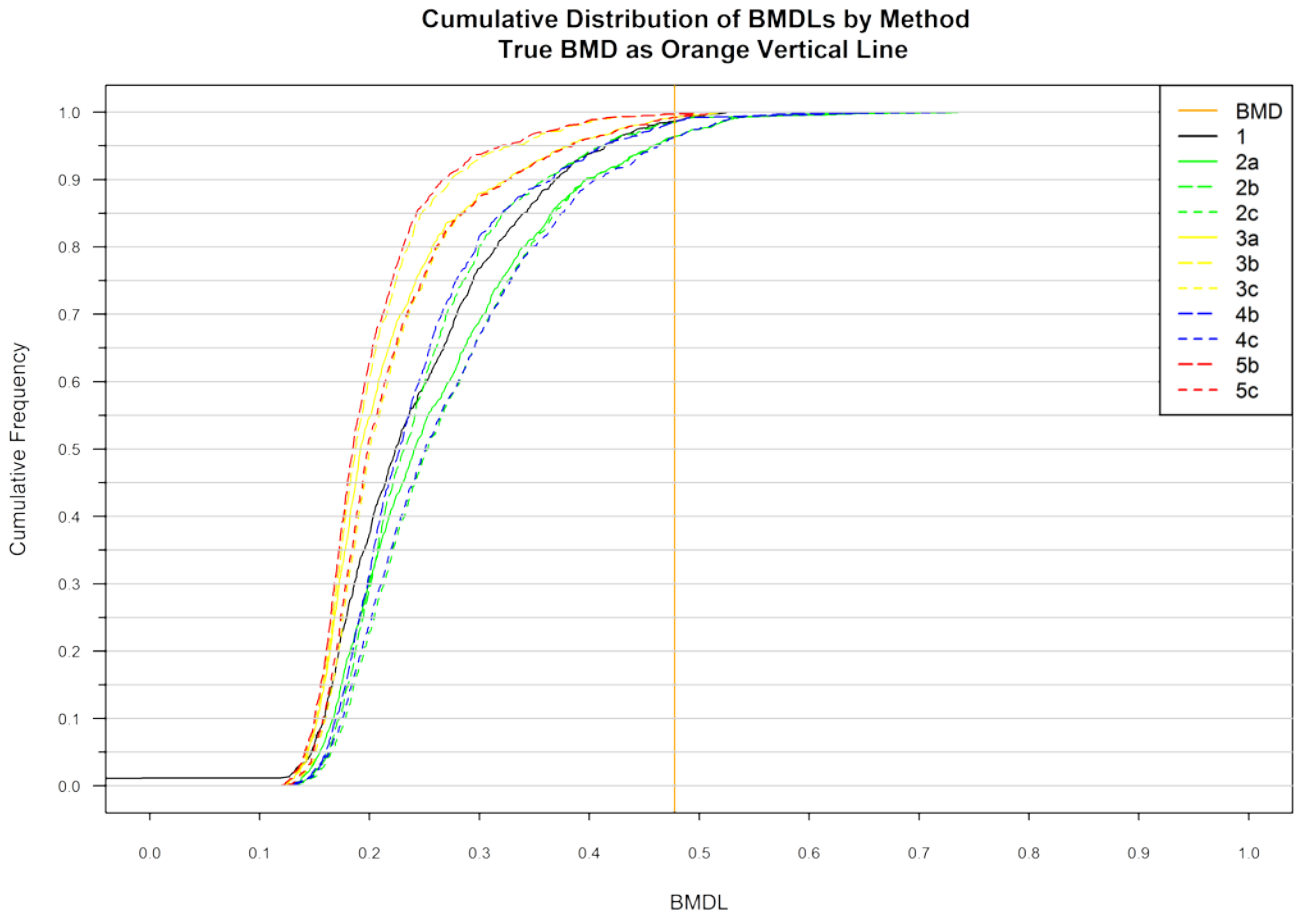
Figure 59: Template h2_lognormal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.419646	0.424498	
50	0.463087	0.473294	0.4777
75	0.502162	0.500561	
IQR	0.0825155	0.076063	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.984		
Poly3	0.974	**2a**	0.973	**3a**	0.988
Power	0.984	**2b**	0.97	**3b**	0.989
Hill	0.981	**2c**	0.973	**3c**	0.988
Exp3	0.983	**4b**	0.96	**5b**	0.987
Exp5	0.981	**4c**	0.968	**5c**	0.987

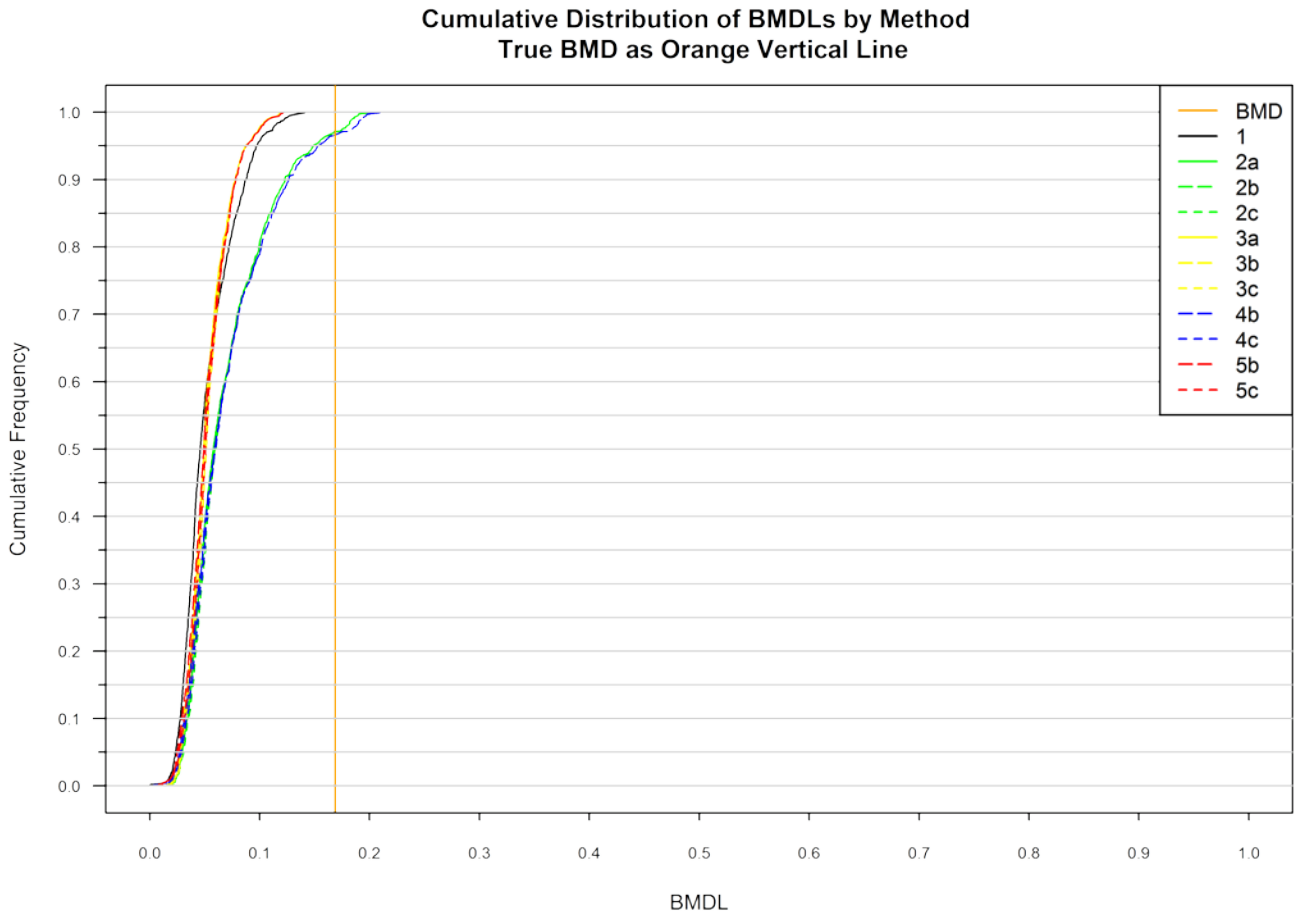
Figure 60: Template h2_lognormal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.300778	0.294058	
50	0.397386	0.397799	0.4777
75	0.504696	0.502542	
IQR	0.203918	0.208483	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.986		
Poly3	0.965	**2a**	0.964	**3a**	0.992
Power	0.977	**2b**	0.985	**3b**	0.997
Hill	0.976	**2c**	0.963	**3c**	0.992
Exp3	0.986	**4b**	0.985	**5b**	0.997
Exp5	0.986	**4c**	0.963	**5c**	0.992

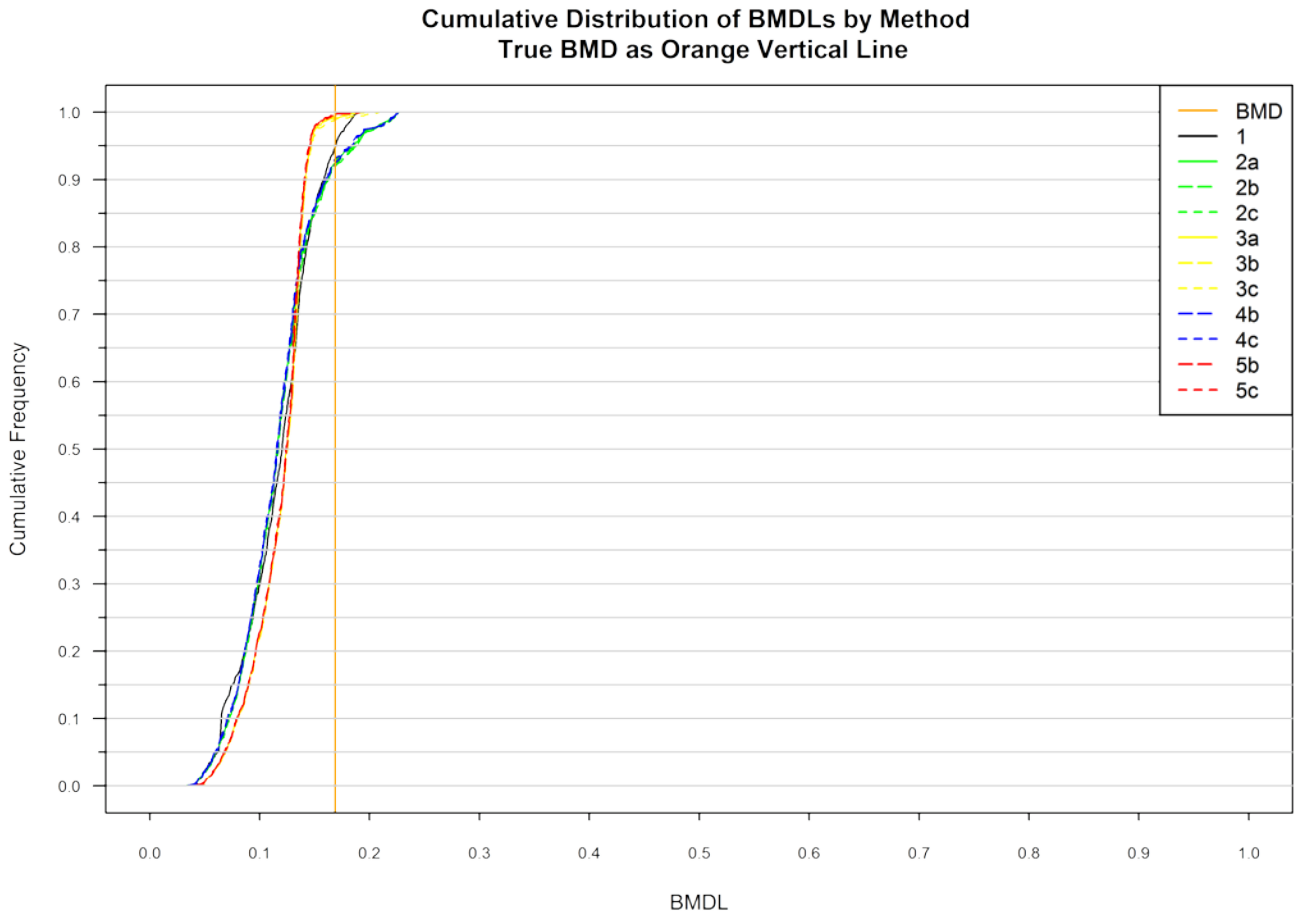
Figure 61: Template h3_lognormal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.101294	0.103851	
50	0.149082	0.154329	0.1688
75	0.197756	0.206451	
IQR	0.0964614	0.1026	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	1		
Poly3	0	**2a**	0.971	**3a**	1
Power	0	**2b**	0.97	**3b**	1
Hill	1	**2c**	0.97	**3c**	1
Exp3	0	**4b**	0.966	**5b**	1
Exp5	1	**4c**	0.966	**5c**	1

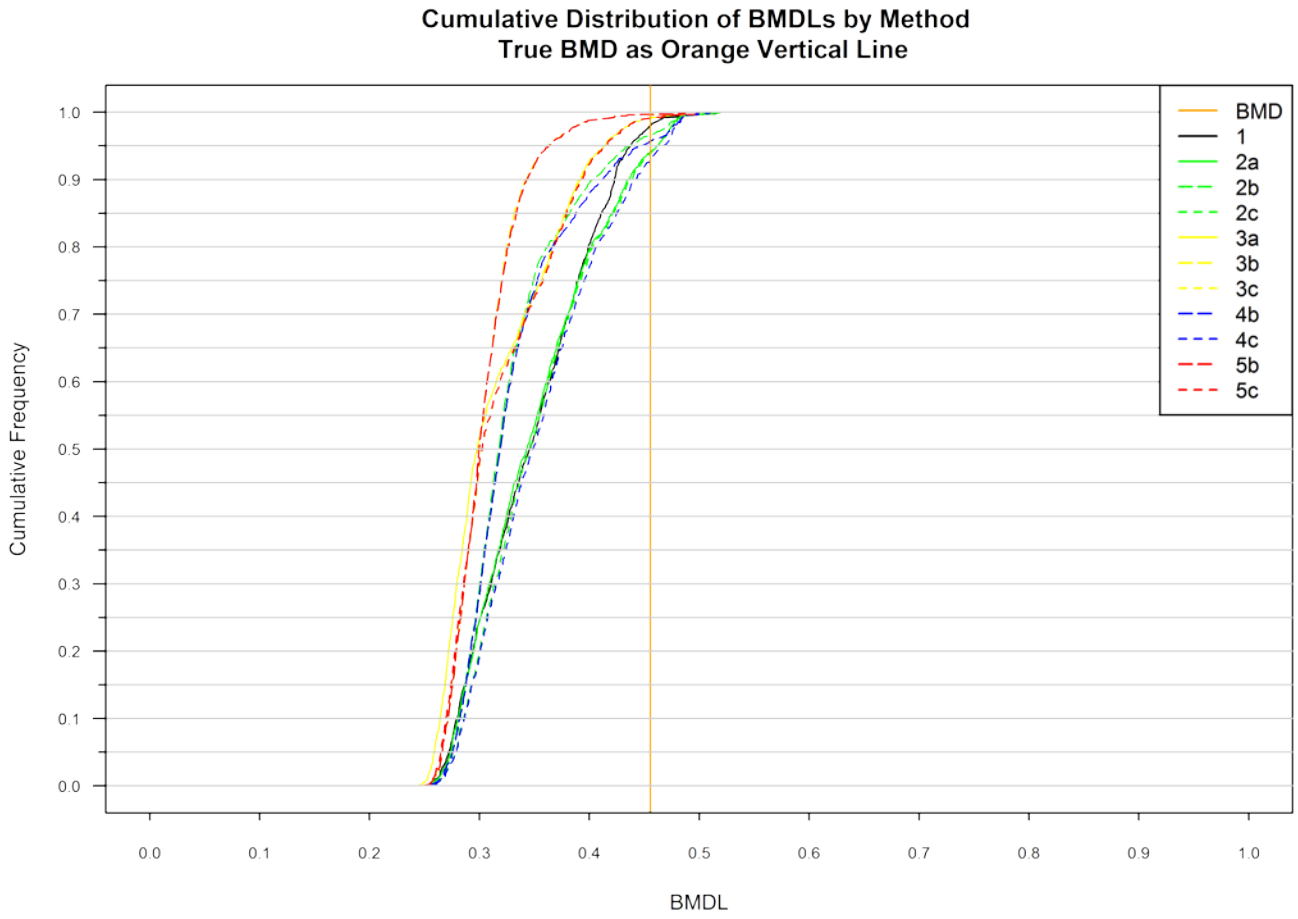
Figure 62: Template h3_lognormal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.145091	0.144347	
50	0.178532	0.176081	0.1688
75	0.22341	0.223368	
IQR	0.0783191	0.0790213	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.948		
Poly3	0	**2a**	0.925	**3a**	0.991
Power	0	**2b**	0.919	**3b**	0.993
Hill	0.968	**2c**	0.922	**3c**	0.988
Exp3	0.483	**4b**	0.923	**5b**	0.997
Exp5	0.948	**4c**	0.93	**5c**	0.995

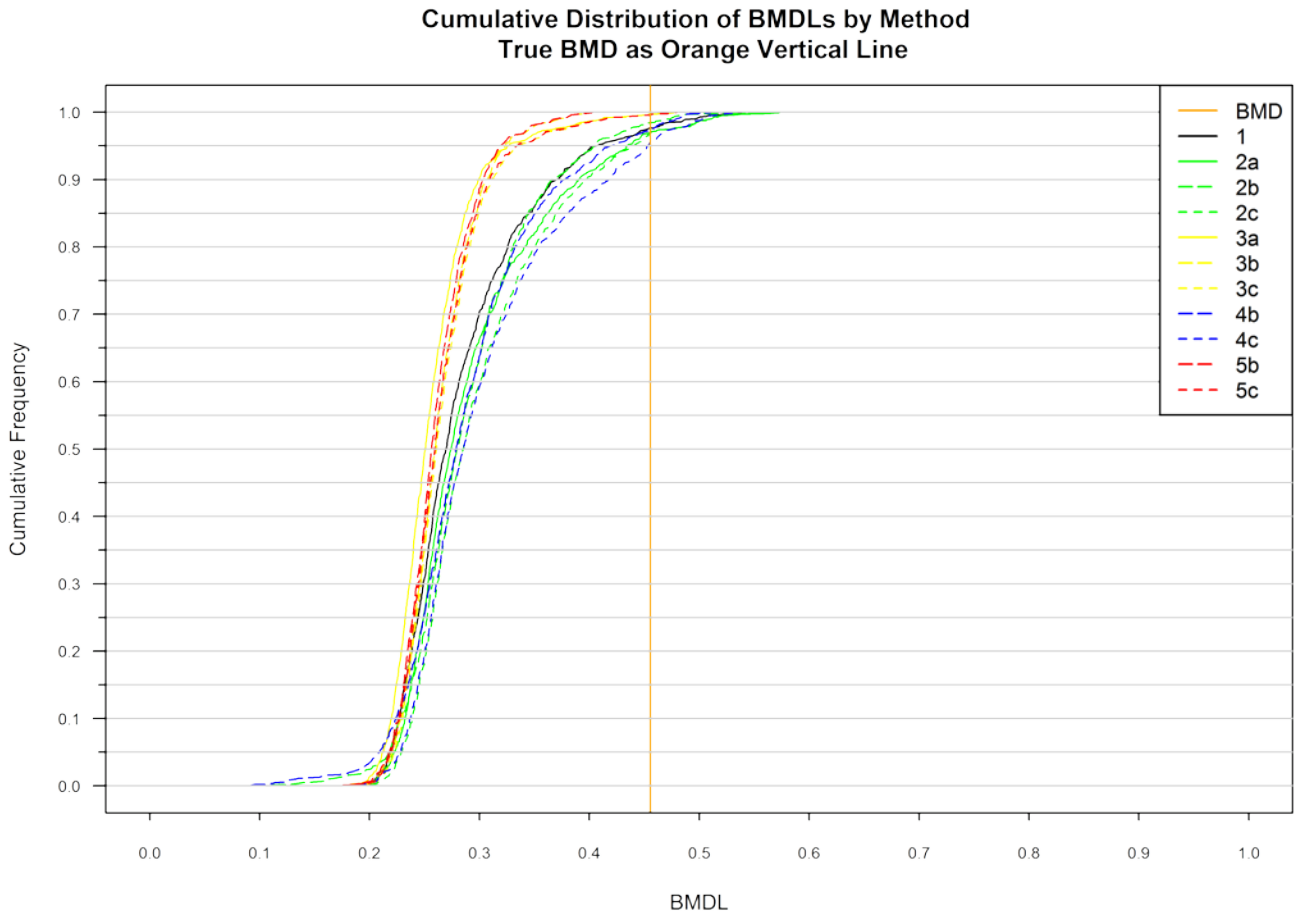
Figure 63: Template h4_lognormal_chronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.361197	0.366285	
50	0.429187	0.437941	0.4556
75	0.47778	0.484914	
IQR	0.116584	0.118629	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.98		
Poly3	0.961	**2a**	0.941	**3a**	0.991
Power	0.982	**2b**	0.965	**3b**	0.996
Hill	0.973	**2c**	0.938	**3c**	0.991
Exp3	0.984	**4b**	0.957	**5b**	0.996
Exp5	0.966	**4c**	0.928	**5c**	0.99

