Published in final edited form as:

Mult Scler. 2010 August ; 16(8): 926–934. doi:10.1177/1352458510373110.

Modeling lesion counts in multiple sclerosis when patients have been selected for baseline activity

CJ Morgan, IB Aban, CR Katholi, and GR Cutter

Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama, USA.

Abstract

The number of new gadolinium-enhancing lesions discovered via magnetic resonance imaging is a well-established outcome for multiple sclerosis studies, especially Phase II Studies. Due to the high cost of magnetic resonance imaging scans, many investigators select participants for the presence of lesions. While this selection procedure is thought to improve the power of inferences, the effect of screening for baseline activity on parameter estimation and interval coverage has not yet been examined. The objective of this study was to investigate the performance of the negative binomial distribution for modeling lesion count data in multiple sclerosis when patients have been selected for activity on a baseline scan. We performed computer simulations to investigate the influence of the screening process on inferences made using a negative binomial model about treatment effects in two independent samples. We also demonstrate how the statistical properties of screening can be incorporated into trial design. We demonstrate that when the negative binomial distribution is used to model lesion counts, while screening for baseline activity improves point estimation, this practice also has the potential to decrease interval coverage and inflate the Type I error rate. For data that is to be modeled using a negative binomial distribution, screening for baseline activity can create a trade-off between cost effectiveness and a higher than desired false positive rate that must be carefully considered in planning Phase II trials.

Keywords

Multiple sclerosis; MRI; lesions; negative binomial; enrichment design

Introduction

The number of new gadolinium-enhancing lesions discovered via magnetic resonance imaging (MRI) is a well-established outcome variable for multiple sclerosis (MS) studies,^{1,2} especially Phase II studies. Unfortunately, MRI is a relatively expensive technique, making the task of accurately calculating power and estimating sample size particularly crucial. Sormani et al. note that the utilization of an appropriate parametric distribution for the number of lesions would improve sample size and power calculations.³ While several count distributions have been proposed to model this type of data, the negative binomial distribution has been consistently found to provide one of the better fits to the data.^{3–5} The negative binomial distribution has also begun to be applied as a common approach to lesion count data^{6–8} and to be used frequently in sample size calculations.⁹

For many clinical trials focusing on relapsing MS, investigators pre-screen subjects for baseline MRI activity (i.e. the presence of one or more lesions; see for example Comi et al.¹⁰), the intuition being that subjects who show no activity on a baseline scan are less likely to show activity in future scans, wasting MRIs and possibly biasing the study towards the null hypothesis. This practice is an example of *enrichment*, which Temple¹¹ describes as the preferential inclusion of patients in a trial who are either more likely to respond to treatment or in whom treatment effects can be more easily detected. This type of enrichment design aims to increase power by conducting analyses on a patient subpopulation with less variability than the general population. Sormani et al.⁹ investigated the result of screening for baseline activity on sample size estimation and found that screening did in fact increase power. However, the effect of this screening procedure on the accuracy of estimates of treatment effects is unclear. In this paper, we investigate the performance of the negative binomial distribution when patients have been selected for activity on a baseline scan. We also illustrate how knowledge of a screening procedure's properties can be useful in trial design.

Methods

We implemented a procedure similar to that of Sormani et al.³ and performed a series of simulations to investigate the effect of screening for baseline activity. We examined the simple setting where subjects are given a single screening scan at baseline, randomized to treatment groups, and then assessed at a later date with a single follow-up scan. We first simulated a baseline scan by drawing the number of lesions for each subject from a negative binomial distribution, denoted by $NB(\mu, \theta)$, with mean μ and variance $\mu + \mu^2/\theta$. To accomplish this, we first simulated a latent mean, λ_i for each subject by drawing from a gamma distribution with mean μ and variance μ^2/θ . We then drew a lesion count for the baseline scan for the i th subject from a Poisson distribution with mean λ_i . We kept sampling until there were 50 subjects having at least one lesion at baseline, and randomly assigned 25 of these to the treatment group and the remaining 25 to the control group.

