The Application of Linear Poisson Models to Changes in ADC Measurements.

Neil Thacker and Paul Tar

Last updated
23 / 09 / 2015
The Application of Linear Poisson Models to Changes in ADC Measurements

N.A. Thacker, P.D. Tar

Abstract

Assessment of tumour response to therapy in pre-clinical animal models is complicated by the presence of “normal” tumour growth. Simple methods for assessment of a response, based upon T tests of median ADC difference, do not take account of these changes so that statistical power is compromised. Alternative approaches, such as the use of measures of heterogeneity, are also likely to have poor statistical power if the data are highly variable. Such problems are exacerbated by the use of rapidly growing tumour models, which are often preferred in order to minimise the duration of the study. When such problems occur, experimental studies must use larger sample sizes (typically tens) in order to reach significance.

We present here what we believe to be an appropriate method for approximating heterogeneous changes of ADC in tumour tissue following treatment, using a statistical approach which supports estimation of volumes and associated errors. Statistical significance for measured change is found to be possible even for individual samples. The approach (LPM, Linear Poisson Models) is based on the construction of a best generalising model using ICA (Independent Component Analysis), with extensions to provide assessments of uncertainty. These methods have been developed and tested by our group over the last five years.

This document is a technical summary of the analysis of pre-clinical ADC data provided by James O’Connor, and should later form the basis of a joint publication. The work has relevance to several projects currently funded at Manchester by Leverhulme, CRUK and the European Commission (IMI).

1 Linear Poisson Models and ADC Measurements

This document outlines a method for applying Linear Poisson Models (LPM) [Tina Memo 2013-006] to tumour ADC distributions in order to model, and then measure, changes in response to cancer treatments. The aim is to quantify, within predictable error bounds, the changing volume of tissue accountable by normal tumour growth and those which can only be interpreted as changes due to treatment. Unfortunately, normal growth and volume changes in the tumour tissue, and the nature of heterogeneous response to therapy, prohibits the use of non-rigid registration to assess any meaningful change. We show here how these problems can be overcome by working directly with histogram distributions of ADC data taken from regions of interest over whole tumours.

Our method begins by constructing a model of untreated tissue variability from histogram control data, then extending the model to account for additional variability observed in treated examples. A successful analysis provides estimates of the quantity of effected tissues, before and after treatment, accompanied by error estimates and a confirmatory goodness-of-fit. We show that with this method, once a model of normal grown is known, the measured errors on these quantities are sufficient to quantify change in individual samples.

The LPM method is an extension of conventional pattern recognition approaches making it possible to build approximate models of sample distributions, from suitable examples, for purposes of quantitation. Unusually for pattern recognition, the analysis has a fully developed error theory, making it possible to make quantitative measurements even where unique (i.e. unambiguous) labels for data classes would not normally be available. This can be considered as the process of extracting areas from significantly overlapping curves in a fit, where although specific parts of the data cannot be attributed to either curve, LPMs can non-the-less fit (via Likelihood estimation) the quantities of each curve which best accounts for the total distribution.

In the context of making ACD measurements, the regression fits a combination of a control distribution (estimated from an untreated cohort) and a treatment-effected distribution (extracted from a treated cohort). Linear components of behaviour are found which predict most accurately the characteristic (heterogeneous) responses of tissues. The total area under the fit associated with each distribution is proportional to the quantity of tissue involved in each class. Estimates of these quantities, and their associated errors, are the primary output of LPMs.