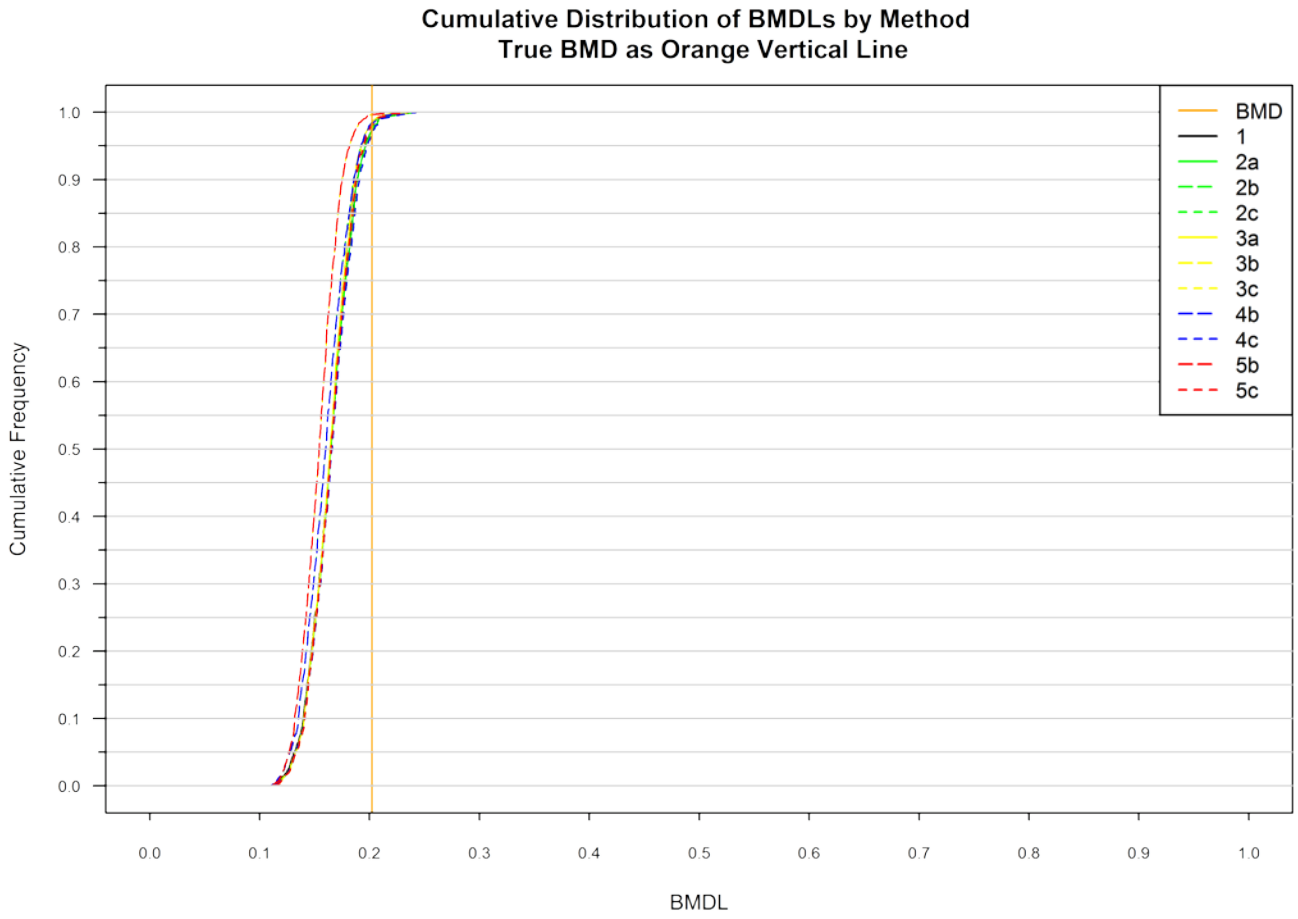
Figure 64: Template h4_lognormal_subchronic; Models fit assuming constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.321858	0.326538	
50	0.379188	0.393375	0.4556
75	0.453001	0.470005	
IQR	0.131143	0.143466	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.976		
Poly3	0.886	**2a**	0.97	**3a**	0.997
Power	0.968	**2b**	0.984	**3b**	0.999
Hill	0.971	**2c**	0.969	**3c**	0.997
Exp3	0.969	**4b**	0.974	**5b**	0.999
Exp5	0.969	**4c**	0.956	**5c**	0.996

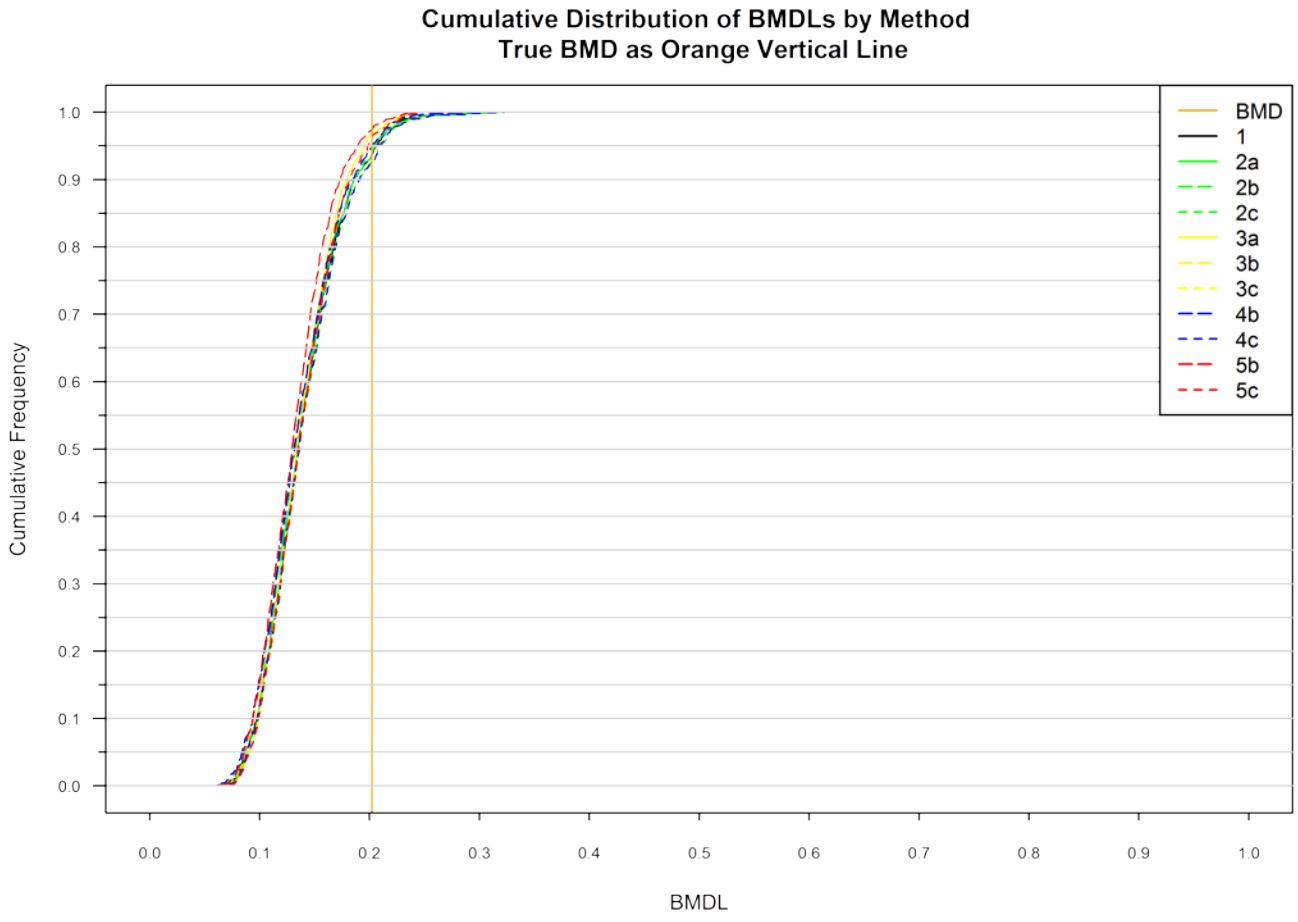
Figure 65: Template w1_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.176174	0.177007	
50	0.19168	0.19213	0.2021
75	0.206559	0.2066	
IQR	0.0303858	0.0295928	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.974		
Poly3	0.955	**2a**	0.973	**3a**	0.984
Power	0.956	**2b**	0.984	**3b**	0.996
Hill	0.912	**2c**	0.967	**3c**	0.981
Exp3	1	**4b**	0.984	**5b**	0.996
Exp5	0.927	**4c**	0.967	**5c**	0.981

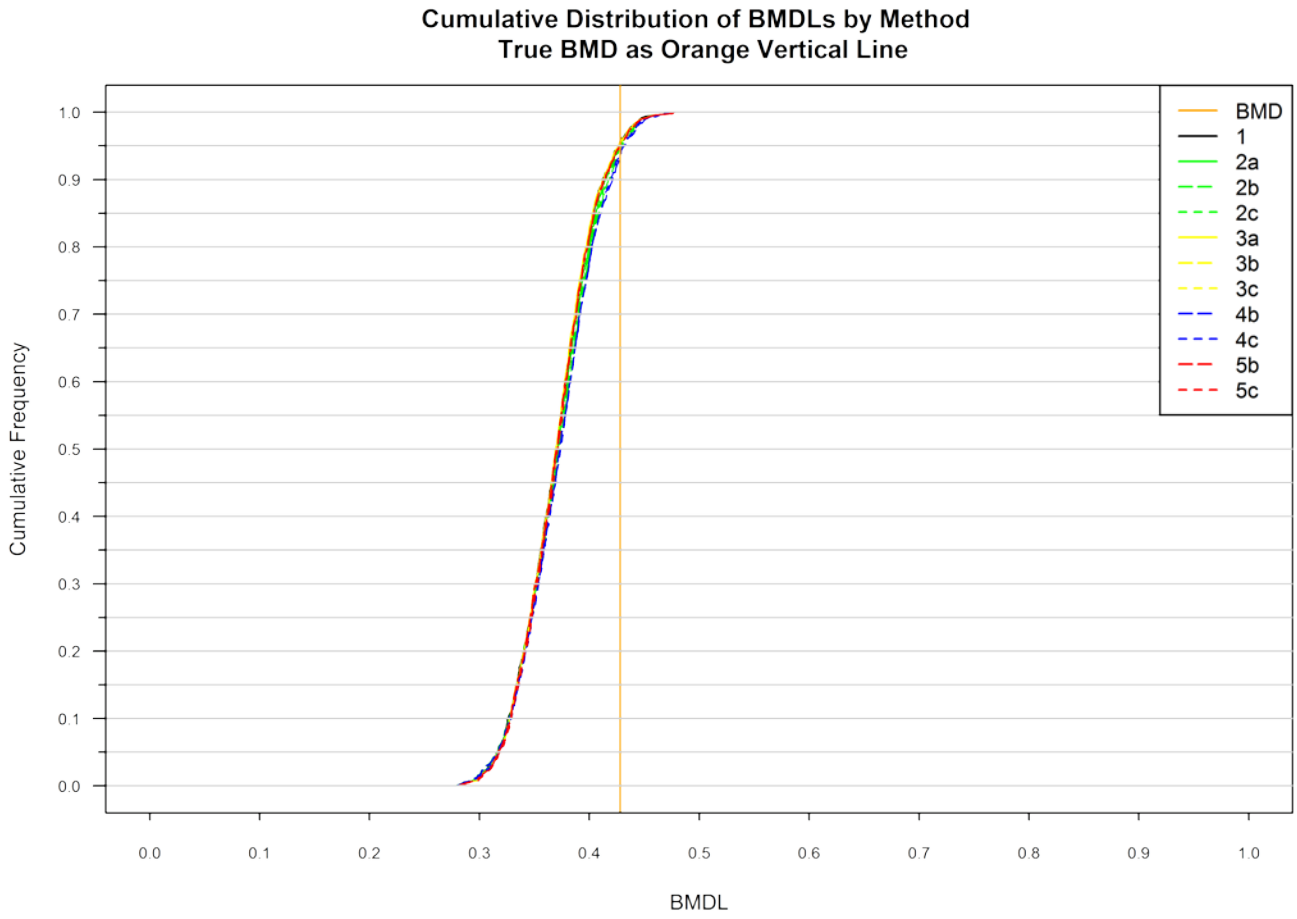
Figure 66: Template w1_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.165195	0.166403	
50	0.191071	0.192183	0.2021
75	0.22384	0.225151	
IQR	0.058645	0.0587472	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.937		
Poly3	0.919	**2a**	0.937	**3a**	0.967
Power	0.934	**2b**	0.948	**3b**	0.974
Hill	0.877	**2c**	0.927	**3c**	0.962
Exp3	0.992	**4b**	0.947	**5b**	0.974
Exp5	0.9	**4c**	0.928	**5c**	0.962

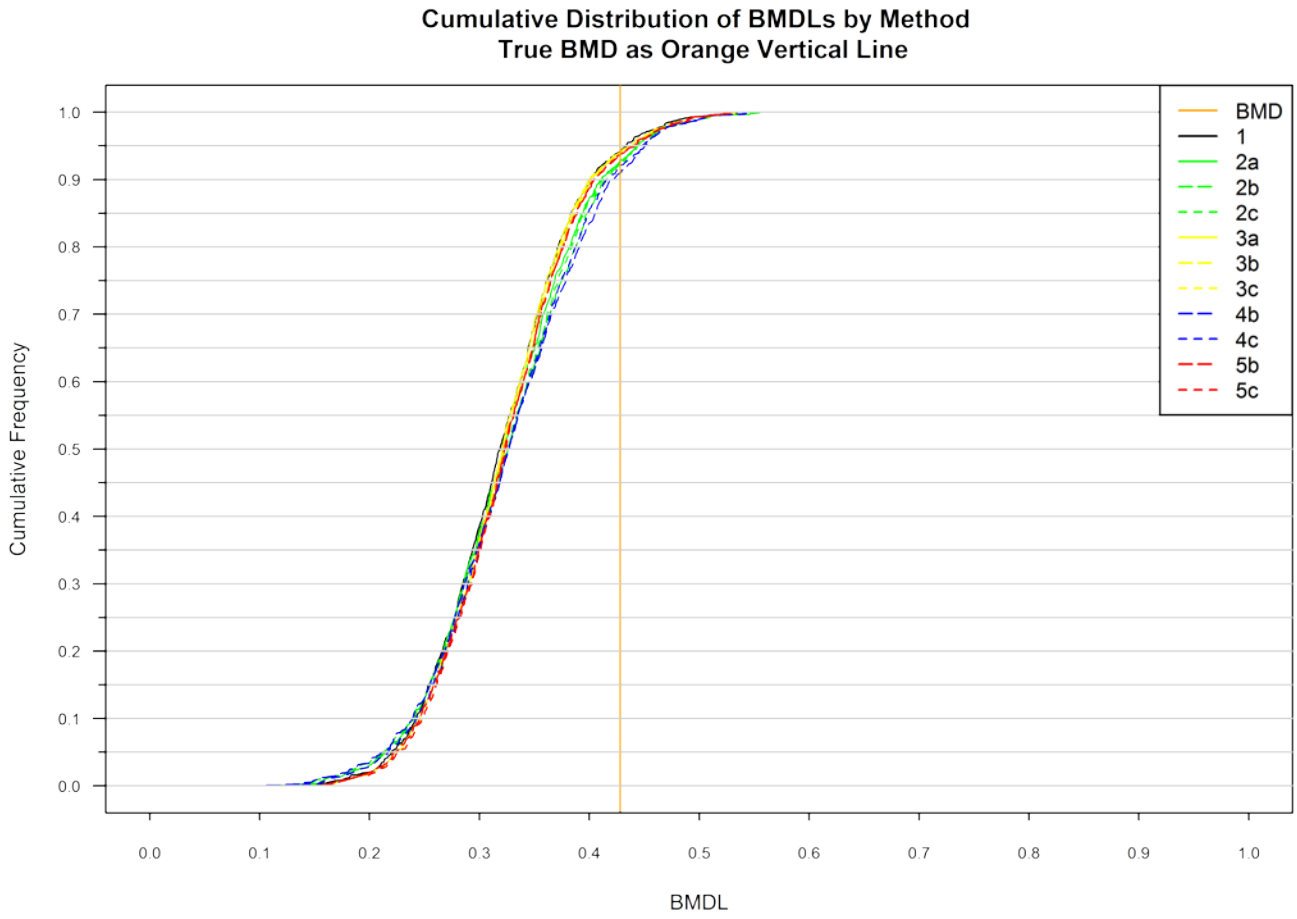
Figure 67: Template w2_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.39898	0.400564	
50	0.424713	0.42754	0.4281
75	0.448081	0.450625	
IQR	0.0491019	0.0500615	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.957		
Poly3	1	**2a**	0.95	**3a**	0.956
Power	0.938	**2b**	0.946	**3b**	0.956
Hill	0.921	**2c**	0.947	**3c**	0.956
Exp3	0.97	**4b**	0.936	**5b**	0.951
Exp5	0.947	**4c**	0.94	**5c**	0.951

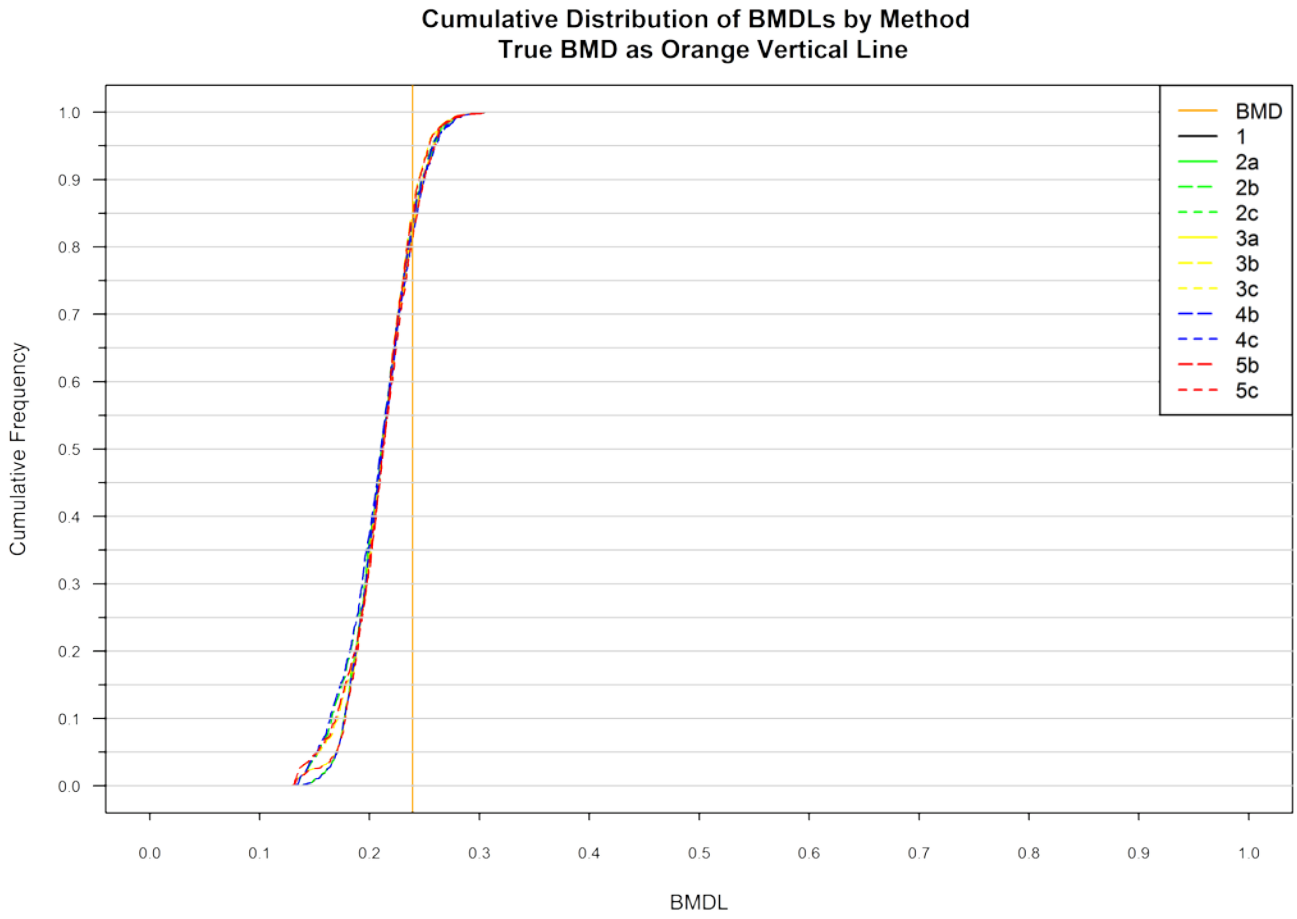
Figure 68: Template w2_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.381787	0.385298	
50	0.426162	0.430857	0.4281
75	0.47272	0.476176	
IQR	0.0909328	0.0908777	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.941		
Poly3	0.999	**2a**	0.929	**3a**	0.941
Power	0.934	**2b**	0.921	**3b**	0.939
Hill	0.93	**2c**	0.926	**3c**	0.939
Exp3	0.953	**4b**	0.91	**5b**	0.937
Exp5	0.941	**4c**	0.921	**5c**	0.937