We then simulated the experimental results. For treated subjects, the number of lesions was drawn from a Poisson distribution with mean $\gamma\lambda_i$. For control subjects, lesion count was drawn from a Poisson distribution with mean λ_i . That is, the mean lesion count for treated subjects is reduced by a factor of γ or, equivalently, treated subjects show a $(1 - \gamma) \times 100\%$ reduction average lesion count compared with controls. We refer to the reduction in average lesion count, $1 - \gamma$, as the *treatment effect*.

Treatment effects were estimated by fitting a negative binomial log-linear model to the experimental results, where the lesion counts for control subjects were assumed to follow a $NB(\mu, \theta)$ distribution and for treated subjects were assumed to follow a $NB(\gamma\mu, \theta)$ distribution. That is, we assumed that the treatment affects only the location parameter, μ , of the negative binomial distribution and that treated and control subjects share common θ . The results of the baseline scan were only used to screen subjects and were not used in estimating treatment effects.

We also repeated the above procedure without selecting subjects based on the results of the baseline scan. In order to investigate how the effects of screening may vary according to patient population, we performed this simulation using several possible parameter values: $\mu = 1, 2, \dots, 12$; $\theta = 0.25, 0.50, 0.75$; and $\gamma = 0.40, 0.50, 0.60$. Full simulation details are provided in the appendix.

Simulation results

In this section, we present the results of our simulations. We will particularly be interested in investigating the amount of bias and variability, as measured by the standard error, of the estimates of the treatment effect. We then examine simulated coverage probabilities which provide information on the differences in the construction of confidence intervals around the treatment effect. Finally, we look at the effect of screening at baseline on the Type I error and power functions which are important considerations in statistical hypothesis testing.

Bias

For each value of γ , μ , and θ , we averaged the estimated value of γ , $\hat{\gamma}$, over 10,000 replications. We then calculated the *average bias* by subtracting the true value of γ from this average. Figure 1 shows the average bias plotted against μ for two values of γ ($\gamma = 0.40, 0.60$) and θ ($\theta = 0.25, 0.75$). Ideally, we want the bias to be zero. In this case, we see that $\hat{\gamma}$ is positively biased, meaning that the reduction in lesions is slightly more than estimated. This bias is present both when patients are screened for baseline activity and when no screening is performed, although the bias is less pronounced in the presence of screening. Thus, while both procedures tend to overestimate γ (and, consequently, underestimate the amount by which lesion count will be reduced), screening for baseline activity results in estimates that are closer to the true treatment effect, while estimates of treatment effects calculated without screening will be more conservative. The bias should decrease to zero when the sample size gets larger because the estimators for both methods are consistent.

Standard error

For each value of γ , μ , and θ , we calculated the standard deviation of the estimated treatment effects obtained from the 10,000 replications. Figure 2 displays these observed standard errors for two values of γ ($\gamma = 0.40, 0.60$) and θ ($\theta = 0.25, 0.75$). We find that screening for baseline activity reduces the variability in the estimate. This decrease in variability reveals another benefit of screening, since smaller standard errors will result in shorter confidence intervals.

Coverage

While smaller confidence intervals are indeed a desirable property, this potential benefit is secondary to verifying the coverage properties of the intervals calculated after screening for baseline activity. We note that for the screened data, our method of analysis does not take the screening procedure into account. That is, we use a negative binomial distribution to analyze these data despite the fact that these data are not distributed as negative binomial. Therefore we might expect the true coverage of a confidence interval calculated by assuming a negative binomial distribution for the data to be less than the nominal level (i.e. a 95% confidence interval for γ might actually be expected to contain γ less than 95% of the time).

For each value of γ , μ , and θ , we calculated a 95% confidence interval for γ and recorded whether this interval contained the true value of γ . The percentages of replications in which these confidence intervals covered the true value of γ are displayed in Figure 3. We see that interval coverage is consistently less than the nominal level of 95% for both the screened and unscreened data. So, as expected, we do see a loss in confidence interval coverage when screening for baseline activity. In the case of the unscreened data, any loss in coverage cannot be due to model misspecification as we used the correct distribution in our analyses. We suspect that the decreased interval coverage here is due to small sample size.