A chi-square per degree of freedom test is applied as a goodness-of-fit check, which can be used to a) ensure that control and treated distributions are being described to comparable levels of fidelity, and b) to scale quantity error covariances to best approximate the effective granularity of the (assumed) underlying Poisson data. We consider these tests to be essential to checking appropriate use of these (or any other) model based analyses.
2 Methodology

The following data are assumed to be available:

- \(N\) examples of ADC distributions for untreated control samples, at time \(t_0\) and at \(t_1\);
- \(M\) examples of ADC distributions for treated samples, at time \(t_0\) and \(t_1\);

The above data are expected to have properties appropriate for the application of Linear Poisson Models:

- The data can be binned to take the form of histograms, where each bin is a Poisson count of voxels associated with a range of ADC values;
- The data are expected to be complex, i.e. not easily describable using simple parametric models such as Gaussians;
- The control data are expected to vary naturally between subjects and over time, as tumours will be of different sizes and compositions;
- The treated data are expected to overlap the control data, but with additional modes of variation indicative of affected tissue.

It is assumed that the binned ADC values are approximately Poisson, which we check by creating Bland-Altman plots of residual distributions. We have used Monte-Carlo simulations to confirm that any double counting or correlation effects (perhaps caused by interpolation and/or smoothing of voxels during data acquisition) can be accounted for with a simple scaling of errors. It is further assumed that ADC histograms can be approximated (to some level of accuracy) as a linear combination of a small number (< \(N\) and < \(M\)) of fixed sub-distributions, as used in LPMs:

\[
H_X \approx M_X = \sum_{k=0}^{n+m-1} P(X|k)Q_k
\]

where \(H_X\) is a histogram bin within the ADC distribution over range \(X\); \(P(X|k)\) is the probability of observing the ADC range of values, \(X\), given linear component \(k\); and \(Q_k\) is the quantity of component \(k\) within the data.

The analysis problem can be summarised as the need to estimate a set of probability distributions, \(P(X|k = 0), P(X|k = 1), \ldots, P(X|k = n - 1)\), which can be linearly combined to described a set of untreated control examples, then learning additional probability distributions, \(P(X|k = n), P(X|k = n + 1), \ldots, P(X|k = n + m - 1)\) which can describe the distribution of affected tissue. The quantities,

\[
Q_{\text{Control}} = Q_0 + Q_1 + \ldots + Q_{n-1} \pm \sigma_{\text{Control}}
\]

and

\[
Q_{\text{Treated}} = Q_n + Q_{n+1} + \ldots + Q_{n+m-1} \pm \sigma_{\text{Treated}}
\]

then need to be estimated from a treated cohort to quantify tissue volume changes and their associated statistical significance.

2.1 Assessing Model Fits and Residual Behaviour

For LPM’s, the Likelihood functions which underpin model selection, parameter estimation and quantitative estimation of errors, is based upon a Poisson model of Histogram formation. The residuals between fitted models and data, \(M_X - H_X\), can reveal valuable information regarding the validity of assumptions made by LPMs. If binned ADC values are indeed Poisson in nature then the residuals should grow proportionally to the square-root of the bin quantity. If the granularity of the individual Poisson events being counted are larger than unity, this too should be revealed as a scaling factor on the square-root dependency. Bland-Altman plots can be constructed and a power-law error model fitted to assess both of these properties [Tina Memo 2015-005 and 2015-006]:

3
\[
\sigma_{M_X - H_X} = a \left( \frac{H_X}{a} \right)^{b/2}
\]
where \(a\) and \(b\) should be unity for non-scaled Poisson bins, and \(a\) should be a variance scaling in cases where only \(b\) is unity. A value of \(b\) deviating too far from unity would suggest that LPMs are inappropriate for the data. This particular parameterisation (as opposed to \(a(H_X)^b\)) is used in order to obtain more efficient estimates of parameter covariances from the fitting process.

A \(\chi^2\) can be computed from the sum of squared residuals, each normalised to the predicted bin errors (assuming independence and non-scaled Poisson counts), to assess the overall quality of a LPM fit to data. As \(\chi^2\) values are computed from Gaussian distributions, the Anscome transform can be applied to approximately convert the Poisson variables into Gaussian ones of fixed variance of one quarter. This approximation then supports the use of a least-squares similarity function, even for small numbers of histogram entries:

\[
\chi^2 = \sum \chi_i^2 = \sum_i \sum_X \left( \frac{\sqrt{H_{Xi} + \frac{3}{8}} - \sqrt{M_{Xi} + \frac{3}{8}}}{\sigma_{Xi}^2} \right)^2
\]

where \(i\) is a sum over histograms; \(X\) a sum over histogram bins; and \(\sigma_{Xi}^2\) is the predicted variance on the transformed residual. The value of \(\sigma_{Xi}^2\) will change depending upon the source of the underlying model, \(M_X\), which will be discussed shortly. However, there will always be a \(\frac{1}{4}\) contribution from the Anscome transformed \(H_X\).