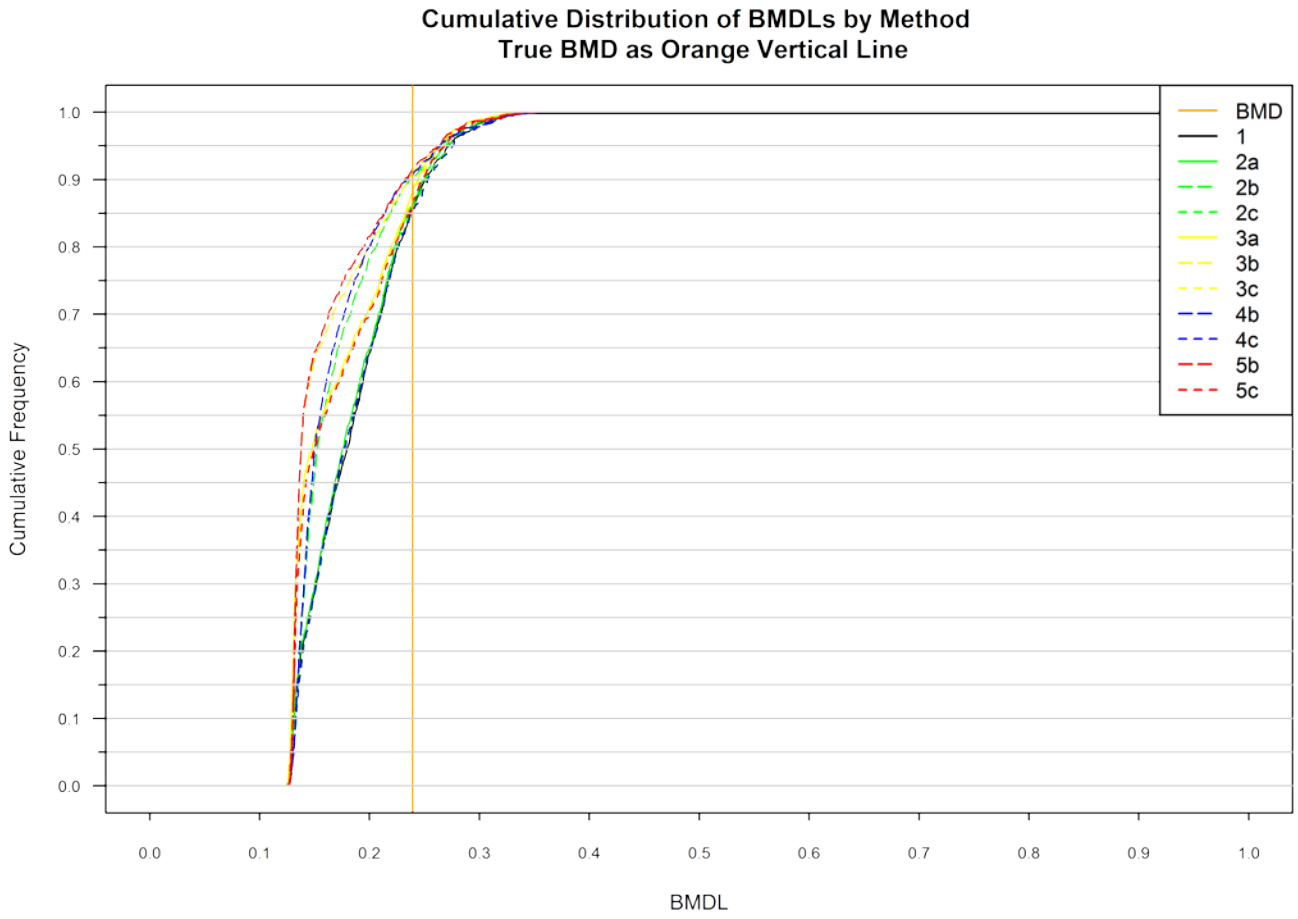
Figure 69: Template w3_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.227419	0.22775	
50	0.247814	0.24835	0.2392
75	0.269499	0.270257	
IQR	0.0420794	0.0425071	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.826		
Poly3	0.995	**2a**	0.827	**3a**	0.824
Power	0.939	**2b**	0.838	**3b**	0.85
Hill	0.876	**2c**	0.826	**3c**	0.823
Exp3	0.108	**4b**	0.834	**5b**	0.844
Exp5	0.237	**4c**	0.813	**5c**	0.81

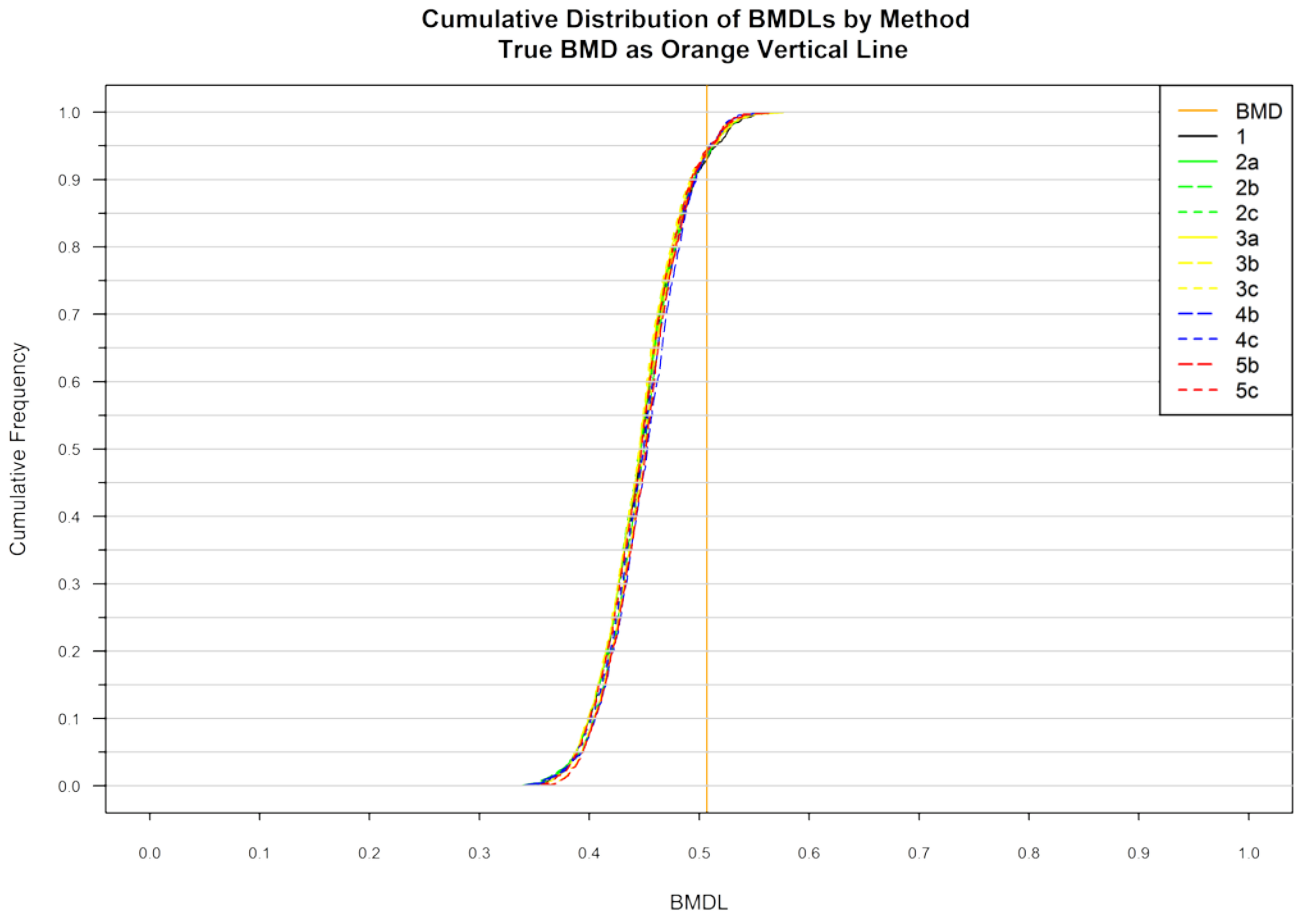
Figure 70: Template w3_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.185504	0.179291	
50	0.245332	0.241669	0.2392
75	0.297298	0.297072	
IQR	0.111794	0.117781	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.859		
Poly3	0.98	**2a**	0.864	**3a**	0.881
Power	0.929	**2b**	0.903	**3b**	0.909
Hill	0.846	**2c**	0.858	**3c**	0.876
Exp3	0.313	**4b**	0.905	**5b**	0.915
Exp5	0.584	**4c**	0.854	**5c**	0.866

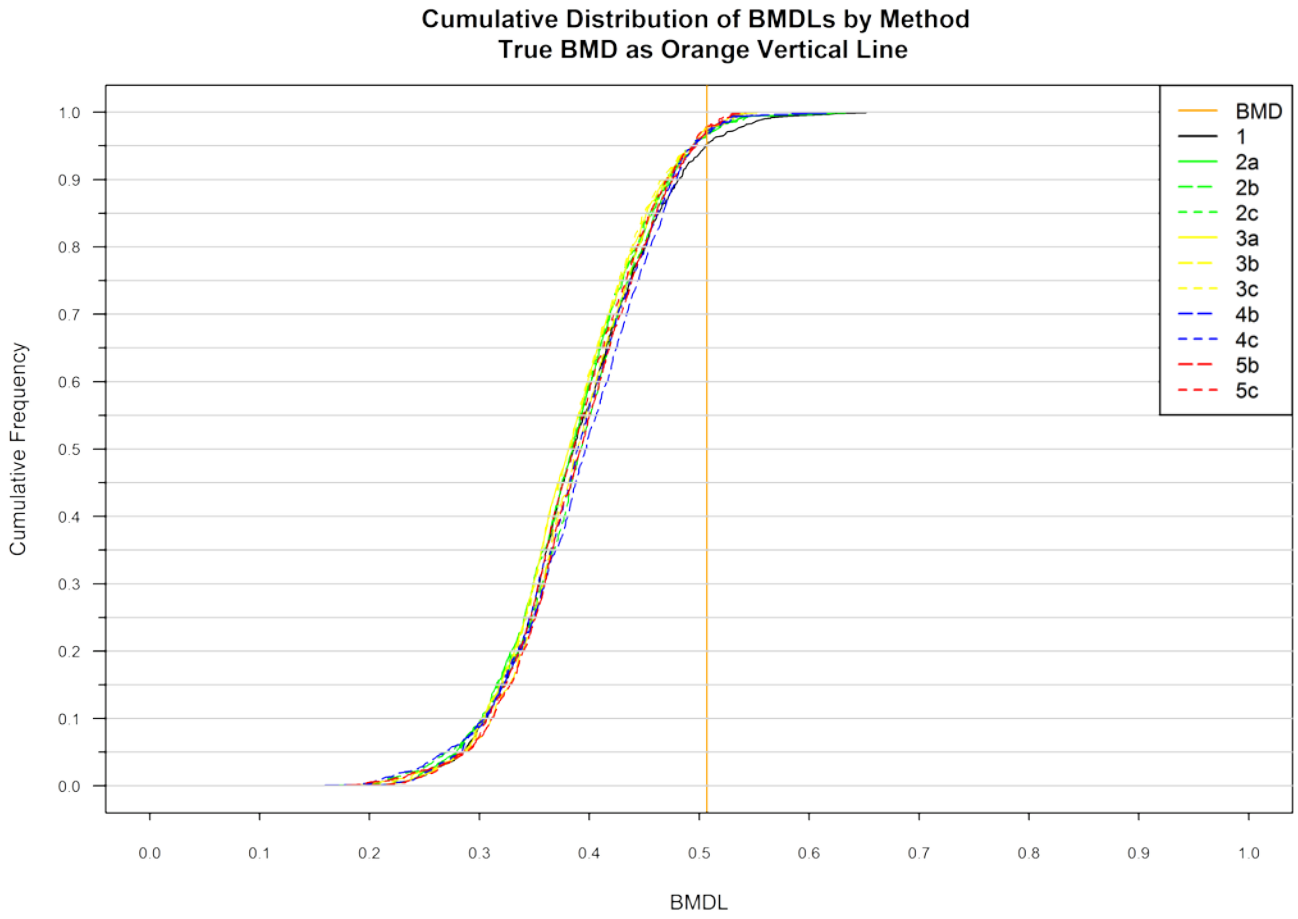
Figure 71: Template w4_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.481565	0.483403	
50	0.508955	0.508025	0.5071
75	0.534072	0.52863	
IQR	0.052507	0.0452269	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.93		
Poly3	0.977	**2a**	0.937	**3a**	0.943
Power	0.929	**2b**	0.934	**3b**	0.94
Hill	0.955	**2c**	0.94	**3c**	0.945
Exp3	0.446	**4b**	0.937	**5b**	0.942
Exp5	0.955	**4c**	0.94	**5c**	0.945

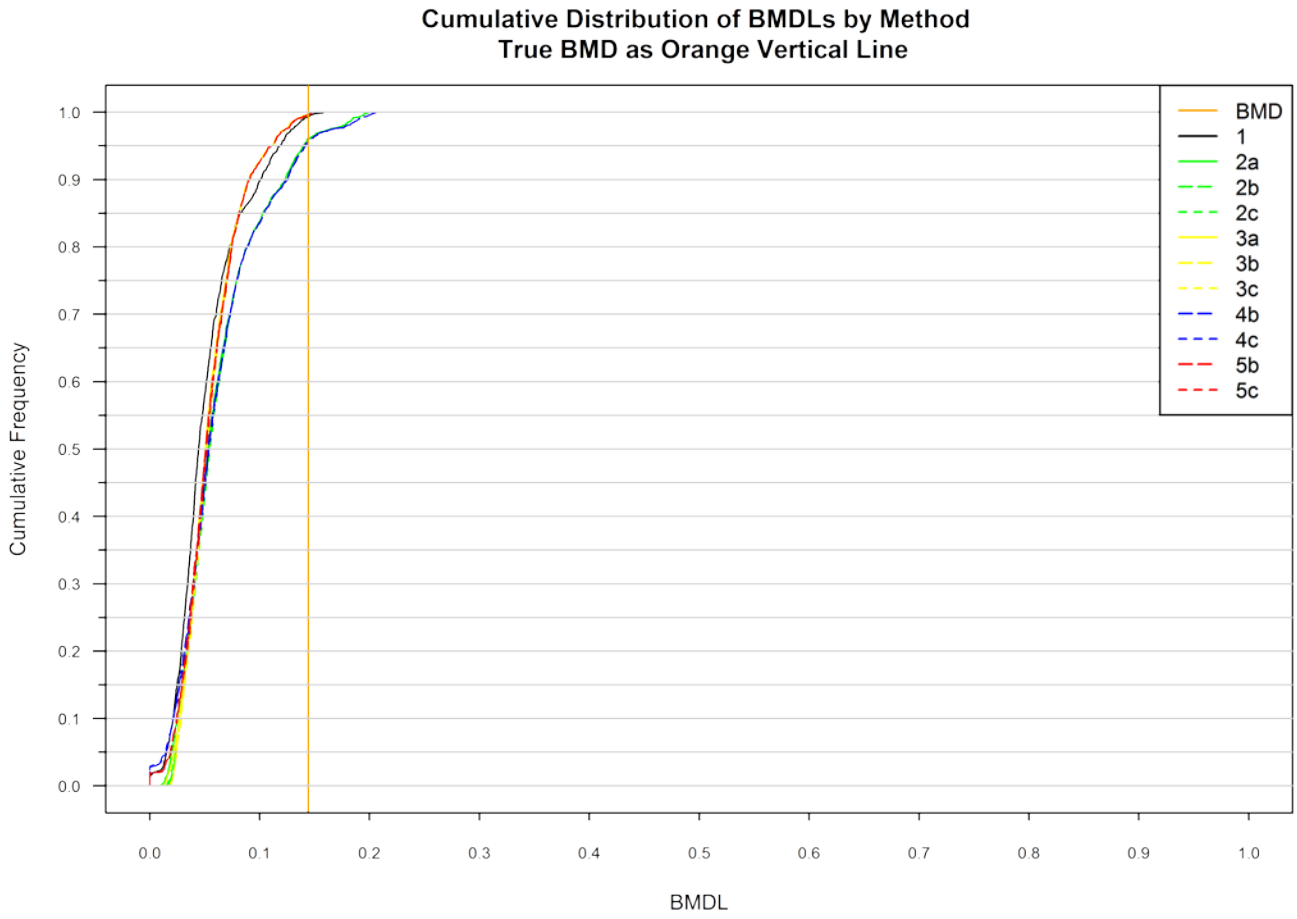
Figure 72: Template w4_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.453489	0.461124	
50	0.499101	0.498702	0.5071
75	0.559228	0.544795	
IQR	0.10574	0.0836716	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.954		
Poly3	0.959	**2a**	0.967	**3a**	0.978
Power	0.952	**2b**	0.964	**3b**	0.972
Hill	0.966	**2c**	0.968	**3c**	0.978
Exp3	0.556	**4b**	0.968	**5b**	0.973
Exp5	0.957	**4c**	0.972	**5c**	0.979

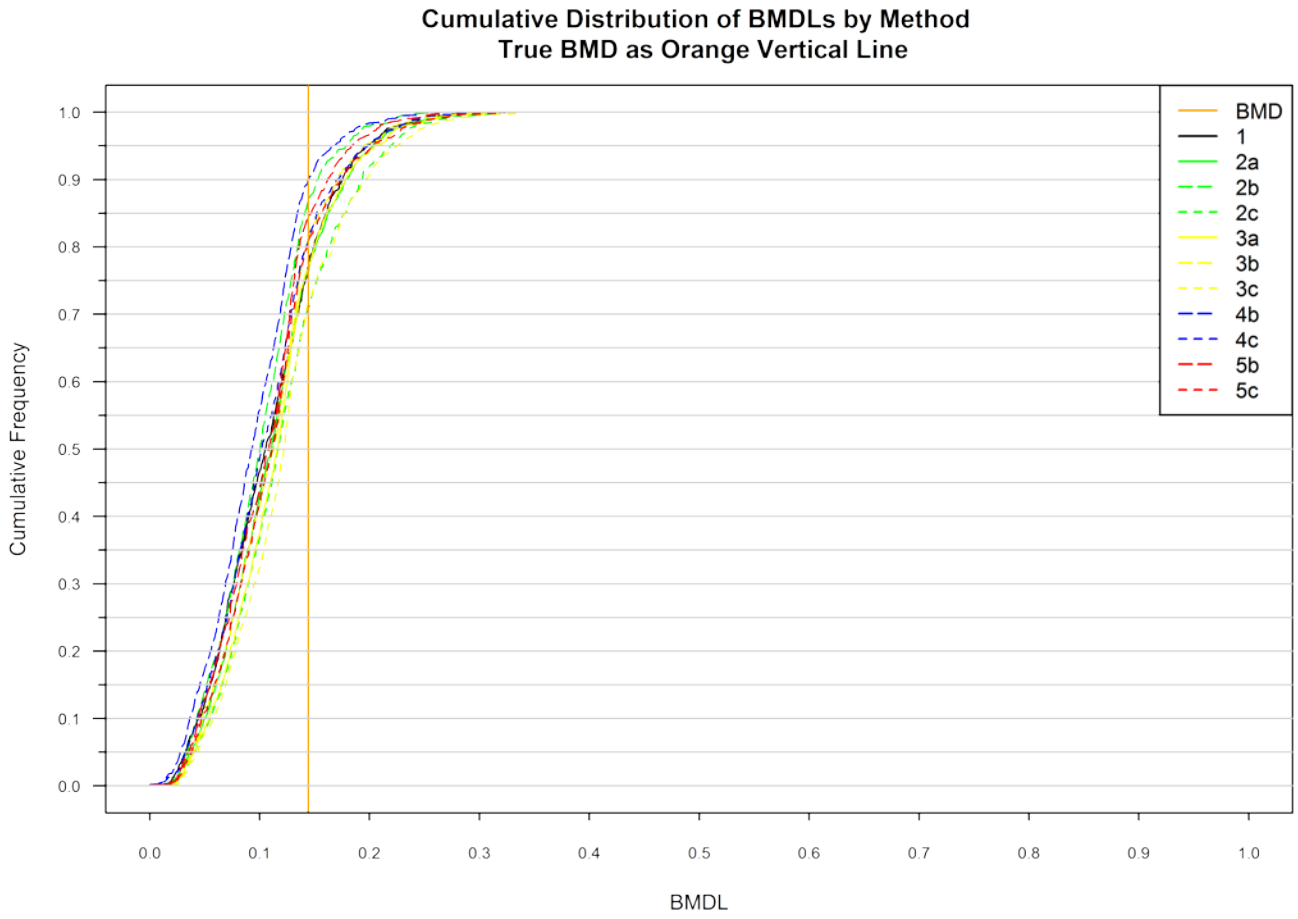
Figure 73: Template h1_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.0833812	0.0838261	
50	0.137823	0.141109	0.1443
75	0.197964	0.205351	
IQR	0.114583	0.121525	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.992		
Poly3	0	**2a**	0.96	**3a**	0.998
Power	0	**2b**	0.96	**3b**	0.998
Hill	0.973	**2c**	0.96	**3c**	0.997
Exp3	0	**4b**	0.957	**5b**	0.995
Exp5	0.999	**4c**	0.957	**5c**	0.995