We see some evidence that the amount of confidence interval coverage lost when screening for baseline activity may depend on the underlying parameters. We find that for small means ($\mu < 2$) interval coverage for the screened data is very close to that of the unscreened data. For example for $\mu = 1$, $\theta = 0.75$, and $\gamma = 0.50$, the confidence intervals coverage is 94.2% for the screened data and 94.0% for the unscreened data. When the mean is larger, we see that interval coverage when screening after baseline activity can be noticeably less than when no screening is performed. To provide another example, for $\mu = 10$, $\theta = 0.75$, and $\gamma = 0.50$, true coverage is 93.9% for the screened data and 95.1% for the unscreened data.

Nevertheless, we note that for all of the parameter values considered, interval coverage was never less than 90% for either procedure, suggesting that confidence interval coverage is relatively robust to the screening procedure.

Type I error

Figure 4 displays the simulated Type I error rates. For the unscreened data and all sets of parameter values, we see that the Type I error rate is larger than the nominal level of $\alpha = 5\%$, suggesting again that a sample larger than $n = 50$ is needed to ensure that the actual Type I error rates are close to the nominal level. For the screened data, we suspect that the increased Type I error rate is likely a result of both small sample size and model misspecification. We find that, for moderate to large mean ($\mu \geq 2$), the Type I error after screening for baseline activity is clearly greater than when no screening is performed. In these cases, screening for baseline activity will increase the probability of erroneously concluding that the treatment effect is statistically significant. For example, for $\mu = 10$ and $\theta = 0.75$, the estimated Type I error rate is 6.4% for the unscreened data and 7.1% for the screened data.

In contrast, it appears that for small mean ($\mu < 2$), the Type I error rate after screening is close to and sometimes even less than that when no screening has been performed. (For $\mu = 1$ and $\theta = 0.75$, the estimated Type I error rate is 6.3% for the unscreened data and 5.7% for the screened data.) In these cases it would appear that there is little to no penalty (i.e. increased Type I error) associated with screening for baseline activity.

Power

We find that screening for baseline activity does result in improved power. For example, for $\mu = 1$, $\theta = 0.75$ and $\gamma = 0.5$, an estimated 180 subjects are required to obtain approximately 80% power, if no screening is performed. In contrast, this level of power can be obtained with a sample size of only 100 if subjects are screened for baseline activity. More detailed results are displayed in the first supplementary figure. These results align well with those reported by Sormani et al.⁹ However, it is important to note that when the mean lesion count is moderate to large, as might be the case if one is investigating intervention in more severe cases, the improvement in power obtained when screening for baseline activity is artificial, as it is likely a result of the inflated Type I error. As discussed above, for small mean lesion count, screening for baseline activity does not appear to inflate the Type I error, so we see a true improvement in power in these cases.

Application

So far, we have focused on assessing whether accurate estimates and proper testing of a treatment effect can be obtained with an enriched sample. In this section, we demonstrate how the above simulation results can be incorporated into trial design. We consider here data reported by Rudick et al.^{12,13} where 172 patients were assigned to either interferon β -1a (AVONEX, Biogen, Inc., Cambridge, MA) or placebo and followed for 2 years. MRI was administered at baseline and annually. Subjects in this study were not selected for baseline

activity. We consider here the lesion counts for Year 2 for only those subjects for whom data is available both at baseline and at Year 2 (81 of 85 subjects in the IFN β -1a group, 82 of 87 in the placebo group). For the sake of illustration, we will design a new study using this data as pilot data. Our task is determining whether this second study should screen subjects for baseline activity.

The first point we need to consider is whether screening for baseline activity will actually enrich our study population. That is, we want to confirm that for this patient population subjects showing at least one lesion on a baseline scan are indeed more likely to respond to treatment than those without any baseline activity. Figure 5 displays the mean lesion count over time by treatment, with subjects separated by baseline activity. We see that subjects exhibiting zero lesions at baseline showed little activity in Years 1 and 2. It follows that a treatment effect will be less detectable in these subjects. Thus, we confirm that screening for baseline activity is likely to enrich our study population and improve power.