Of more value is the \(\chi^2\) per degree of freedom, which should be unity in the case of non-scaled Poisson data. Values above unity can be used to scale error covariances and should closely correspond to the value of \(a\) from the above Bland-Altman power-law model fitting (assuming \(b = 1.0\)). The number of degrees of freedom, \(D\), is defined as the number of data points fitted minus the number of parameters computed from those data points. The degree of freedom calculation (along with the calculation of \(\sigma_{Xi}^2\)) depends upon the source of the underlying model components, which will be discussed in the relevant sections that follow.

The total \(\chi^2\) for a cohort of histograms is of use, but \(\chi^2_D\) for each individual histogram is perhaps more useful to provide a subject-by-subject assessment of model fits (and therefore potential error scaling). This gives a set of useful values:

\[
\chi^2_D = \frac{1}{D} \chi^2
\]

and

\[
\chi^2_{Di} = \frac{1}{D_i} \chi^2_i
\]

where \(D = \sum_i D_i\).

### 2.2 Building a Control Model

A control LPM describes the ADC distributions of \(N\) normal, untreated tumours at two time points a fixed period apart. Building an approximating control model involves populating \(N\) histograms, each containing \(c\) number of bins (the issue of selecting an appropriate number of bins is discussed in section 4). The histograms must contain 2 dimensions, i.e. ADC verses time, with 2 respective time bins for measurements taken at \(t_0\) and \(t_1\). Populating joint histograms containing ADC and time information allows temporal correlations to be learnt as well as correlations across different ACD ranges.

A LPM estimates a set of \(n\) PMFs, \((P(X|k))\), which are capable of approximating the \(N\) control histograms with the best generalising accuracy, using histogram-specific weighting quantities. As there are \(N\) histograms with \(c\) bins, and \(n\) linear components, there are \(Nc\) weighting quantities and \(cn\) PMF bins which will be estimated. The number of degrees-of-freedom for a \(\chi^2_D\) computed from the control model is therefore:

\[
D = cN - n(c + N)
\]

---

1This parametric form is chosen for its statistical characteristics, i.e. the noise distributions for \(a\) and \(b\) are found to be Gaussian.
where $cN$ gives the total number of data points (bins in all histograms); $nc$ is the total number of PMF bins estimated; and $nN$ is the number of quantities estimated, one quantity per component, per histogram. As the $\chi^2$ for the control model is fitted to the control data (as opposed to the treated data, which is an independent dataset) then the $\sigma^2_{X_i}$ term is simply $\frac{1}{4}$.

The number of model components required to best describe the data, whilst not over-fitting, is determined using a leave-one-out generalisation procedure. $N - 1$ examples can be used to train a LPM, which should then be applied to the remaining example to confirm a representative model can be learned. This can be repeated for all leave-one-out combinations in order to find the best generalising model order (i.e. one that gives the smallest total residual variance). A final LPM should be trained using all $N$ examples, with the final model being applied to treated data for making ADC change measurements. The final model will contain $n$ linear components, where $n \ll N$. In the case where the LPM solutions result in $n$ being close to $N$ (i.e. each control example is statistically unique), we would conclude there is too much variability in the data to construct a useful common set of base components.