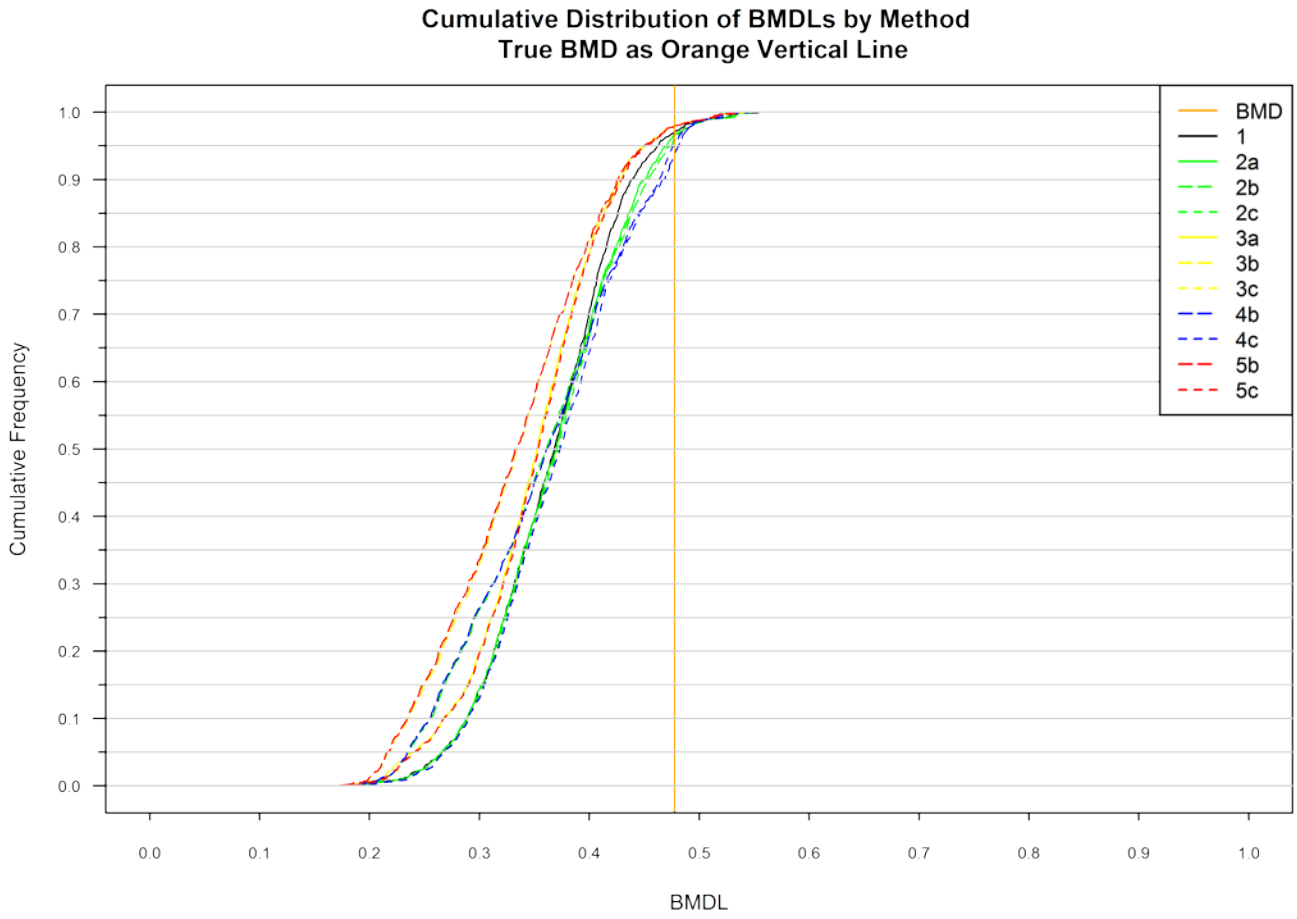
Figure 74: Template h1_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.13163	0.124518	
50	0.176081	0.160355	0.1443
75	0.229743	0.224512	
IQR	0.0981124	0.0999938	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.196	**1**	0.772		
Poly3	0.196	**2a**	0.764	**3a**	0.769
Power	0.196	**2b**	0.866	**3b**	0.812
Hill	0.922	**2c**	0.709	**3c**	0.716
Exp3	0.011	**4b**	0.9	**5b**	0.845
Exp5	0.94	**4c**	0.809	**5c**	0.803

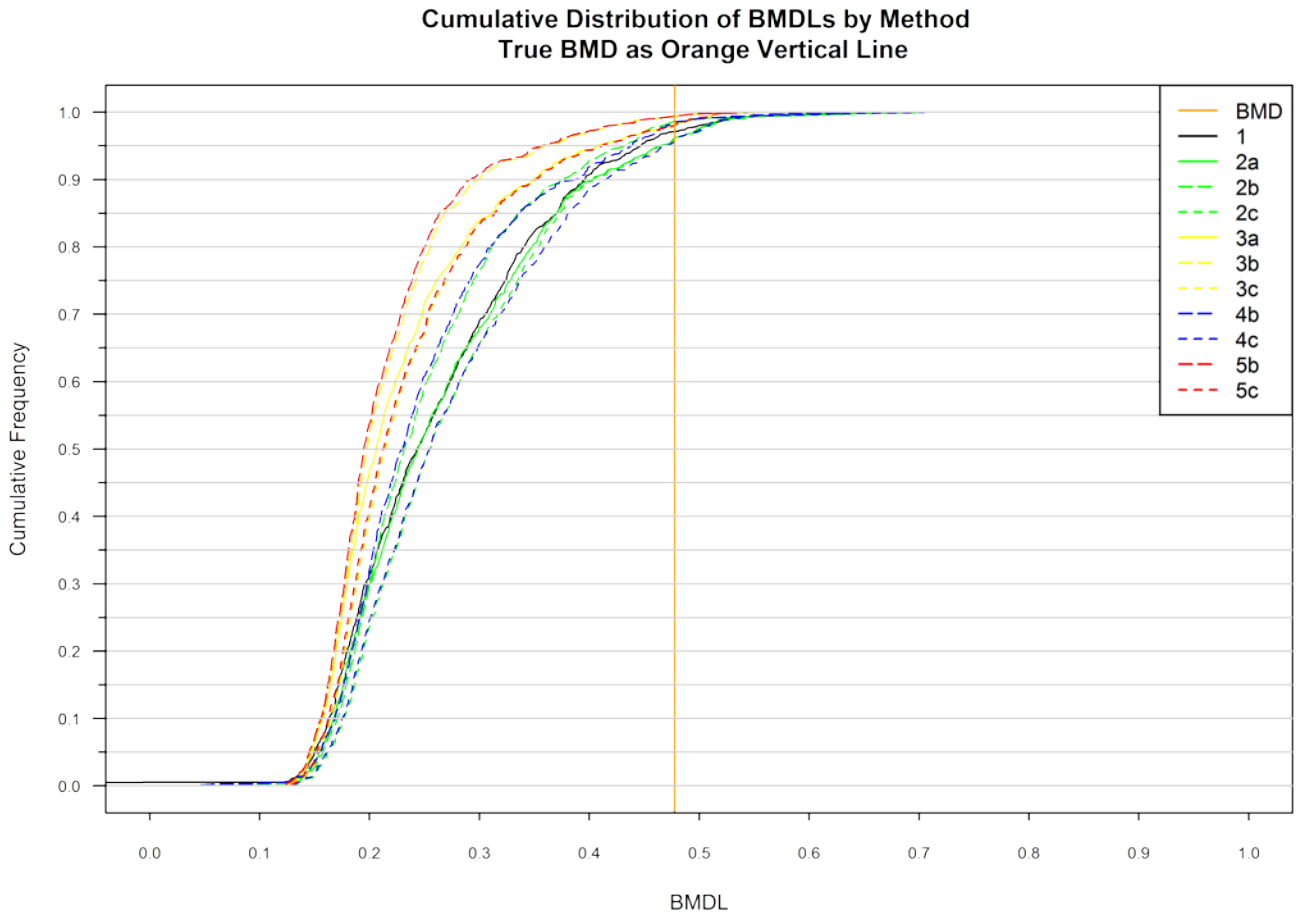
Figure 75: Template h2_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.424172	0.426072	
50	0.468375	0.478108	0.4777
75	0.505995	0.501588	
IQR	0.081823	0.0755159	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.971		
Poly3	0.965	**2a**	0.966	**3a**	0.98
Power	0.972	**2b**	0.96	**3b**	0.98
Hill	0.966	**2c**	0.966	**3c**	0.98
Exp3	0.971	**4b**	0.938	**5b**	0.98
Exp5	0.965	**4c**	0.953	**5c**	0.98

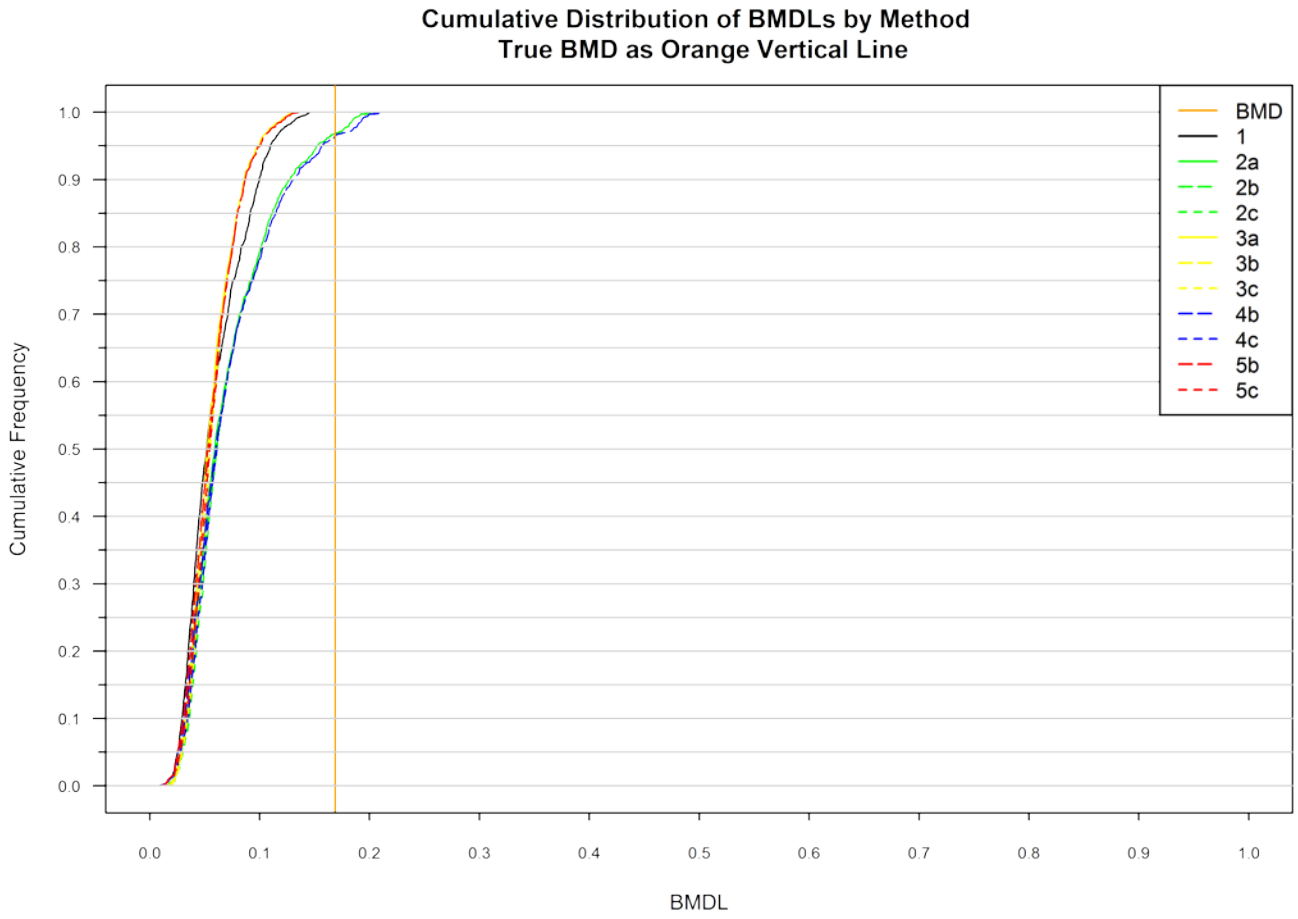
Figure 76: Template h2_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.308684	0.305247	
50	0.413429	0.41995	0.4777
75	0.51256	0.50846	
IQR	0.203875	0.203213	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.971		
Poly3	0.955	**2a**	0.962	**3a**	0.982
Power	0.965	**2b**	0.986	**3b**	0.994
Hill	0.965	**2c**	0.96	**3c**	0.981
Exp3	0.97	**4b**	0.986	**5b**	0.994
Exp5	0.97	**4c**	0.956	**5c**	0.98

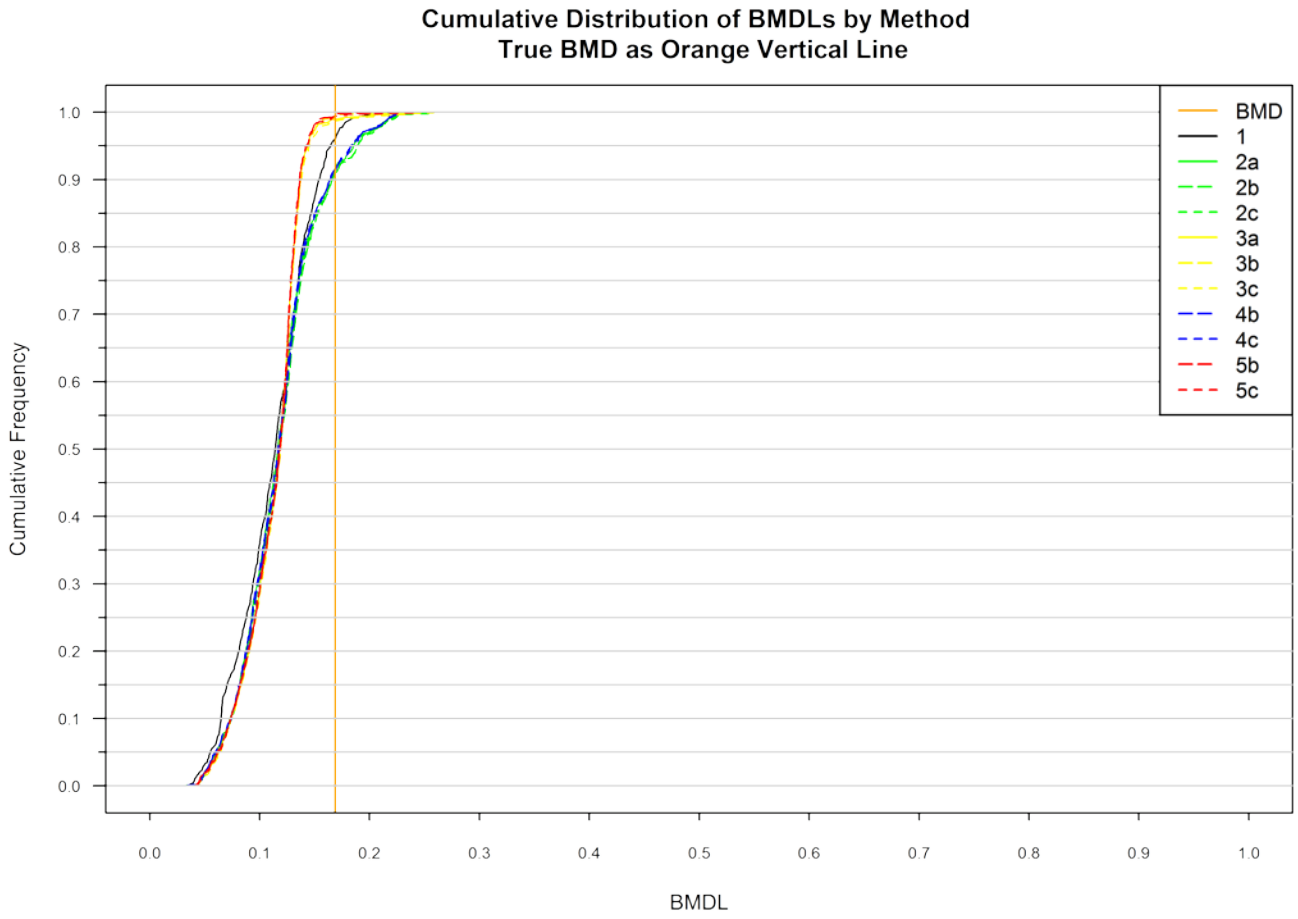
Figure 77: Template h3_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.100372	0.10284	
50	0.149778	0.154509	0.1688
75	0.19806	0.206278	
IQR	0.0976883	0.103438	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	1		
Poly3	0	**2a**	0.969	**3a**	1
Power	0	**2b**	0.968	**3b**	1
Hill	0.999	**2c**	0.968	**3c**	1
Exp3	0.064	**4b**	0.964	**5b**	1
Exp5	1	**4c**	0.964	**5c**	1

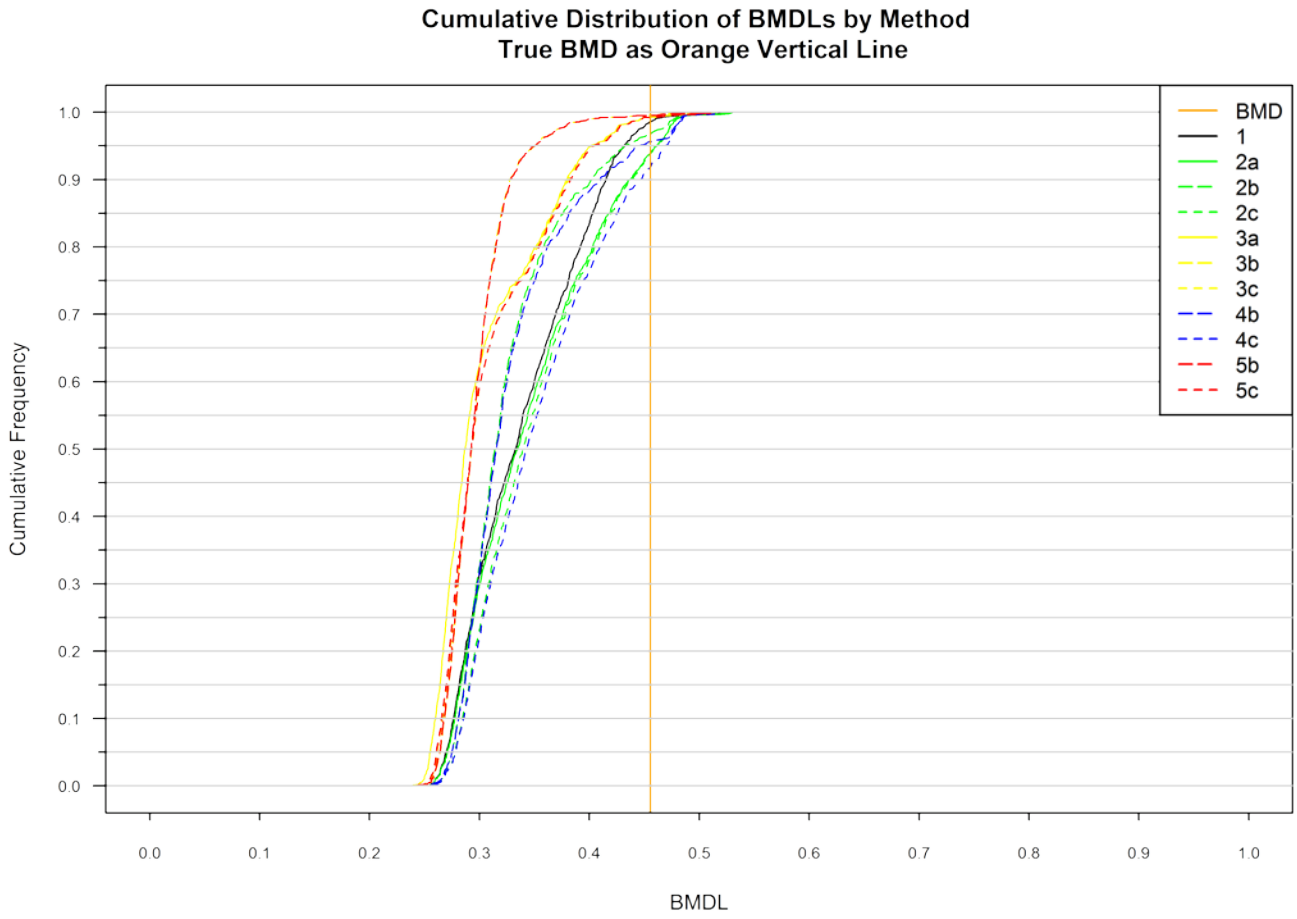
Figure 78: Template h3_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.144591	0.143486	
50	0.177564	0.176425	0.1688
75	0.224162	0.224681	
IQR	0.0795715	0.0811951	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.016	**1**	0.96		
Poly3	0.016	**2a**	0.918	**3a**	0.989
Power	0.016	**2b**	0.906	**3b**	0.992
Hill	0.978	**2c**	0.909	**3c**	0.987
Exp3	0.365	**4b**	0.913	**5b**	0.997
Exp5	0.97	**4c**	0.916	**5c**	0.994