Next, we determine the extent to which our estimation of a treatment effect will be affected by screening. Aban et al.⁶ fit a negative binomial distribution to the lesion counts for the Year 2 scan and estimated the model parameters as $\hat{\mu} = 1.646$, $\hat{\theta} = 0.256$, and treatment effect $\hat{\gamma} = 0.495$. As the mean lesion count is small, we expect little loss of coverage as a result of the decision to screen. We repeated the simulation procedure described above using these estimated parameters. For these parameters, we estimate that a nominal 95% confidence interval will have a true coverage of 93.4% if subjects are screened for baseline activity, and a true coverage of 94.8% if no screening is performed. We feel that this difference in coverage is small and conclude that the effect of screening on confidence interval coverage is slight, but that it indeed increases our chance of a Type I error.

Finally, we evaluate the cost effectiveness of an enrichment design versus using a non-enriched sample. From our simulation procedure we find that, for these parameters, in order to have 80% power to detect the predicted treatment effect, we would require 308 participants if no screening is performed and 100 participants if subjects are screened for baseline activity. For the latter study, these 100 participants do not include the individuals who might be given a screening scan only to fail to meet the entry criterion.

While there is no way to know for certain how many subjects must be scanned and rejected before recruitment for the study is complete, we can use the negative binomial distribution to estimate the probability a subject will exhibit zero lesions on the screening scan and in turn use this probability to estimate the number of subjects we will be required to scan. If we assume that lesion count on the baseline scan is distributed as a negative binomial with the parameters estimated by Aban et al.,⁶ then we expect 59.8% of subjects to present with zero lesions at baseline (see Notes). With this rate, on average we will have to scan and reject 149 subjects before reaching the desired sample size of 100. We can say with 95% confidence that the number of rejected subjects will be between 112 and 187.

Thus we conclude that for a study where subjects are given an MRI scan at baseline, randomized to treatment groups and then scanned again at a later date to assess treatment effectiveness, 616 scans (two per subject) are needed to obtain 80% power if subjects are not screened for baseline activity. If subjects are screened, then the same level of power can be obtained with an estimated 349 scans (two scans for each of the 100 individuals included in the study plus the estimated 149 scans of rejected subjects; a 95% confidence interval for the

Notes

Note that this percentage is fairly close to the 54.9% of untreated subjects with zero lesions in the pilot data. The observed rate of zero lesions can also be used to estimate the number of subjects who will need to be given a screening scan. We use the model-based estimate of 59.8% as it is larger and therefore will give a more conservative estimate of this number.

number of scans needed to complete the trial is 312 to 387). In this case we find that the improvement in power offered by screening is enough to offset the increased recruitment effort necessary to collect an enriched sample, and we concluded that a trial that incorporates screening for baseline activity is indeed more cost effective than a trial that does not screen.

Together with the conclusions that the use of screening appears both to enrich the study population and to have little effect on the accuracy of effect size estimation, we make the decision to screen subjects for baseline activity in our hypothetical trial.

Discussion

MRI is an expensive tool for identifying lesions in the brains of patients with multiple sclerosis. Pre-screening subjects for baseline activity is used to improve the power of studies and consequently decrease the total number of MRI scans needed to demonstrate treatment efficacy. While we consider here only one, simplistic screening mechanism, we acknowledge that investigators may use a variety of entrance criteria when recruiting for a clinical trial. For example, an investigator might administer several screening scans, requiring activity on at least one in order for admittance to the trial. These selection methods are all forms of enrichment designs. While the specific conclusions reached in this paper may not be directly applicable to all situations, the take-home message remains the same: which-ever method of enriching the study population is selected, the effect of this enrichment on recruitment effort and accuracy of estimation must be carefully weighed against the likely improvement in power and potential increase of Type I error and the cost of mistakes.

For all parameter values considered, we found that when the negative binomial distribution is used to model lesion counts, screening for baseline activity reduced bias and variability in estimates of the treatment effect. When the mean lesion count is small, we found that confidence interval coverage and Type I error for screened data was comparable to that of the unscreened data. For moderate to large mean, however, we found that screening for baseline activity had an adverse effect on estimation, namely confidence interval coverage and Type I error. An increase in Type I error can be costly by falsely encouraging drug development and wasting valuable resources in a large-scale Phase III trial. Although the screening process decreases interval coverage when the treatment effect is estimated via a negative binomial model, the effects are relatively minor (i.e. coverage is still greater than 90%), indicating that the negative binomial distribution is robust in the face of screening, and suggesting that the effect size confidence interval should be carefully considered before moving forward to a Phase III trial.