Several methods can be used to eliminate problematic data, as often arises with poorly controlled data acquisition, particularly involving biological samples. If $\chi^2_D$ values of unity are not achieved, then further steps may be required. If there are problems with a specific leave-one-out combination then an example may be excluded as an outlier in order to continue with the construction of a model. If there are consistently large $\chi^2_D$ values across all examples then the predicted Poisson errors assumed on histogram bins may require scaling. This can be achieved simply by multiplying errors by:

$$\sigma^\prime_{Control} = \sigma_{Control} \chi_D$$

3 Building a Treatment Model

$M$ examples of treated tumours must be provided, with $t_0$ giving ACD distributions before treatment and $t_1$ giving distributions a fixed period afterwards (same period as found between $t_0$ and $t_1$ in the control group). These should form $M$ 2-dimensional histograms with binning common with the control histograms.

The $n$ components from the control LPM is first fitted to the $M$ treatment examples, which are only expected to account for the untreated parts of the data. Choosing $n$ with the best approximation is achieved using the generalisation test specified earlier. The additional variability in data caused by treatment can be learnt by adding $m$ additional linear components, whilst keeping the $n$ control components in the model unchanged. If successful, $m + n \ll M$ whilst providing a goodness-of-fit comparable to that achieved in the control group i.e. we assume that the combined model should approximate control and treatment data equally well.

As there are now $n$ PMFs which have been estimated previously from control data and an additional $m$ PMFs estimated from the treated data, the number of degrees-of-freedom becomes:

$$D = cM - cm - M(n + m)$$

where $cM$ is the number of data points (all bins); $cm$ is the number of estimated PMF bins describing treatment variability; and $M(n + m)$ are the weighting quantities estimated. As the original $n$ control components were estimated from an independent training data set, the errors from that training data propagate to the estimates of $\sigma^2_{X_i}$:

$$\sigma^2_{X_i} = \frac{1}{4} + \sum_{k=0}^{n-1} \frac{Q^2_{ki}}{4M_{X_i}} \sigma^2_P(X|k)$$

The final $\chi^2_D$ values can be used to scale estimated volume errors on a case-by-case basis.

4 Selecting Appropriate Histogram Bins

Previous work has shown that the resolution of ADC data is not expected to be much better than ±6%. A binning resolution approximately matching the resolution of the data is sought. If too low, the histogram binning will loose information regarding the shape of the ADC distribution. If too high, bins will be too poorly populated for the
statistical approximations used. Even when using Anscomb transforms, we need to avoid a large number of single figure histogram entries.

A satisfactory number of bins \( c \) can be determined by running the analysis end-to-end using different binnings. A range of binning resolutions is sought over which statistically consistent volume and error estimates can be achieved. Any result from this range will give statistically valid summaries of the data.

5 Results

Two tumour datasets (HCT and Lovo) were used to test the proposed method. These contained 13 control + 15 treated examples in the HCT set, and 8 control + 10 treated in the Lovo set. A leave-one-out model selection process, minimising the total residual variance as a function of the number of model components, shows that the HCT data is the more complex of the two sets. HCT data required \( n = 4 \) control components and \( m = 5 \) additional treatment components. Lovo data required only \( n = 2 \) control components and a further \( m = 2 \) for treated. The model selection curves can be seen in figures 1(a), 1(b), 2(a) and 2(b).

The residuals in both sets were similar, revealing approximately Poisson behaviour with a variance scaling factor of around 10. This can be seen from the Bland-Altman error model fit parameters in figure 3 and also the \( \chi^2_D \) values in figure 4. Note that the Bland-Altman error parameter \( b \) is close to unity, indicating a Poisson power-law error growth (as required by LPM assumptions), and error parameter \( a \) and \( \chi^2_D \) are equivalent within errors.

The selected binning \( c = 200 \) was chosen from a range of options including 50, 100, 150, 200 and 250. Figure 5 shows a scatter plot of estimated volumes (\( Q_{Control} \) as computed from the treated HCT data) as a function of binning. The x-axis gives volumes computed using the preferred 200 bins, whereas the y-axis gives the alternative binning volumes. It can be seen that binnings of 50 and 100 are too low, but binnings 150, 200 and 250 are self-consistent in terms of estimated volumes and size of errors. Figure 6 compares the volume error estimates, again showing that 50 and 100 binnings are in poorer agreement than the others.