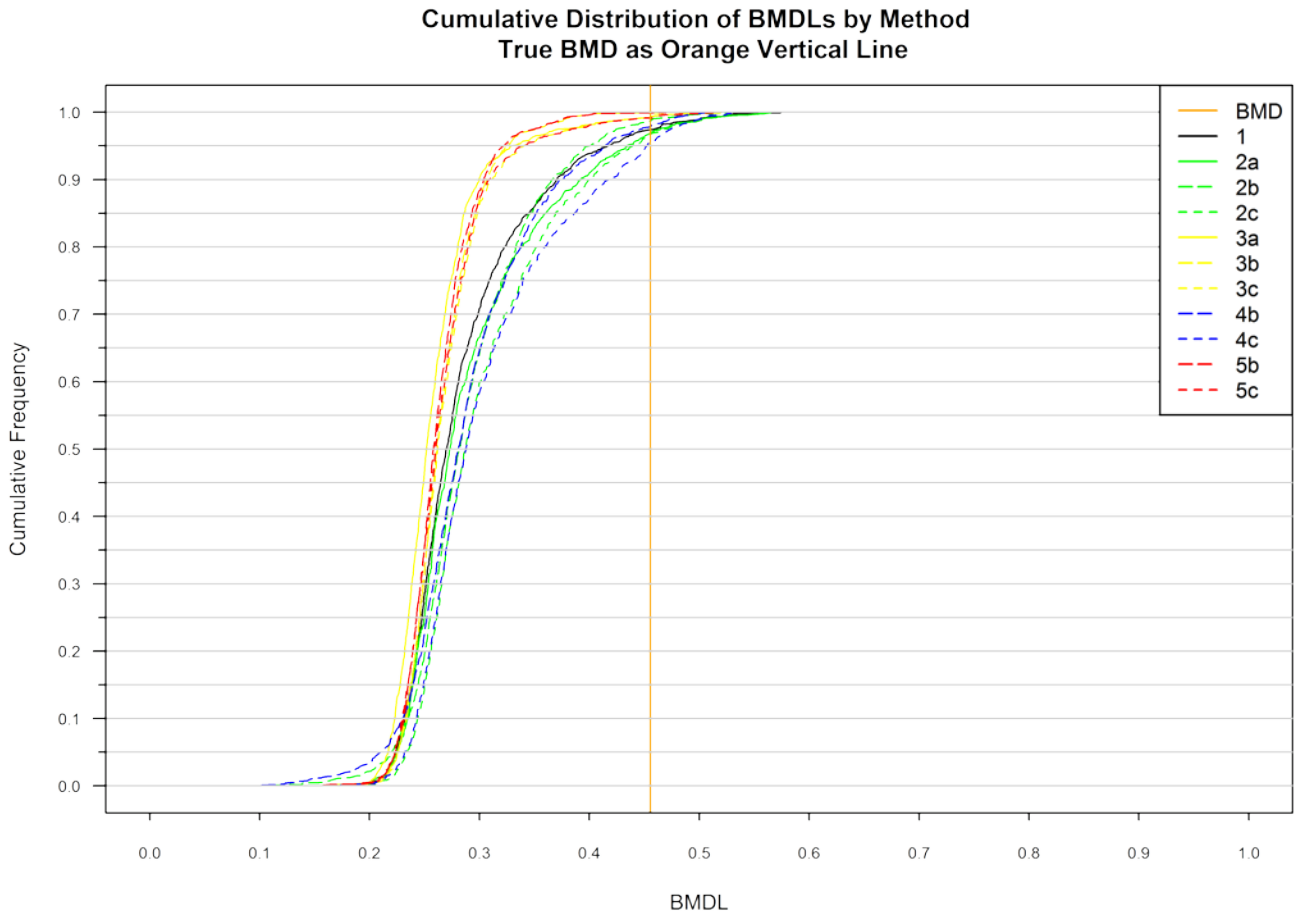
Figure 79: Template h4_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.346389	0.35019	
50	0.415115	0.424361	0.4556
75	0.472987	0.482987	
IQR	0.126598	0.132797	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.986		
Poly3	0.969	**2a**	0.941	**3a**	0.992
Power	0.987	**2b**	0.968	**3b**	0.995
Hill	0.979	**2c**	0.938	**3c**	0.992
Exp3	0.981	**4b**	0.956	**5b**	0.995
Exp5	0.971	**4c**	0.918	**5c**	0.992

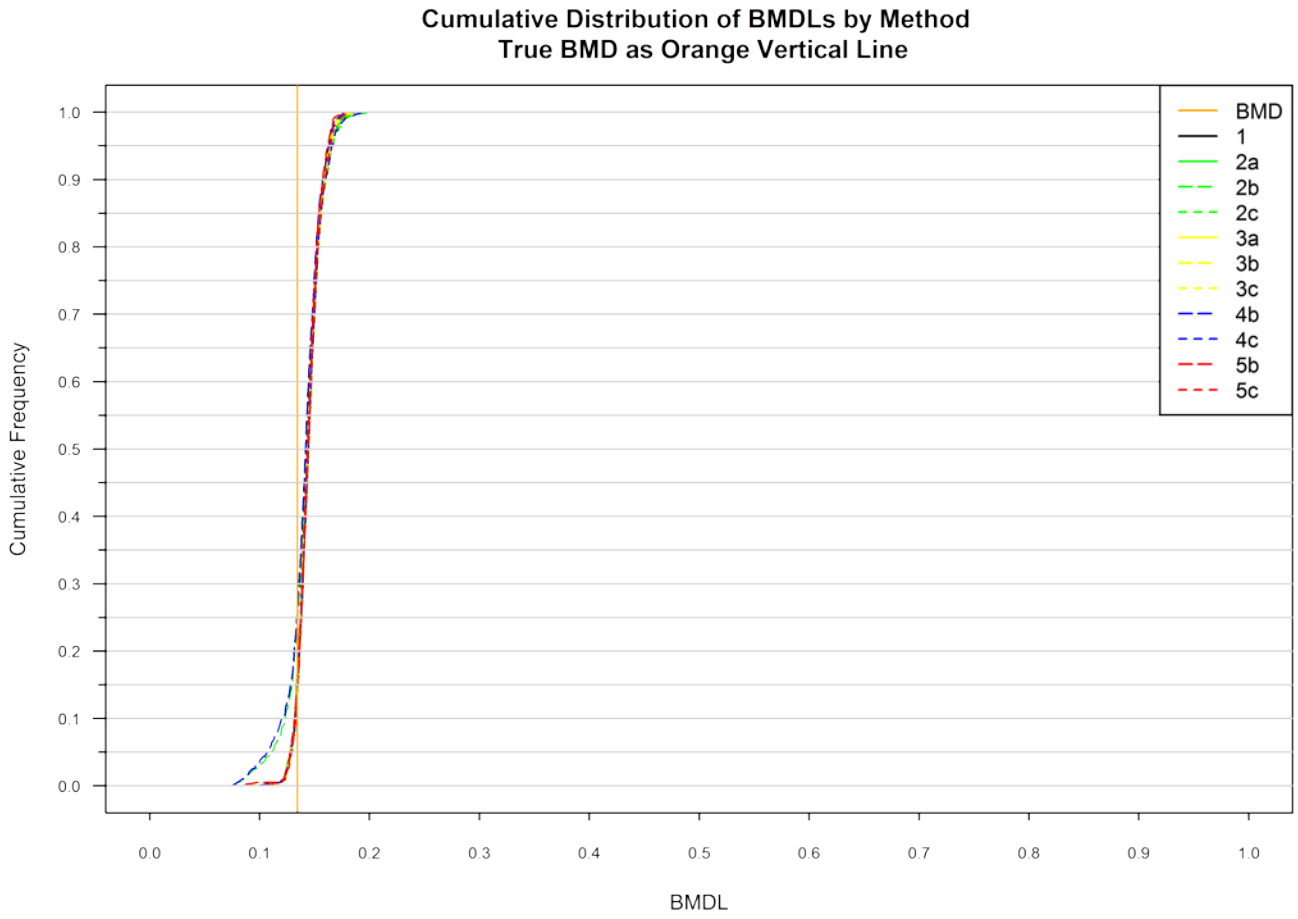
Figure 80: Template h4_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.322756	0.325515	
50	0.372426	0.388331	0.4556
75	0.449126	0.463604	
IQR	0.12637	0.13809	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.974		
Poly3	0.887	**2a**	0.968	**3a**	0.993
Power	0.97	**2b**	0.987	**3b**	0.998
Hill	0.964	**2c**	0.967	**3c**	0.992
Exp3	0.969	**4b**	0.978	**5b**	0.998
Exp5	0.968	**4c**	0.956	**5c**	0.991

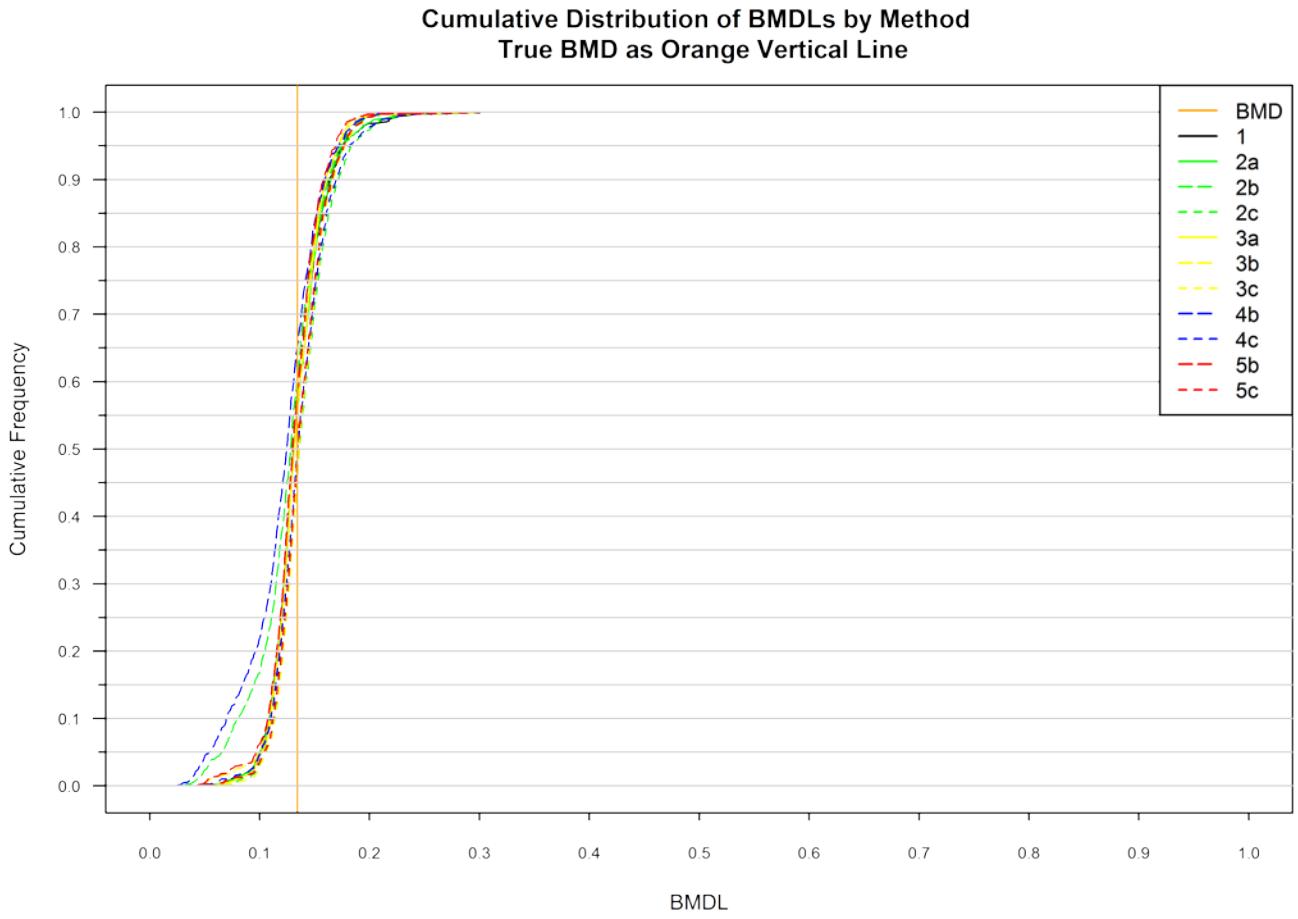
Figure 81: Template p1_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.153559	0.153268	
50	0.161469	0.161303	0.1345
75	0.169948	0.169612	
IQR	0.0163889	0.0163438	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.194	**1**	0.176		
Poly3	0.156	**2a**	0.166	**3a**	0.164
Power	0.164	**2b**	0.246	**3b**	0.156
Hill	0.604	**2c**	0.141	**3c**	0.13
Exp3	0	**4b**	0.26	**5b**	0.161
Exp5	0.694	**4c**	0.146	**5c**	0.14

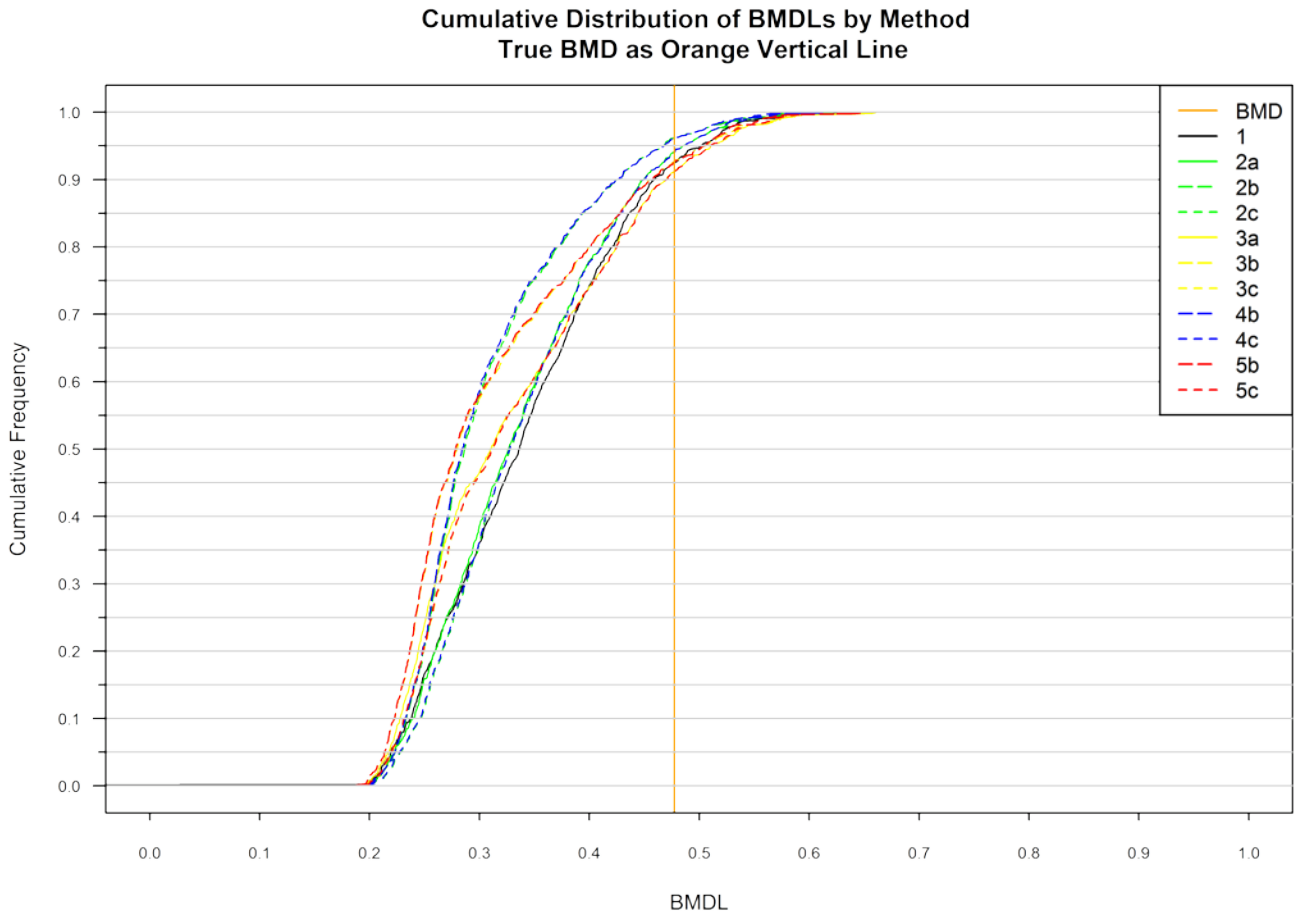
Figure 82: Template p1_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.152987	0.151599	
50	0.17129	0.170004	0.1345
75	0.198557	0.196814	
IQR	0.0455704	0.045215	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.645	**1**	0.561		
Poly3	0.548	**2a**	0.549	**3a**	0.562
Power	0.566	**2b**	0.606	**3b**	0.555
Hill	0.749	**2c**	0.462	**3c**	0.458
Exp3	0.014	**4b**	0.655	**5b**	0.59
Exp5	0.667	**4c**	0.501	**5c**	0.503

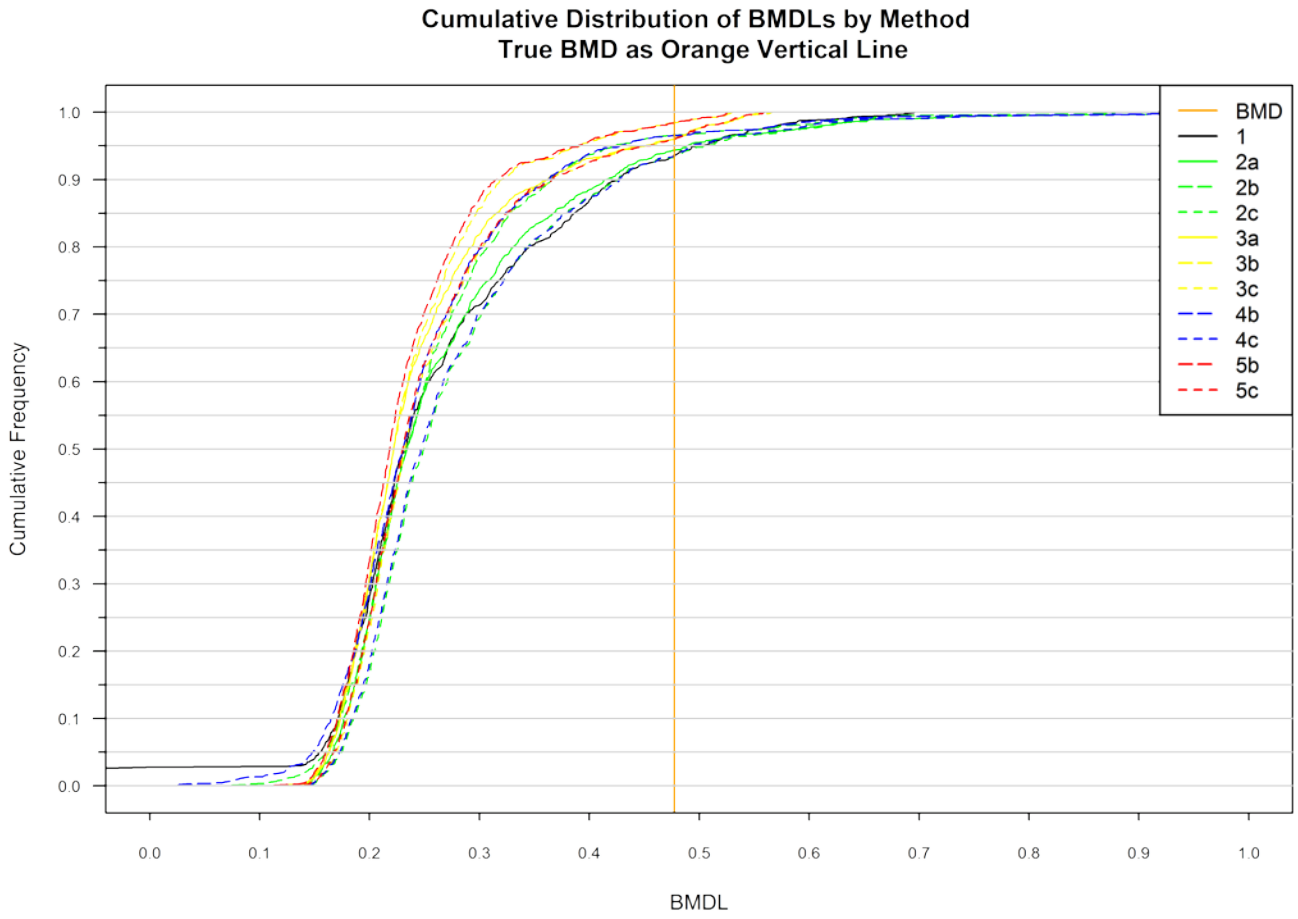
Figure 83: Template p2_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.370432	0.368339	
50	0.462292	0.462792	0.4775
75	0.537635	0.537827	
IQR	0.167203	0.169488	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.925		
Poly3	0.929	**2a**	0.94	**3a**	0.911
Power	0.919	**2b**	0.961	**3b**	0.927
Hill	0.916	**2c**	0.941	**3c**	0.911
Exp3	0.925	**4b**	0.96	**5b**	0.926
Exp5	0.923	**4c**	0.94	**5c**	0.911

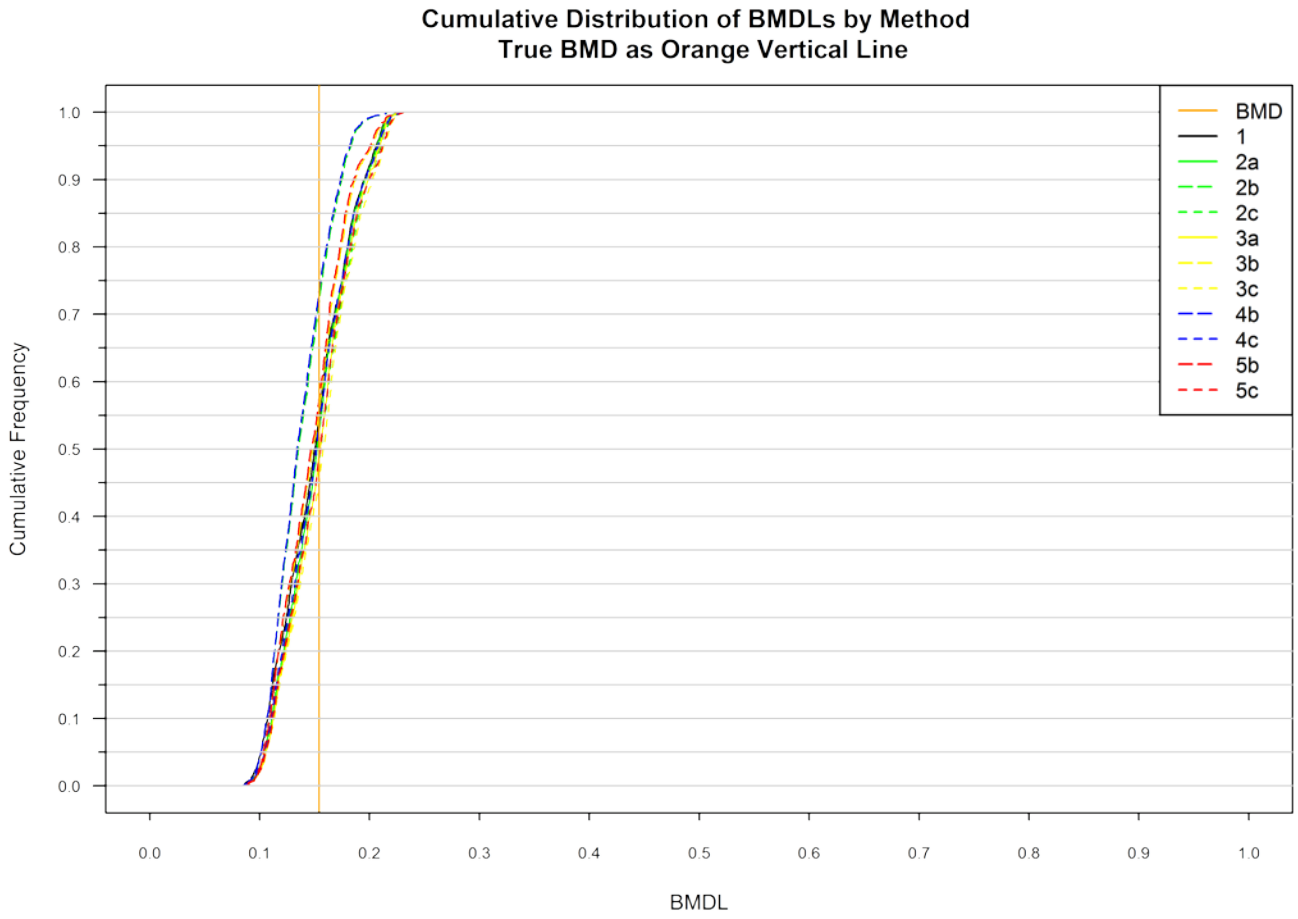
Figure 84: Template p2_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.303709	0.300626	
50	0.408924	0.410344	0.4775
75	0.57092	0.570033	
IQR	0.267211	0.269406	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.936		
Poly3	0.912	**2a**	0.943	**3a**	0.96
Power	0.896	**2b**	0.966	**3b**	0.984
Hill	0.897	**2c**	0.934	**3c**	0.959
Exp3	0.924	**4b**	0.965	**5b**	0.984
Exp5	0.926	**4c**	0.936	**5c**	0.959