In this paper we investigate the effect of screening for baseline activity on inferences about treatment effects in a specific setting: when the negative binomial distribution is used to model lesion counts and make statistical inferences. We emphasize here that the results presented in this paper are only applicable to data that can appropriately modeled using a negative binomial distribution.

In our analyses, we also make the further assumption that treated and control subjects can be adequately modeled using negative binomial distributions with common θ . While some studies have found that θ did not appear to differ by much among treated and control subjects,^{5,8} other studies have concluded that θ can differ significantly for treated and control subjects.^{6,7} Aban et al.⁶ conducted a simulation study and concluded that negative binomial-based methods are fairly robust to the violation of the assumption of common θ for treated and control subjects, as long as the parameter θ for one treatment group is no more

than twice that of the other. We believe that the reasonability of this assumption will depend on the setting and recommend examining each data set carefully to determine, first, whether the application of the negative binomial distribution is appropriate and, second, whether common θ for treated and control subjects can be assumed before applying the results from this paper.

The availability of negative binomial-based generalized linear models not only facilitates the identification of treatment effects via p -values as is expected in clinical research, but also allows the estimation of the effect and its confidence interval as well as detection of changes in those treatment effects over time. Using these techniques we may isolate treatment lags and in future trials where head-to-head therapies are compared, we may be able to assess which drugs are superior based on longer-term results. These later temporal effects of treatment fidelity are very important in treatments for chronic diseases, which may involve decades of therapy.

Screening for baseline activity reduces the bias and variability in the estimation of treatment effects but carries a potential cost of slightly higher Type I errors. Thus, depending on the patient population, screening for baseline activity may create a trade-off between cost effectiveness and conservative inference that must be carefully weighed. This approach, when combined with negative binomial-based modeling, is an appropriate tool for the analysis of MRI-monitored multiple sclerosis studies. Nevertheless, p -values alone should not be cause for moving from Phase II to Phase III.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Richard A Rudick, MD (Director of Mellen Center, Cleveland Clinic Foundation, Cleveland Ohio) for sharing the data used in this paper. The data used in this paper is from a clinical trial supported by NINDS RO1 26321, NMSS RG3099, and Biogen, Inc., Cambridge, MA.

References

1. McFarland HF, Frank JA, Albert PS, et al. Using gadolinium-enhanced magnetic resonance imaging lesions to monitor disease activity in multiple sclerosis. *Ann Neurol*. 1992; 32:758–766. [PubMed: 1471866]
2. Sormani MP, Filippi M. Statistical issues related to the use of MRI data in multiple sclerosis. *J Neuroimaging*. 2007; 17:56S–59S. [PubMed: 17425737]
3. Sormani MP, Bruzzi P, Miller DH, Gasperini C, Barkhof F, Filippi M. Modelling MRI enhancing lesion counts in multiple sclerosis using a negative binomial model: implications for clinical trials. *J Neurol Sci*. 1999; 163:74–80. [PubMed: 10223415]
4. Sormani MP, Bruzzi P, Rovaris M, et al. Modelling new enhancing MRI lesion counts in multiple sclerosis. *Mult Scler*. 2001; 7:298–304. [PubMed: 11724445]
5. van den Elskamp IJ, Knol DL, Uitdehaag BMJ, Barkhof F. The distribution of new enhancing lesion counts in multiple sclerosis: further explorations. *Mult Scler*. 2009; 15:42–49. [PubMed: 18845655]
6. Aban IB, Cutter GR, Mavinga N. Inferences and power analysis concerning two negative binomial distributions with an application to MRI lesion counts data. *Comput Stat Data Anal*. 2009; 53:820–833. [PubMed: 19177180]
7. Sormani MP, Bruzzi P, Beckmann K, et al. The distribution of magnetic resonance imaging response to interferon β -1b in multiple sclerosis. *J Neurol*. 2005; 252:1455–1458. [PubMed: 16021360]