Figures 7 to 10(b) show the composition of each tumour. Figures 7 and 9 show the composition of HCT and Lovo control distributions. The other figures show the composition of HCT and Lovo treated distributions, containing as a subset, the original control components. Figures 8(b) and 10(b) summary the quantity of tissue accountable by normal growth versus treatment changes. The errors were scaled by the case-by-case \( \chi^2_{Di} \).

Figures 11(a) and 11(b) show the probability distribution of control ADC values in the HCT set at times \( t = 0 \) and \( t = 1 \), respectively. Figures 12(a) and 12(b) show the corresponding treatment HCT components. The general trends shows that ADC values increase with treatment, giving a long tail to the distributions. The effects of treatment within the Lovo set are similar, but much simplified.

6 Discussion

Our model can be described as an ICA of histogram correlation between two measured time-points. The ICA components extracted during this analysis can be interpreted as characteristic changes in homogenous regions within otherwise heterogenous tumour tissue. This is a mathematical embodiment of the description normally used by biologists to describe the system. For example, component 2 (yellow) in figures 11 (a) and (b) can be interpreted as tissue which is becoming more necrotic, and thereby changing its state to larger values of ADC. The extracted distributions and variation in their contributions to individual samples provide a wealth of information regarding heterogeneity. The statistical model is directly applicable for other uses, such as Monte-Carlo modelling of tumour growth and treatment.

The construction of a pattern recognition system, which describes intrinsic variability of data distributions, would normally require significant amounts of data. One possible barrier to the successful application of LPMs to ADC distributions is simply that insufficient examples may be available to describe the variabilities seen. As very limited data is typically available in related studies, the adoption of a ‘leave-one-out’ approach to the evaluation of each sample is recommended. This can be applied both to the treatment and non-treatment datasets in order to assess the effectiveness of statistical separability.

The amount of data required to build models will typically be greater than that required to construct a T-test, however once constructed the statistical power will be much greater for data sets with complex behaviours (where a T-test will treat all sources of variation as stochastic noise). Whereas conventional analyses assume that the result of a successful tumour treatment will be an increase in median ADC, the proposed analysis makes no assumptions regarding the expected change, or indeed volume change in the tumours. Both of these factors are incorporated
automatically into the analysis, estimated from sample data, along with the associated effects of adding parameters 
to the predictive model. By doing so the approach is sensitive to (extracts information relevant to) both changes 
in volume as well as ADC.

Whilst the most obvious interpretation of scale factor $a$ is the inability to describe the sets of histograms using the 
assumed linear model, another interpretation is an “effective degree of freedom” scaling, i.e. there are 10 times 
fewer independent measurements samples than would be expected by counting the voxel entries. Such an effect 
might occur due to correlations introduced during MR image acquisition and formation, generally as a result of 
ttempts to speed up acquisition rates while maintaining subjective image quality. Sampling on an $n^3$ k-space 
and interpolating to a $(2n)^3$ image would generate a scale factor of 8. Sampling using sparse k-space acquisition 
and image smoothing will have similar effects. Generally any attempts to “improve” the image subsequent to the 
original measurement may convince an observer but will not improve an objective analysis.

It should be remembered that because this is a linear system the model is likely to be degenerate (i.e. not a unique 
description). However, the overall linear model of variation is expected to be stable, so that estimates of treated 
and untreated volume should be consistent with the estimated errors. Once processed, the origins of the original 
data as ADC distributions can be largely disregarded, a more informative summary is available in the form of 
estimated quantities of responding tissue volumes. Estimation errors are available for subsequent analysis, such 
as the construction of p-values. It is these estimates which are not available in off-the-shelf pattern recognition 
techniques.