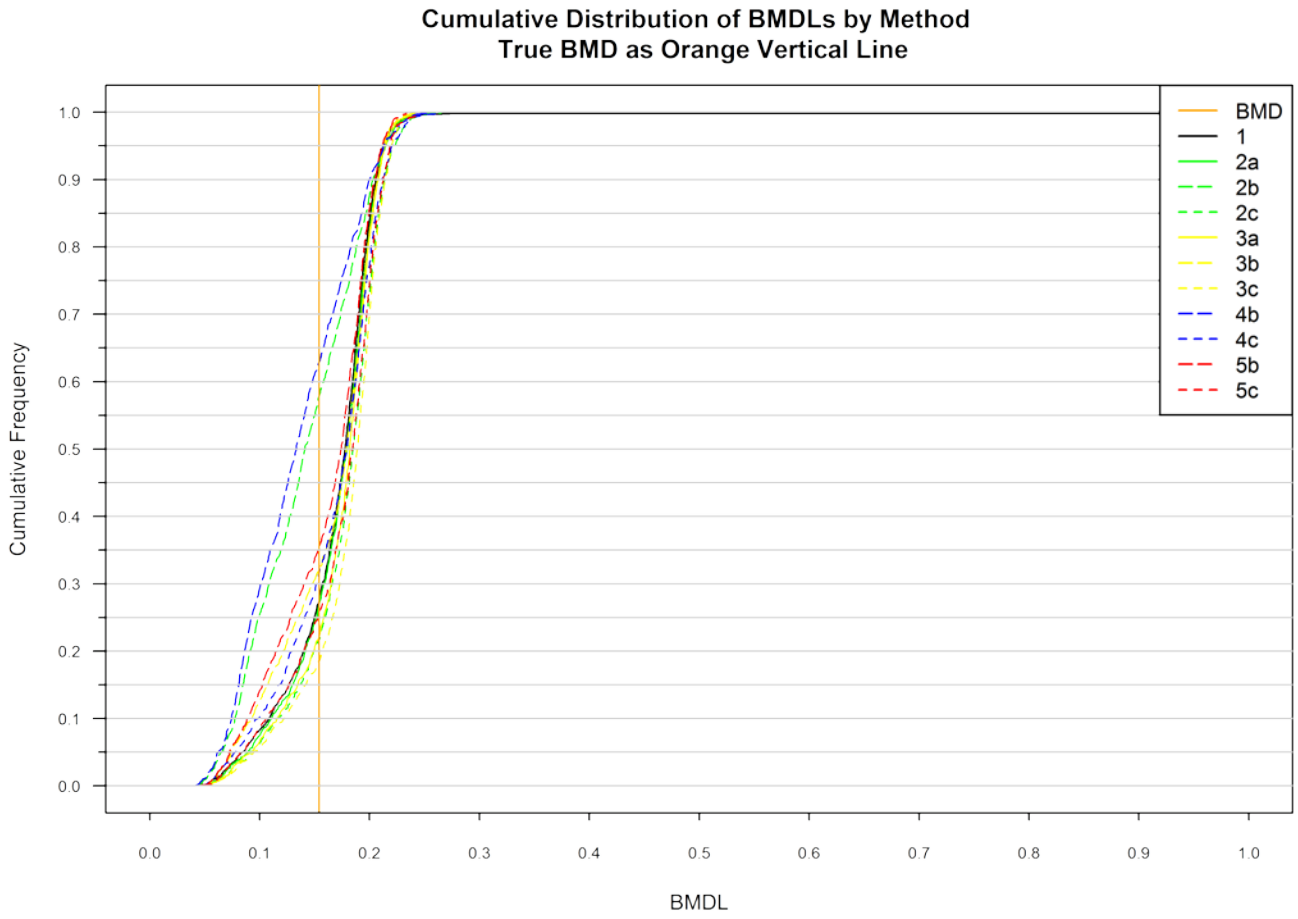
Figure 85: Template p3_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.159689	0.158098	
50	0.181911	0.180631	0.1541
75	0.202974	0.200859	
IQR	0.0432851	0.0427616	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.543		
Poly3	0	**2a**	0.527	**3a**	0.49
Power	0	**2b**	0.715	**3b**	0.553
Hill	0.879	**2c**	0.481	**3c**	0.451
Exp3	0.042	**4b**	0.735	**5b**	0.577
Exp5	0.885	**4c**	0.539	**5c**	0.489

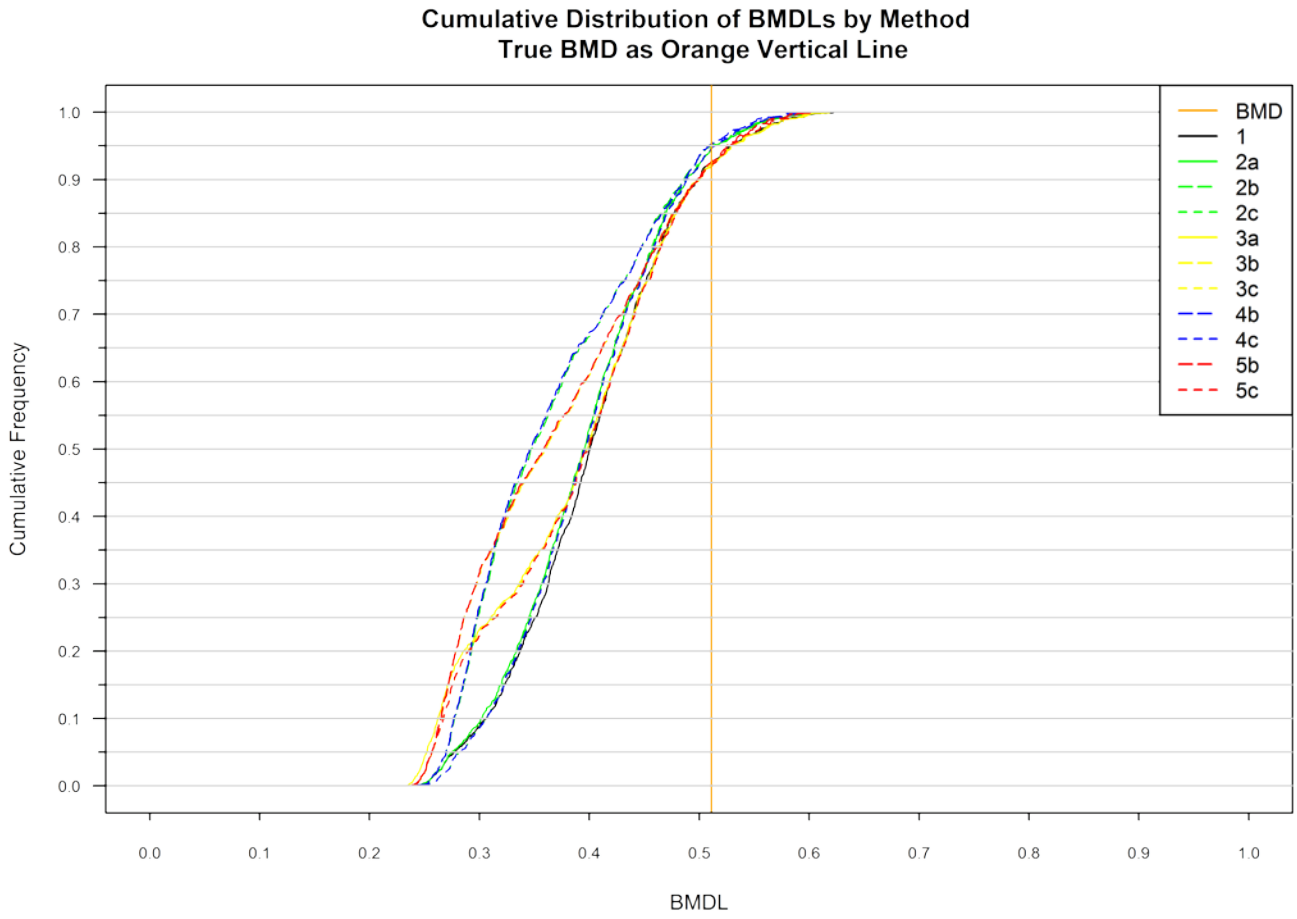
Figure 86: Template p3_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.18706	0.179529	
50	0.210786	0.207417	0.1541
75	0.229675	0.228489	
IQR	0.0426157	0.0489601	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.276		
Poly3	0	**2a**	0.27	**3a**	0.222
Power	0	**2b**	0.579	**3b**	0.321
Hill	0.848	**2c**	0.217	**3c**	0.186
Exp3	0.56	**4b**	0.629	**5b**	0.355
Exp5	0.85	**4c**	0.317	**5c**	0.256

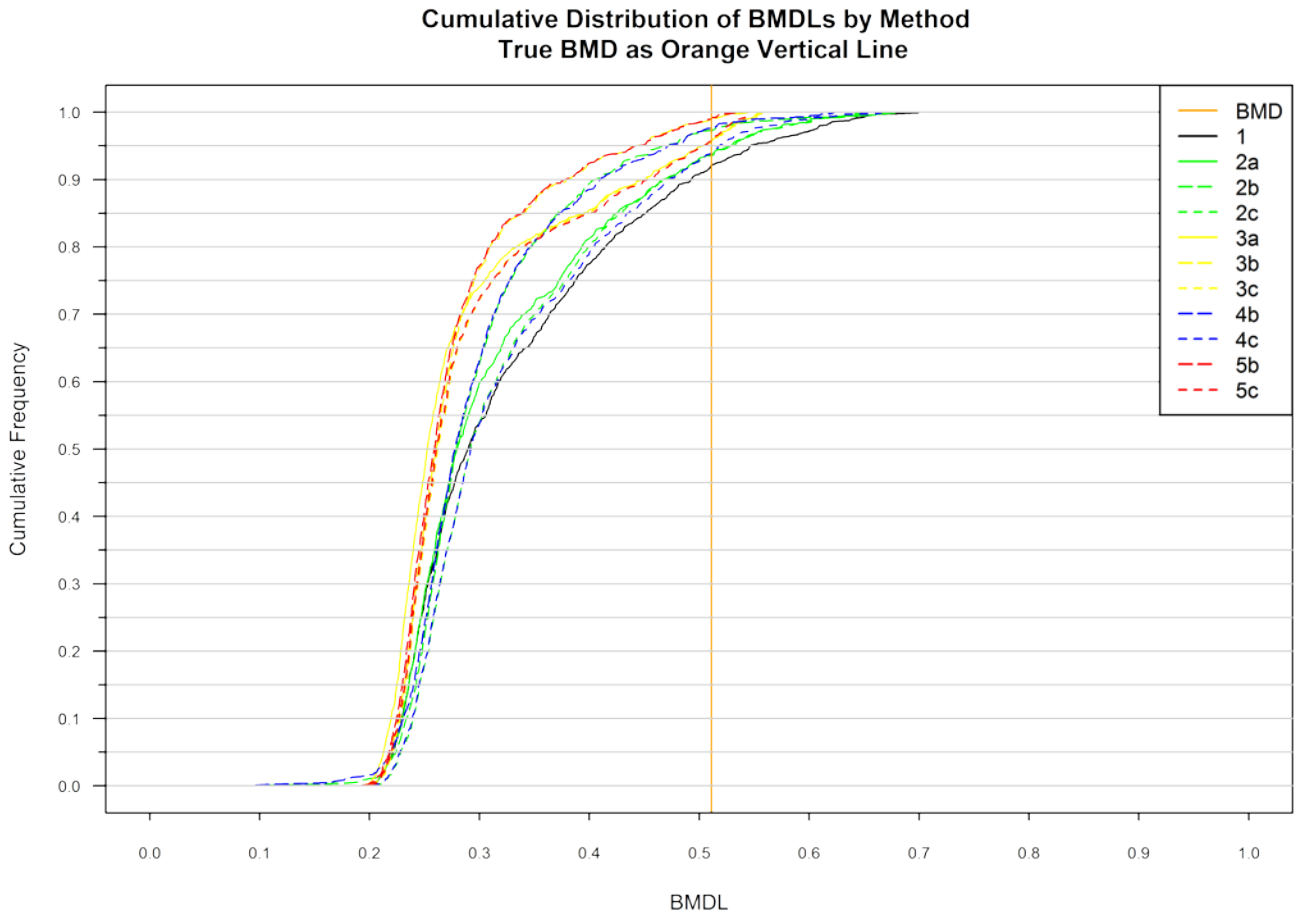
Figure 87: Template p4_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.450839	0.450706	
50	0.509469	0.509173	0.5112
75	0.565634	0.559834	
IQR	0.114795	0.109128	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.925		
Poly3	0.915	**2a**	0.946	**3a**	0.919
Power	0.926	**2b**	0.949	**3b**	0.923
Hill	0.935	**2c**	0.949	**3c**	0.92
Exp3	0.91	**4b**	0.952	**5b**	0.926
Exp5	0.934	**4c**	0.945	**5c**	0.919

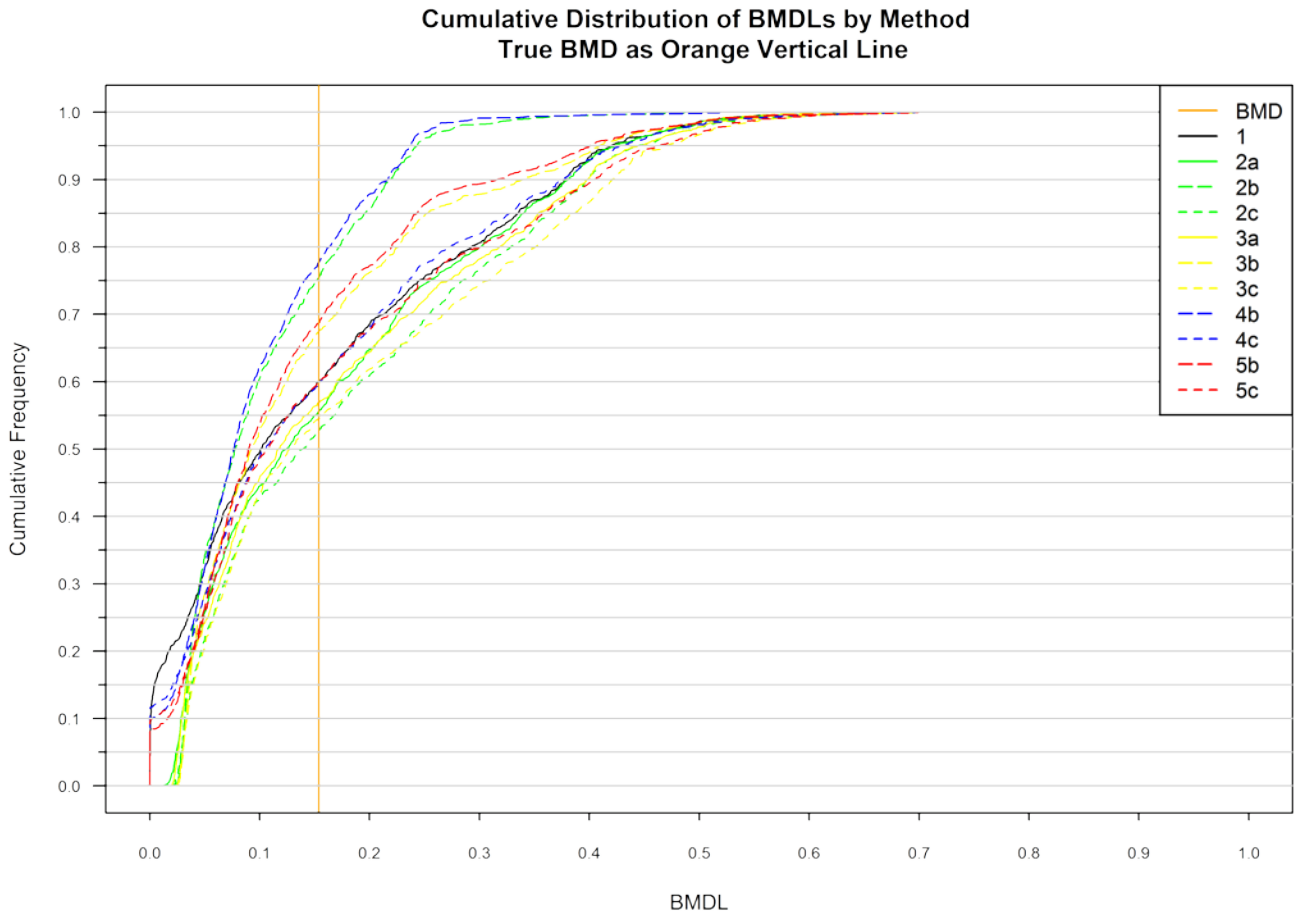
Figure 88: Template p4_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.342916	0.341238	
50	0.440979	0.445106	0.5112
75	0.583854	0.576371	
IQR	0.240937	0.235134	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	1	**1**	0.921		
Poly3	0.908	**2a**	0.937	**3a**	0.958
Power	0.911	**2b**	0.972	**3b**	0.989
Hill	0.932	**2c**	0.935	**3c**	0.956
Exp3	0.909	**4b**	0.976	**5b**	0.989
Exp5	0.916	**4c**	0.938	**5c**	0.959

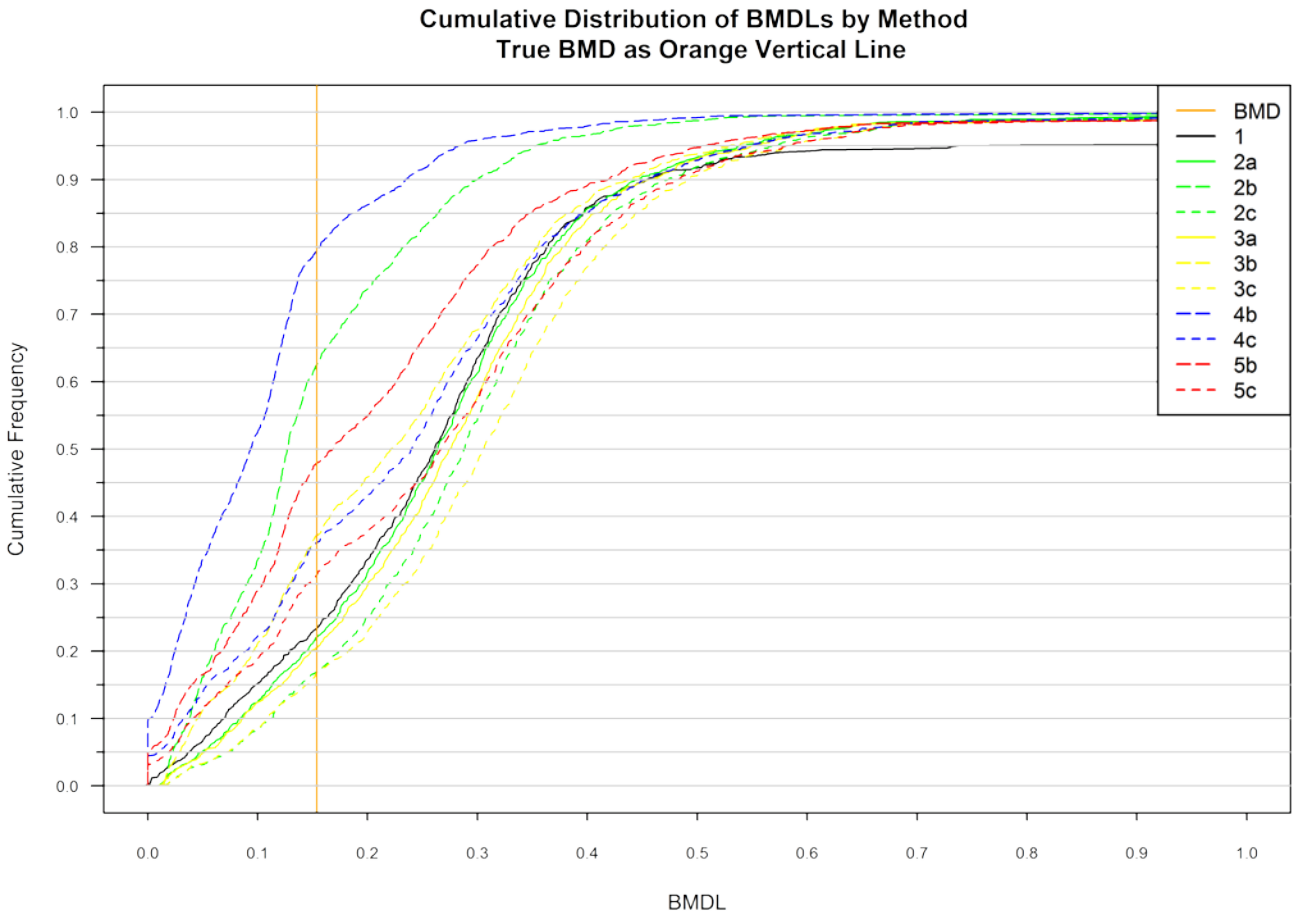
Figure 89: Template e1_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.171188	0.160622	
50	0.23601	0.234485	0.154
75	0.366385	0.317186	
IQR	0.195198	0.156564	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.601		
Poly3	0	**2a**	0.557	**3a**	0.569
Power	0	**2b**	0.754	**3b**	0.675
Hill	0.905	**2c**	0.528	**3c**	0.545
Exp3	0	**4b**	0.777	**5b**	0.687
Exp5	0.901	**4c**	0.596	**5c**	0.6