8. Sormani MP, Bruzzi P, Comi G, Filippi M. The distribution of the magnetic resonance imaging response to glatiramer acetate in multiple sclerosis. *Mult Scler.* 2005; 11:447–449. [PubMed: 16042228]
9. Sormani MP, Miller DH, Comi G, et al. Clinical trials of multiple sclerosis monitored with enhanced MRI: new sample size calculations based on large data sets. *J Neurol Neurosurg Psychiatry.* 2001; 70:494–499. [PubMed: 11254773]
10. Comi G, Filippi M, Wolinsky JS. and the European/Canadian Glatiramer Acetate Study Group. European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of glatiramer acetate on magnetic resonance imaging-measured disease activity and burden in patients with relapsing multiple sclerosis. *Ann Neurol.* 2001; 49:290–297. [PubMed: 11261502]
11. Temple RJ. Special study designs: early escape, enrichment, studies in non-responders. *Commun Stat Theory Meth.* 1994; 23:499–531.
12. Rudick RA, Cutter G, Baier M, et al. Use of the multiple sclerosis functional composite to predict disability in relapsing MS. *Neurology.* 2001; 56:1324–1330. [PubMed: 11376182]
13. Jacobs LK, Cookfair DL, Rudick RA, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). *Ann Neurol.* 1996; 39:285–294. [PubMed: 8602746]

Appendix

Simulation procedure

We repeated the following procedure 10,000 times:

1. Create a pool of potential subjects by drawing their latent means from the gamma distribution:

$$\lambda_k^* \sim \text{Gamma}(\theta, \theta/\mu), k=1, \dots, N.$$

(N should be sufficiently large that at least 50 of the potential subjects will exhibit baseline activity. We used $N=2000$.)

2. Simulate the results of the baseline scan:

$$b_k^* \sim \text{Poisson}(\lambda_k^*), k=1, \dots, N.$$

3. Out of the subjects such that $b_k^* > 0$, randomly select 25 and assign them to the control group and label their latent means as $\lambda_1, \dots, \lambda_n$. Randomly assign another 25 subjects to the treatment group and denote their latent means as $\lambda_{n+1}, \dots, \lambda_{2n}$.

4. Simulate the experimental results:

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i), i=1, \dots, n \\ y_i &\sim \text{Poisson}(\gamma\lambda_i), i=n+1, \dots, 2n \end{aligned}$$

We estimated γ by calculating the ratio of the sample mean for controls to the sample mean for treated subjects. The standard error of this estimate can be obtained by fitting a negative binomial log-linear model to the experimental results. For each replication, we estimated γ and then recorded whether the 95% confidence interval for γ contained the true value of γ . For each value of μ , θ , and γ , we counted the number of replications for which the null hypothesis $H_0: \gamma = 1$ was rejected at a significance level of $\alpha = 5\%$. We performed this procedure for $\mu = 1, 2, \dots, 12$; $\theta = 0.25, 0.50, 0.75$; and $\gamma = 0.40, 0.50, 0.60$. In order to estimate the Type I error rate, we also performed this procedure for the values of μ and θ

listed above and $\gamma = 1$; we then calculated the proportion of replications in which the null hypothesis of no treatment effect was erroneously rejected.

For the purposes of comparison, we repeated this procedure without screening for baseline activity. That is, in Step 3, 50 subjects were selected from the entire pool of potential subjects, rather than just those with positive baseline values.

Figure 1. Average bias in estimated values of γ .(A) $\gamma = 0.4$, $\theta = 0.25$, (B) $\gamma = 0.4$, $\theta = 0.75$, (C) $\gamma = 0.5$, $\theta = 0.25$, and (D) $\gamma = 0.5$, $\theta = 0.75$.

Figure 2. Observed standard error of the estimates of γ .(A) $\gamma = 0.4$, $\theta = 0.25$, (B) $\gamma = 0.4$, $\theta = 0.75$, (C) $\gamma = 0.5$, $\theta = 0.25$, and (D) $\gamma = 0.5$, $\theta = 0.75$.

Figure 3.

Estimated true interval coverage.

(A) $\gamma = 0.4, \theta = 0.25$, (B) $\gamma = 0.4, \theta = 0.75$, (C) $\gamma = 0.5, \theta = 0.25$, and (D) $\gamma = 0.5, \theta = 0.75$.

Figure 4.
Simulated Type I error.
(A) $\theta = 0.25$, (B) $\theta = 0.50$, and (C) $\theta = 0.75$.

Figure 5.
Mean lesion count (with standard error bars) by time, treatment, and baseline activity.