As we have mentioned earlier, the aim of this analysis is to build models of variation and to assess heterogenous 
changes in tissue, without the need to perform a pixel by pixel registration. Arguably a spatial map's only purpose 
is to support determination of a quantity estimate. This can be done by working directly with histograms and 
estimating the most predictive model of variation. Although it is logically possible to compute probabilities of 
association of the incoming tumour data with responding and non-responding parts of the model, for purposes of 
constructing an image, in circumstances of high ambiguity such an image would potentially be highly misleading. 
However, whilst a spatial map cannot be directly constructed, the extracted models might be of use to provide 
constraints to algorithms for non-rigid alignment as an alternative to more arbitrary models of image smoothness.

For drug trials, it is important that statistical methods can be “powered” in advance. This requirement poses a 
significant challenge when the results are dependant upon the complexity of biological response (which is unknown). 
This will be true for any analysis, including T-tests of mean ADC difference. The two tumor models presented 
here show the differences in complexity which might arise. However, the normal group variation can be assessed 
in advance of the treatment group, allowing some form of prediction. The control model could be used in isolation 
simply to identify outliers following treatment, in the form of a hypothesis test. For the data in this study this 
would provide almost unambiguous detection of change in all samples.

Whilst the statistical power of the LPM seems impressive, the ability to unambiguously detect response to treatment 
in individual samples raises another issue. In a highly variable system, how many samples are required to truly 
understand response? It would be incorrect, for example, to observe a 6 S.D. effect in one sample and then conclude 
that the drug always works. Provided the highly unambiguous assessment of individuals seen in this work can be 
relied upon, the appropriate statistical model for powering a study reverts to binomial statistics ($n$ responders 
from $N$ samples).

It is also important that statistical methods can deal with all data provided, (i.e. there are no special cases which 
require exclusion). Normally poorly behaved data might be interpreted as a measurement failure rather than a 
genuine variation, particularly in biological studies with methodological challenges. In this work all samples can 
indeed be included during model building, with the possible consequence of inflating estimated errors. Leave-one 
out tests will however indicate problems if there are too few samples to adequately model all possible forms of 
variation. When assessing a treatment response, if this too is effectively an outlier, then the goodness of fit test 
will show this and the estimated errors will be scaled accordingly. In summary, highly variable data is correctly 
summarised as useless in measurement terms. A conventional PR approach would be unable to do this. It should 
be remembered that the only question we can really expect to answer is; can we detect a change in tumour 
development? Bad quality data will of course prevent us from doing this.

Building prior models of both control and treatment could also be considered as the basis for personalised therapy. 
Although we would not expect a mouse model to be directly applicable to human data, so requiring a clinical study 
or the accumulation of data from clinical work. Not only can the model be used to assess the nature of response 
in comparison to that expected, fitting the ADC distribution using only data from the first time point could be 
used to predict the expected response, i.e. do we expect this tumour to respond appropriately. The limits on the 
success of this approach would depend upon two issues, the conformity of the biology to the assumed model and 
the quality of ADC data. For use in multiple centres this would require careful standardisation of well designed 
MR protocols and data analysis, along with appropriate use of QA (Quality Assurance) and QC (Quality Control).
7 Conclusions

We have shown that for MR based ADC data, the complex issue of heterogeneous tumour response can be addressed by constructing models of variation. New tools for quantitative modelling have been developed and applied, which support tasks such as estimation of volumetric response to drug therapy. The constructed models have immediate uses as a description of variation for understanding the role of heterogeneity in tumour development. They can also be used to quantitatively assess change, with a level of statistical power which is much greater (by around a factor of 5 on S.D. of measured difference) than the standard methods. This is sufficient to be able to quantify changes in individual subjects, opening the possibility of personalised drug therapy.

For pre-clinical work, the methods are more complicated than those currently used and more work will be needed to investigate issues such as; powering a study, how best to use the volumetric estimates and the consequences of unstable data. Meanwhile, these models would have great value in assessing the performance of simpler statistical analyses using Monte-Carlo (see for example Tina Memo 2012-007).