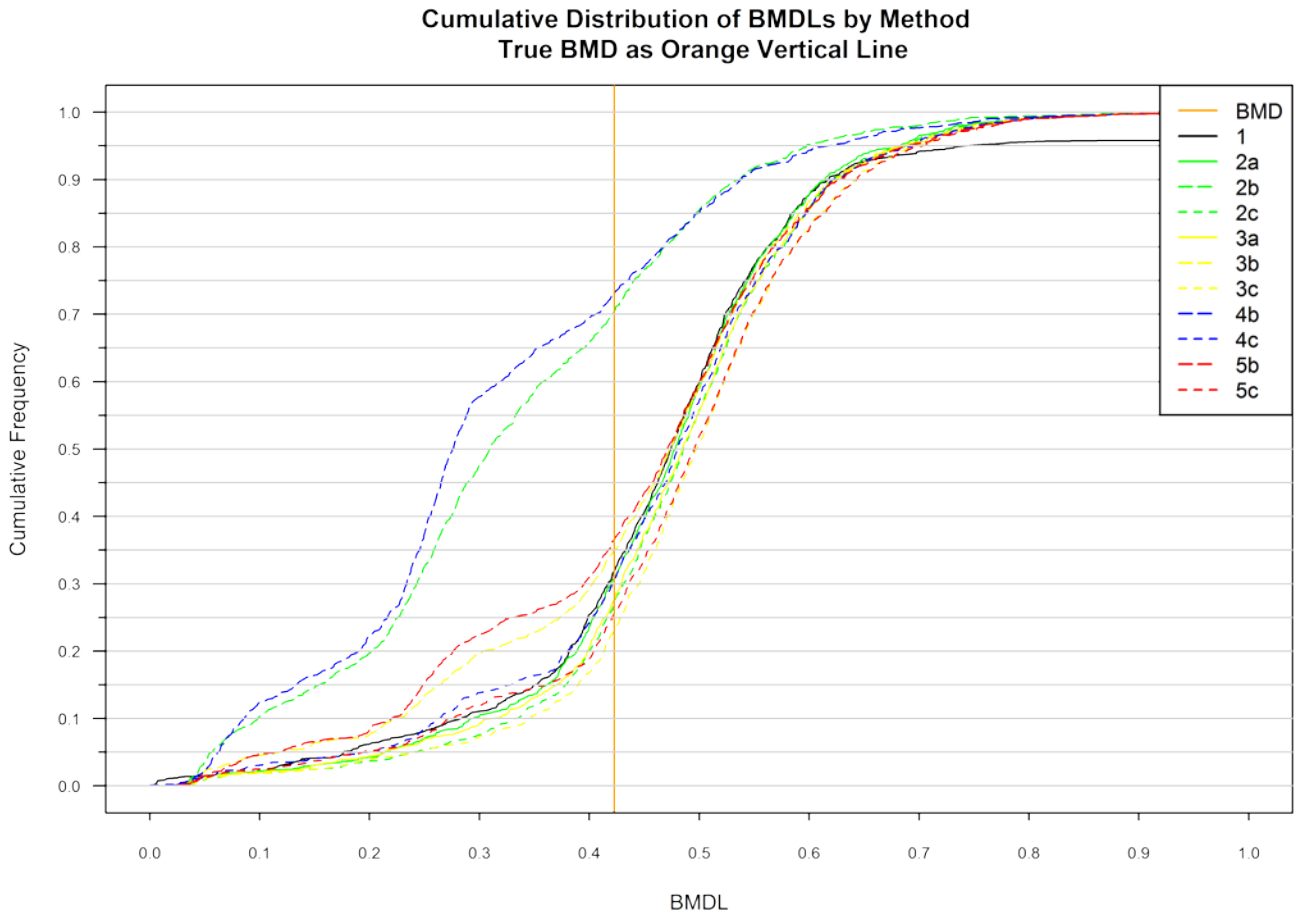
Figure 90: Template e1_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.289409	0.194234	
50	0.452251	0.411534	0.154
75	0.64184	0.647884	
IQR	0.352431	0.453649	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.004	**1**	0.234		
Poly3	0.005	**2a**	0.221	**3a**	0.206
Power	0.002	**2b**	0.625	**3b**	0.372
Hill	0.847	**2c**	0.171	**3c**	0.168
Exp3	0.01	**4b**	0.795	**5b**	0.479
Exp5	0.896	**4c**	0.361	**5c**	0.313

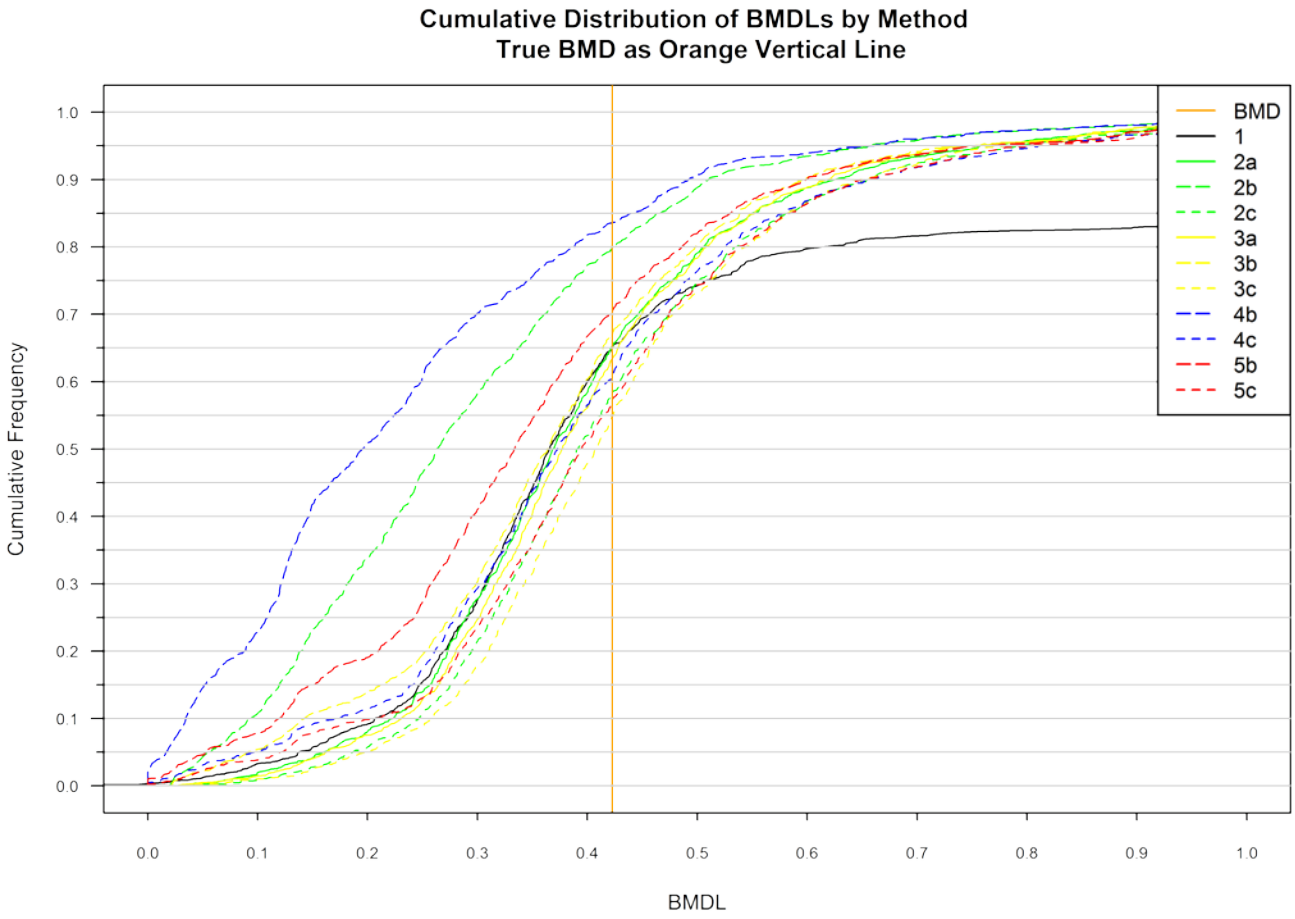
Figure 91: Template e2_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.540035	0.537044	
50	0.645226	0.65591	0.4225
75	0.771952	0.789654	
IQR	0.231917	0.25261	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.119	**1**	0.318		
Poly3	0.116	**2a**	0.308	**3a**	0.277
Power	0.117	**2b**	0.704	**3b**	0.351
Hill	0.898	**2c**	0.271	**3c**	0.229
Exp3	0.036	**4b**	0.732	**5b**	0.365
Exp5	0.941	**4c**	0.304	**5c**	0.258

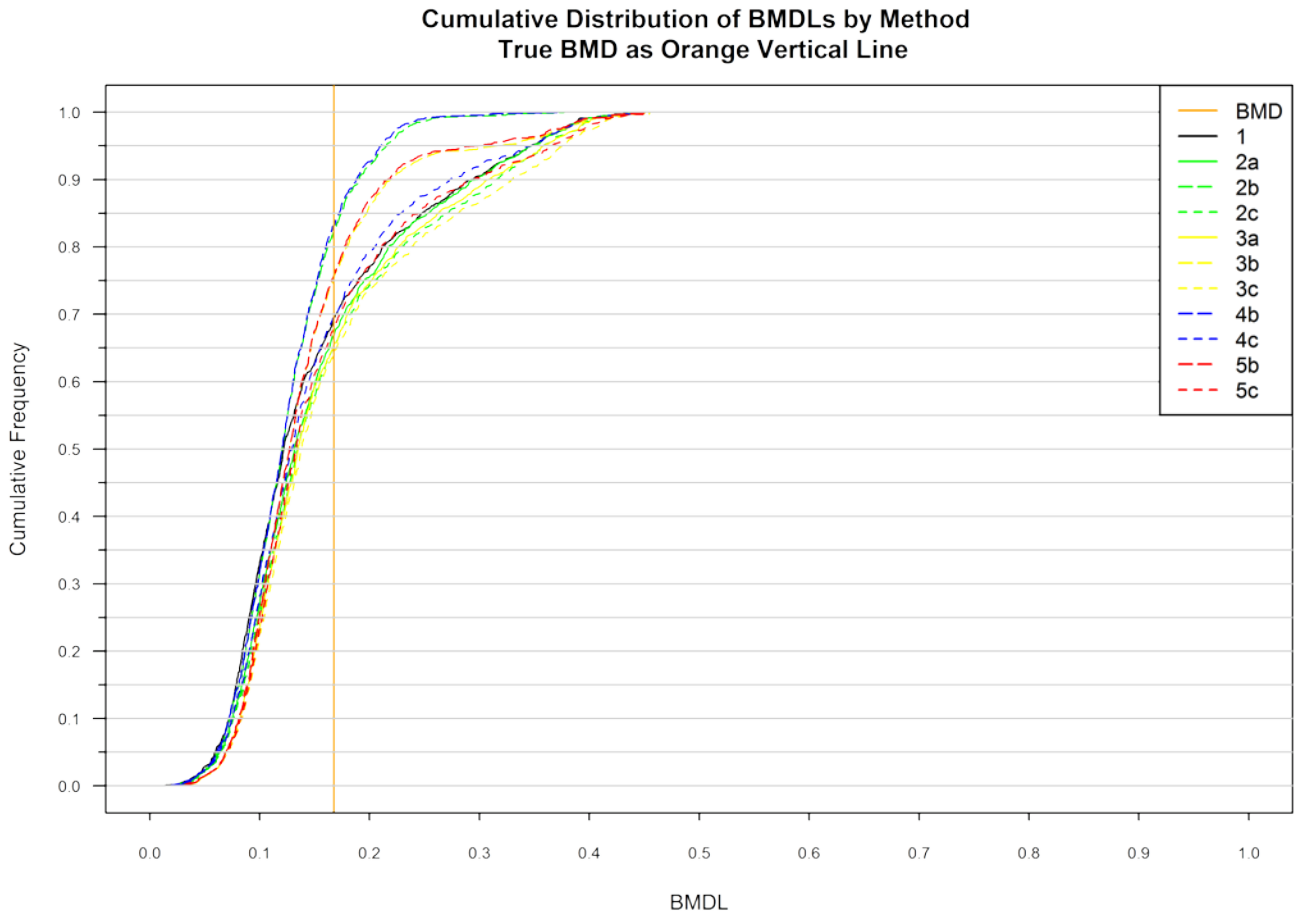
Figure 92: Template e2_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.486407	0.482779	
50	0.680443	0.700083	0.4225
75	1.07779	1.0326	
IQR	0.59138	0.549819	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.571	**1**	0.651		
Poly3	0.535	**2a**	0.653	**3a**	0.634
Power	0.546	**2b**	0.797	**3b**	0.67
Hill	0.737	**2c**	0.581	**3c**	0.548
Exp3	0.468	**4b**	0.836	**5b**	0.702
Exp5	0.899	**4c**	0.609	**5c**	0.573

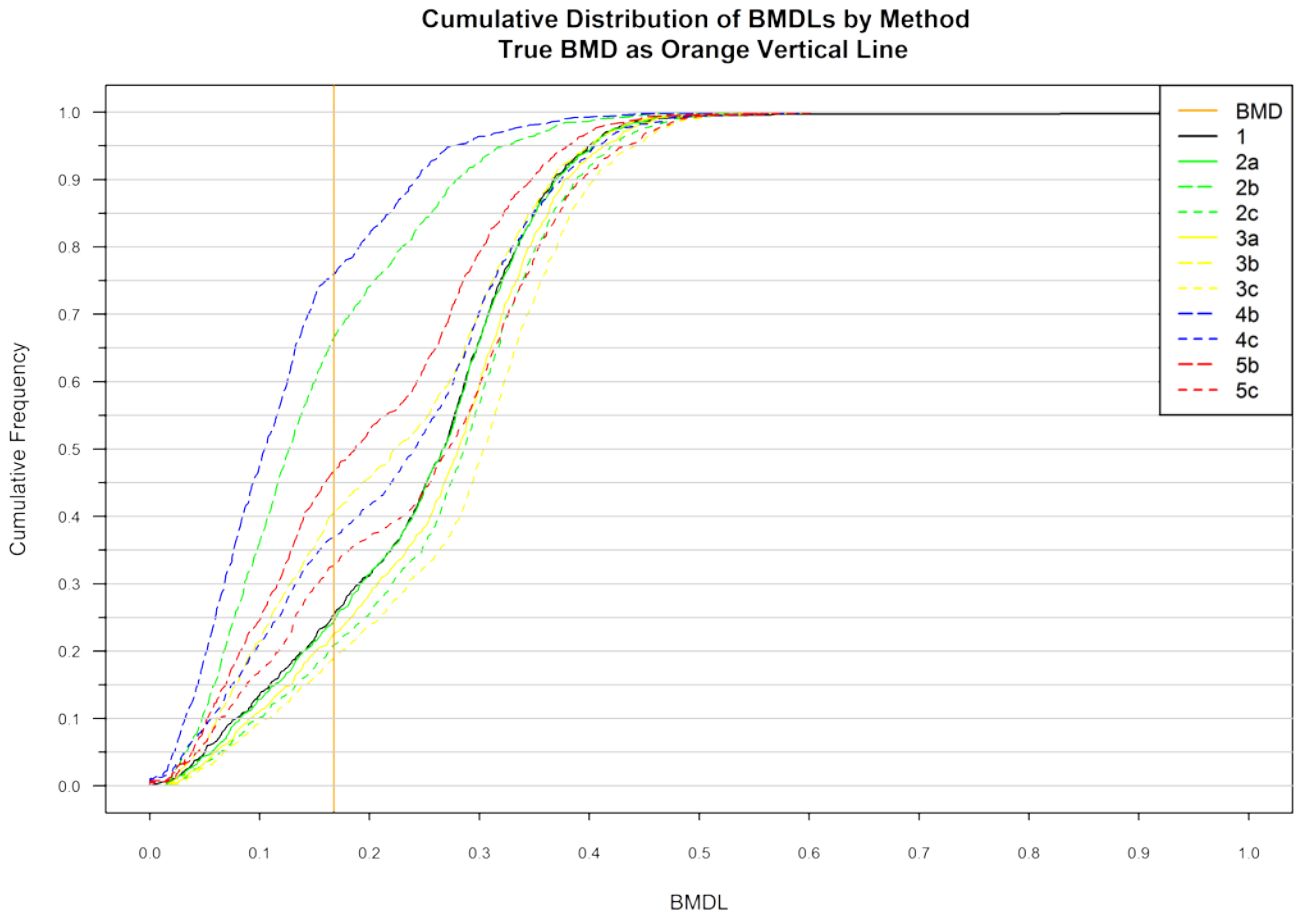
Figure 93: Template e3_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.155965	0.15494	
50	0.202898	0.199268	0.1675
75	0.261512	0.24923	
IQR	0.105547	0.0942898	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.694		
Poly3	0	**2a**	0.668	**3a**	0.654
Power	0	**2b**	0.824	**3b**	0.751
Hill	0.88	**2c**	0.65	**3c**	0.641
Exp3	0.004	**4b**	0.832	**5b**	0.759
Exp5	0.882	**4c**	0.698	**5c**	0.68

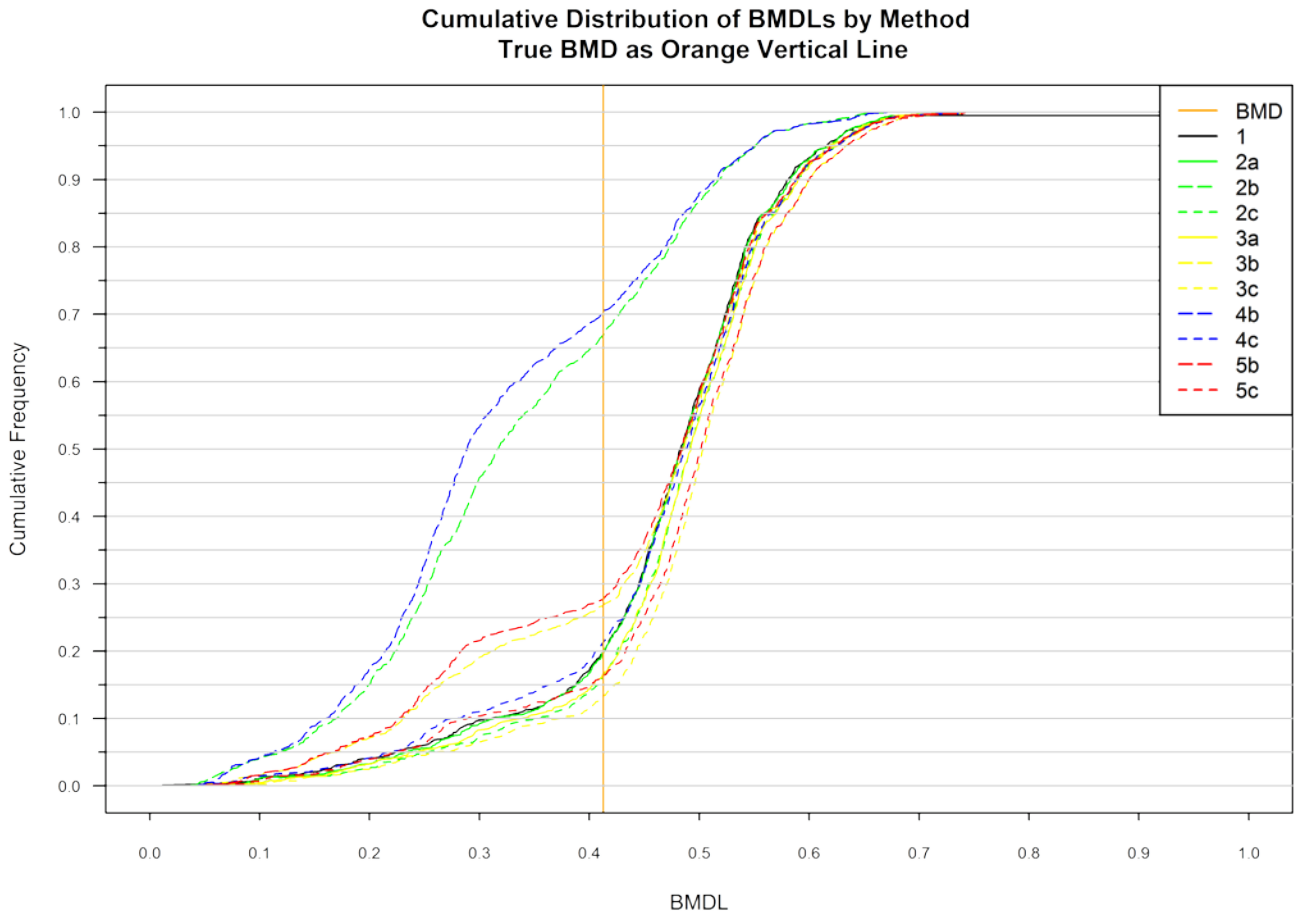
Figure 94: Template e3_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.264619	0.209358	
50	0.368416	0.340183	0.1675
75	0.454675	0.436225	
IQR	0.190055	0.226867	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0	**1**	0.254		
Poly3	0	**2a**	0.244	**3a**	0.225
Power	0	**2b**	0.665	**3b**	0.405
Hill	0.862	**2c**	0.209	**3c**	0.189
Exp3	0.008	**4b**	0.76	**5b**	0.466
Exp5	0.875	**4c**	0.369	**5c**	0.329