Figure 1: (a) The optimum number of components required to give the most generalisable model for HCT treated data, as assessed by a leave-one-out selection process. The y-axis shows the variance on residuals averaged over all leave-one-out combinations. (b) The optimum number of components required to give the most generalisable model for HCT control data, as assessed by a leave-one-out selection process. The y-axis shows the variance on residuals averaged over all leave-one-out combinations.

Figure 2: (a) The optimum number of components required to give the most generalisable model for Lovo control data, as assessed by a leave-one-out selection process. The y-axis shows the variance on residuals averaged over all leave-one-out combinations. (b) The optimum number of components required to give the most generalisable model for Lovo treated data, as assessed by a leave-one-out selection process. The y-axis shows the variance on residuals averaged over all leave-one-out combinations.
Figure 3: The power-law error model parameters estimated from Bland-Altman plots of residuals for the 4 Linear Poisson Models. Parameter $b$ (power law $\frac{0.5}{b}$), being close to unity, shows that in all models the data behaves Poisson-like, giving an approximate square-root growth on bin errors. Parameter $a$ (scaling on variance), is consistently around 10 for all models showing that each is being modelled to a similar level of accuracy.

Figure 4: The $\chi^2_D$ values for each model, consistent with a variance scaling of around 10. The red diamond and yellow triangle highlight 2 possible outliers in the HCT treated data. Note that these values corroborate the Bland-Altman parameter $a$ values from figure 3.

Figure 5: The self-consistency of estimated volumes and errors across a range of possible ADC binning resolutions. The x-axis shows estimated volumes, $Q_{\text{Control}}$, for histograms with 200 bins (the preferred binning). The y-axis shows corresponding volume estimates for alternative binning resolutions. Histograms using 50 and 100 bins fall outside the stable range, whereas binning of 150 to 250 are tightly clustered and statistically equivalent. Error bars on the 50 and 100 bin points have been omitted for clarity.
Figure 6: The self-consistency of estimated volume errors across a range of possible ADC binning resolutions. The x-axis shows volume errors, $\sigma_{Control}$, for histograms with 200 bins (the preferred binning). The y-axis shows corresponding error estimates for alternative binning. Histograms with 50 to 100 bins are generally larger and more variable than higher resolutions up to 250.

Figure 7: The composition of the HCT control data, i.e. quantities $Q_0, \ldots, Q_{n-1}$, estimated for each of the $N$ control subjects.

Figure 8: (a) The composition of the HCT treated data, i.e. quantities $Q_0, \ldots, Q_{n+m-1}$, estimated for each of the $M$ treated subjects. (b) The composition summary of the HCT treated data, i.e. quantities $Q_{Control}, Q_{Treated}$, estimated for each of the $M$ treated subjects.
Figure 9: The composition of the Lovo control data, i.e. quantities $Q_0, \ldots, Q_{n-1}$, estimated for each of the $N$ control subjects.

Figure 10: (a) The composition of the Lovo treated data, i.e. quantities $Q_0, \ldots, Q_{n+m-1}$, estimated for each of the $M$ treated subjects. (b) The composition summary of the Lovo treated data, i.e. quantities $Q_{\text{Control}}, Q_{\text{Treated}}$, estimated for each of the $M$ treated subjects.

Figure 11: (a) Probability distributions for LPM components in HCT control data at time $t = 0$. (b) Probability distributions for LPM components in HCT control data at time $t = 1$. 
Figure 12: (a) Probability distributions for LPM components in HCT treated data at time $t = 0$. (b) Probability distributions for LPM components in HCT treated data at time $t = 1$.

Figure 13: (a) Probability distributions for LPM components in Lovo control data at time $t = 0$. (b) Probability distributions for LPM components in Lovo control data at time $t = 1$.

Figure 14: (a) Probability distributions for LPM components in Lovo treated data at time $t = 0$. (b) Probability distributions for LPM components in Lovo treated data at time $t = 1$. 