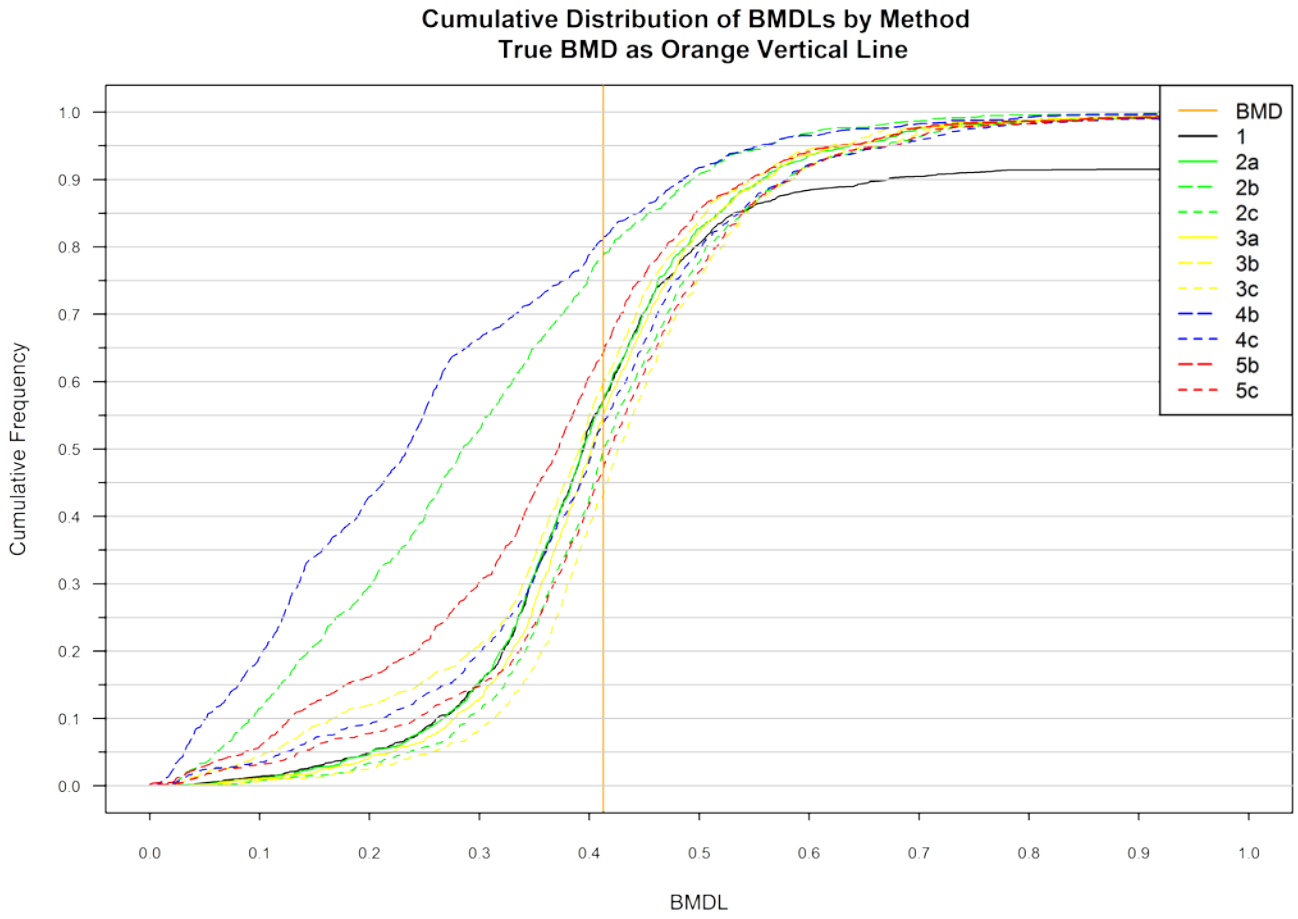
Figure 95: Template e4_lognormal_chronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.533157	0.526414	
50	0.60137	0.60161	0.4126
75	0.680067	0.681783	
IQR	0.14691	0.155369	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.011	**1**	0.198		
Poly3	0.011	**2a**	0.198	**3a**	0.167
Power	0.011	**2b**	0.669	**3b**	0.268
Hill	0.938	**2c**	0.164	**3c**	0.135
Exp3	0.093	**4b**	0.704	**5b**	0.278
Exp5	0.939	**4c**	0.213	**5c**	0.165

Figure 96: Template e4_lognormal_subchronic; Models fit assuming non-constant variance



Percentiles and Inter-quartile Range for BMD Estimates	BMD-Averaging Methods	Model-Averaging Methods	True BMD
25	0.486769	0.473158	
50	0.60089	0.60175	0.4126
75	0.780708	0.81835	
IQR	0.293939	0.345191	

Model	Coverage	Method	Coverage	Method	Coverage
Linear	0.41	**1**	0.572		
Poly3	0.382	**2a**	0.572	**3a**	0.551
Power	0.384	**2b**	0.787	**3b**	0.599
Hill	0.798	**2c**	0.498	**3c**	0.433
Exp3	0.548	**4b**	0.811	**5b**	0.643
Exp5	0.899	**4c**	0.537	**5c**	0.471

APPENDIX A. BIBLIOGRAPHY OF LITERATURE ON MODEL AVERAGING

Bailer, A. J., Noble, R. B., & Wheeler, M. W. (2005). Model uncertainty and risk estimation for experimental studies of quantal responses. *Risk Analysis*, *25*(2), 291-299.

Barbieri, M. M., & Berger, J. O. (2004). Optimal predictive model selection. *Annals of Statistics*, 870-897.

Brown, P. J., Vannucci, M., & Fearn, T. (2002). Bayes model averaging with selection of regressors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(3), 519-536.

Buchholz, A., Holländer, N., & Sauerbrei, W. (2008). On properties of predictors derived with a two-step bootstrap model averaging approach—a simulation study in the linear regression model. *Computational Statistics & Data Analysis*, *52*(5), 2778-2793.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.

Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging* (Vol. 330). Cambridge: Cambridge University Press.

Clyde, M. A. (1999). Bayesian model averaging and model search strategies. *Bayesian statistics*, *6*, 157.

Clyde, M. A., Ghosh, J., & Littman, M. L. (2011). Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics*, *20*(1).

de-G. Acquah, H. (2010). Comparison of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in Selection of an Asymmetric Price Relationship. *J. Develop. Agricult. Econom.* **2**, 1-6.

Domingos, P. (1997). Why Does Bagging Work? A Bayesian Account and its Implications. In *KDD* (pp. 155-158). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.40.1298&rep=rep1&type=pdf

Domingos, P. (2000). Bayesian averaging of classifiers and the overfitting problem. In *ICML* (pp. 223-230). http://homes.cs.washington.edu/~pedrod/papers/mlc00b.pdf

Faes, C., Aerts, M., Geys, H., & Molenberghs, G. (2007). Model averaging using fractional polynomials to estimate a safe level of exposure. *Risk Analysis*, *27*(1), 111-123.

Fraley, C., Raftery, A. E., & Gneiting, T. (2010). Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Monthly Weather Review*, *138*(1), 190-202.

George, E. (1999). Discussion of "Model averaging and model search strategies" by M. Clyde. In *Bayesian Statistics 6–Proceedings of the Sixth Valencia International Meeting*.

Greenland, S. (1999). Multilevel modeling and model averaging. *Scandinavian journal of work, environment & health*, 43-48.

Hansen, B. E. (2007). Least Squares Model Averaging. *Econometrica*, 1175-1189.

Hansen, B. E., & Racine, J. S. (2012). Jackknife model averaging. *Journal of Econometrics*, *167*(1), 38-46.

Hjort, N. L., & Claeskens, G. (2003). Frequentist model average estimators. *Journal of the American Statistical Association*, *98*(464), 879-899.

Hjort, N. L., & Claeskens, G. (2006). Focused information criteria and model averaging for the Cox hazard regression model. *Journal of the American Statistical Association*, *101*(476), 1449-1464.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382-401.

Kang, S. H., Kodell, R. L., & Chen, J. J. (2000). Incorporating model uncertainties along with data uncertainties in microbial risk assessment. *Regulatory Toxicology and Pharmacology*, *32*(1), 68-72.

Kapetanios, G., Labhard, V., & Price, S. (2008). Forecasting using Bayesian and information-theoretic model averaging: An application to UK inflation. *Journal of Business & Economic Statistics*, *26*(1), 33-41.

Kim, S. B., Kodell, R. L., & Moon, H. (2014). A diversity index for model space selection in the estimation of benchmark and infectious doses via model averaging. *Risk Analysis*, *34*(3), 453-464.

Leeb, H., & Pötscher, B. M. (2006). Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics*, 2554-2591.

Ley, E., & Steel, M. F. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of applied econometrics*, *24*(4), 651-674.

Liang, H., Zou, G., Wan, A. T., & Zhang, X. (2011). Optimal weight choice for frequentist model average estimators. *Journal of the American Statistical Association*, *106*(495).

Magnus, J. R., Powell, O., & Prüfer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics*, *154*(2), 139-153.

Minka, T. P. (2000). Bayesian model averaging is not model combination. *Available electronically at http://www. stat. cmu.edu/minka/papers/bma.html*, 1-2.

Montgomery, J. M., Hollenbach, F. M., & Ward, M. D. (2012). Improving predictions using ensemble Bayesian model averaging. *Political Analysis*, *20*(3), 271-291.

Montgomery, J. M., & Nyhan, B. (2010). Bayesian model averaging: Theoretical developments and practical applications. *Political Analysis*, *18*(2), 245-270.

Moon, H., Kim, H. J., Chen, J. J., & Kodell, R. L. (2005). Model averaging using the Kullback information criterion in estimating effective doses for microbial infection and illness. *Risk Analysis*, *25*(5), 1147-1159.

Moon, H., Kim, S. B., Chen, J. J., George, N. I., & Kodell, R. L. (2013). Model uncertainty and model averaging in the estimation of infectious doses for microbial pathogens. *Risk Analysis*, *33*(2), 220-231.

Morales, K. H., Ibrahim, J. G., Chen, C. J., & Ryan, L. M. (2006). Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. *Journal of the American Statistical Association*, *101*(473), 9-17.

Namata, H., Aerts, M., Faes, C., & Teunis, P. (2008). Model averaging in microbial risk assessment using fractional polynomials. *Risk Analysis*, *28*(4), 891-905.

Neuman, S. P. (2003). Maximum likelihood Bayesian averaging of uncertain model predictions. *Stochastic Environmental Research and Risk Assessment*, *17*(5), 291-305.

Parrish, M. A., Moradkhani, H., & DeChant, C. M. (2012). Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. *Water Resources Research*, *48*(3).

Piegorsch, W. W. (2014). Model Uncertainty in Environmental Dose–Response Risk Analysis. *Statistics and Public Policy*, *1*(1), 78-85.

Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, *53*(5), 793-808.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179-191.

Rajagopal, R., & Del Castillo, E. (2005). Model-robust process optimization using Bayesian model averaging. *Technometrics*, *47*(2), 152-163.

Rao, J. S., & Tibshirani, R. (1997). The out-of-bootstrap method for model averaging and selection. *University of Toronto*.
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.35.7953&rep=rep1&type=pdf

Richards, S. A., Whittingham, M. J., & Stephens, P. A. (2011). Model selection and model averaging in behavioral ecology: the utility of the IT-AIC framework. *Behavioral Ecology and Sociobiology*, *65*(1), 77-89.

Ritz, C., Gerhard, D., & Hothorn, L. A. (2013). A unified framework for benchmark dose estimation applied to mixed models and model averaging. *Statistics in Biopharmaceutical Research*, *5*(1), 79-90.

Schomaker, M., Wan, A. T., & Heumann, C. (2010). Frequentist model averaging with missing observations. *Computational Statistics & Data Analysis*, *54*(12), 3336-3347.

Singh, A., Mishra, S., & Ruskauff, G. (2010). Model averaging techniques for quantifying conceptual model uncertainty. *Groundwater*, *48*(5), 701-715.

Shao, K. (2012). A comparison of three methods for integrating historical information for Bayesian model averaged benchmark dose estimation. *Environmental toxicology and pharmacology*, *34*(2), 288-296.

Shao, K., & Gift, J. S. (2014). Model uncertainty and Bayesian model averaged benchmark dose estimation for continuous data. *Risk Analysis*, *34*(1), 101-120.

Shao, K. & Small. M. J. (2011). Potential Uncertainty Reduction in Model-Averaged Benchmark Dose Estimates Informed by an Additional Dose Study. *Risk Analysis* **31**, 1561-1575.

Shao, K., & Small, M. J. (2012). Statistical evaluation of toxicological experimental design for Bayesian model averaged benchmark dose estimation with dichotomous data. *Human and Ecological Risk Assessment: An International Journal*, *18*(5), 1096-1119.

Symonds, M. R., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*(1), 13-21.

Volinsky, C. T., Madigan, D., Raftery, A. E., & Kronmal, R. A. (1997). Bayesian model averaging in proportional hazard models: assessing the risk of a stroke. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *46*(4), 433-448.

Vrugt, J. A., Diks, C. G., & Clark, M. P. (2008). Ensemble Bayesian model averaging using Markov chain Monte Carlo sampling. *Environmental fluid mechanics*, *8*(5-6), 579-595.

Vrugt, J. A., & Robinson, B. A. (2007). Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resources Research*, *43*(1).

Wan, A. T., Zhang, X., & Zou, G. (2010). Least squares model averaging by Mallows criterion. *Journal of Econometrics*, *156*(2), 277-283.

Wang, D., Zhang, W., & Bakhai, A. (2004). Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. *Statistics in medicine*, *23*(22), 3451-3467.

Wang, H., Zhang, X., & Zou, G. (2009). Frequentist model averaging estimation: a review. *Journal of Systems Science and Complexity*, *22*(4), 732-748.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of mathematical psychology*, *44*(1), 92-107.

West, R. W., Piegorsch, W. W., Peña, E. A., An, L., Wu, W., Wickens, A. A., ... & Chen, W. (2012). The impact of model uncertainty on benchmark dose estimation. *Environmetrics*, *23*(8), 706-716.

Wheeler, M. W., & Bailer, A. J. (2007). Properties of Model-Averaged BMDLs: A Study of Model Averaging in Dichotomous Response Risk Estimation. *Risk Analysis* **27**, 659–670.

Wheeler, M. W., & Bailer, A. J. (2009). Comparing Model Averaging With Other Model Selection Strategies for Benchmark Dose Estimation. *Environ. Ecol. Stat.* **16**, 37-51.

Wheeler, M., & Bailer, A. J. (2012). Monotonic Bayesian semiparametric benchmark dose analysis. *Risk Analysis*, *32*(7), 1207-1218.

Wheeler, M. W., & Bailer, A. J. (2013). An empirical comparison of low-dose extrapolation from points of departure (PoD) compared to extrapolations based upon methods that account for model uncertainty. *Regulatory Toxicology and Pharmacology*, *67*(1), 75-82.

White, R. H., Cote, I., Zeise, L., Fox, M., Dominici, F., Burke, T. A., ... & Samet, J. M. (2009). State-of-the-science workshop report: issues and approaches in low-dose–response extrapolation for environmental health risk assessment. *Environ Health Perspect*, *117*(2), 283-7.

Wintle, B. A., McCarthy, M. A., Volinsky, C. T., & Kavanagh, R. P. (2003). The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, *17*(6), 1579-1590.

Zhang, X., & Liang, H. (2011). Focused information criterion and model averaging for generalized additive partial linear models. *The Annals of Statistics*, *39*(1), 174-200.

APPENDIX B. FORMAT OF AN AVG FILE FOR INPUT INTO CMODELAVG.EXE

```
        > Line 1: Model Names
Linear Polynomial3 Power Hill Exponential3 Exponential5     [list models available]
        > Line 2: Model Weights: 0 ≤ wt(i) ≤ 1; Σwt(i) = 1
0.5 0.5 0 0 0 0                                             [in this example, linear and Poly3
                                                            models would be run with equal prior
                                                            wt]
        > Line 3: Restricted: 0=No, 1=Yes
0 1 1 1 1 1                                                 [same number of flags as in Line 2, in
                                                            same order]
        > Line 4: Options: a b c d e f g h
                > 4a: Number bootstrap iterations
                > 4b: Confidence limit
                > 4c: BMR Type:       0=Absolute Deviation, 1=Standard Deviation
                                      2=Relative Deviation, 3=Point Estimate (currently only
                                      relative deviation, "2," is available)
                > 4d: BMRF
                > 4e: Distribution: 0=Normal, 1=Lognormal   (currently only "0," Normal, is
        available)
                > 4f: Constant Variance: 0=No, 1=Yes
                > 4g: Random Seed: 16 hex bytes, or "0" for automatic selection of seed)
                > 4h: Adverse Direction: -1=Down, 1=Up
1000 0.95 2 0.1 0 1 0 1                                     [in this example, a 95% lower bound on
                                                            a relative risk of 0.01 is computed
                                                            using 1000 bootstrap iterations;
                                                            models assume constant variance;
                                                            random seed is picked automatically,
                                                            and the adverse direction is "up"]
        > Line 5: File name, path, or other identifier(s)
D:\Projects\ModelAvg\ModelAvg\Data\Continuous4.dax
        > Line 6: Data Column Headers
Dose N Mean Std
        > Line 7+: Data (in order given by Headers; 1 line per dose group)
0       10      1.61    0.12
35      10      1.62    0.13
105     10      1.71    0.11
316     10      1.91    0.15
625     10      2.5     0.13


        .
```

APPENDIX C. EXAMPLE CSV FILE (REAL_DATA_UP.CSV) FOR RUNNING WITH MATEST2.SH

real data up																									
unknown BMD																									
1																									
		dose	N	mean	std	dose	N	mean	std	dose	N	mean	std	dose	N	mean	std	dose	N	mean	std	dose	N	mean	std
20	unknown	0	10	1.037	0.015	1	10	1.05	0.01	3	10	1.052	0.01	9	10	1.066	0.01								
21	unknown	0	20	0.92	0.09	10	20	1.05	0.2	33	20	1.01	0.09	100	20	1.07	0.2								
22	unknown	0	20	0.9	0.2	10	20	0.99	0.09	33	20	1.08	0.2	100	20	1.08	0.1								
27	unknown	0	10	87	9	0.1	10	88	28	1	10	92	6	10	10	87	6								
30	unknown	0	6	3.5	1.7	0.1	6	4.5	3.8	1	5	3.7	3.7	10	5	12.9	2								
74	unknown	0	14	11	1	50	14	11.8	1	250	14	11.7	0.9	1000	14	12.4	1								
75	unknown	0	10	1.74	0.37	50	10	2.14	0.57	250	10	2.38	0.65	1000	10	2.7	0.78								
76	unknown	0	14	6.43	0.39	50	14	6.49	0.48	250	14	6.74	0.37	1000	14	6.86	0.58								
77	unknown	0	10	1.27	0.14	50	10	1.4	0.25	250	10	1.45	0.21	1000	10	1.6	0.31								
78	unknown	0	14	1.5	0.11	50	14	1.57	0.14	250	14	1.62	0.1	1000	14	1.66	0.17								
82	unknown	0	10	1.2	0.4	5	10	3	1	16	10	4.4	1.3	50	10	10.1	1.2								
83	unknown	0	10	1.6	0.8	5	10	3.2	0.9	16	10	3.9	1.3	50	10	10.5	1.5								
84	unknown	0	10	0.6	0.2	5	10	0.9	0.6	16	10	3.2	0.7	50	10	10.1	2								
85	unknown	0	10	2.1	1.2	5	10	2.3	0.6	16	10	3.7	0.2	50	10	9.6	2.5								
86	unknown	0	10	0.7	0.6	5	10	1.6	0.4	16	10	2.1	1.3	50	10	5.8	1.7								
87	unknown	0	10	1.3	0.9	5	10	0.8	0.5	16	10	2	0.6	50	10	5.1	0.8								
92	unknown	0	10	11.615	0.363	125	10	11.99	0.268	250	10	11.747	0.253	500	10	12.332	0.292	1000	10	13.289	0.52	1500	10	12.941	0.417
93	unknown	0	10	6.079	0.172	125	10	6.134	0.225	250	10	6	0.158	500	10	6	0.269	1000	10	6.841	0.162	1500	10	7.402	0.208
95	unknown	0	10	45.08	2.8	62.5	10	44.2	2.7	125	10	44.24	2.1	250	10	49.28	4.08	500	10	51.43	3.16				
97	unknown	0	10	48.49	3.54	62.5	10	46.65	2.12	125	10	48.57	2.12	250	10	55.03	2.8	500	10	59.61	3.26				

Note: File name MUST be the same as the name in cell A1 (though the name in A1 need not have the underscores).