# Evidence synthesis for count distributions based on heterogeneous and incomplete aggregated data

Christian Röver[1] Stefan Andreas[2,3] Tim Friede[1]

November 19, 2014

**Abstract**

The analysis of count data is commonly done using Poisson models. Negative binomial models are a straightforward and readily motivated generalization for the case of overdispersed data, i.e., when the observed variance is greater than expected under a Poissonian model. Rate and overdispersion parameters then need to be considered jointly, which in general is not trivial. Here we are concerned with evidence synthesis in the case where the reporting of data is rather heterogeneous, i.e., events are reported either in terms of mean event counts, the proportion of event-free patients, or rate estimates and standard errors. Either figure carries some information about the relevant parameters, and it is the joint modeling that allows for coherent inference on the parameters of interest. The methods are motivated and illustrated by a systematic review in chronic obstructive pulmonary disease.

## 1 Introduction

Count data commonly occur as endpoints in clinical trials, for example, when one is interested in inferring rates of recurrent events. Accordingly, these types of data are frequently encountered in meta analysis, when results from different studies are integrated. The Poisson distribution is often applicable when modeling event counts that are associated with a certain *rate* which then is usually the quantity of interest. The negative binomial distribution arises as a straightforward generalisation of the Poisson distribution (Lawless, 1987); it results as a marginal distribution of counts when the corresponding "Poissonian" rate is not fixed, but is associated with some uncertainty taking the mathematical form of a Gamma distribution. The introduced additional variability, the *overdispersion*, induces extra variation in the data and uncertainty in resulting estimates. The consideration of overdispersion is often useful and necessary in order to account for heterogeneity of some kind in the data.

Clinical trials in the context of *chronic obstructive pulmonary disease (COPD)*, are commonly concerned with count data. The progress of COPD is characterized by recurrent *exacerbations*, periods of rapid worsening of the disease. The

[1]Department of Medical Statistics, University Medical Center Göttingen, Humboldtallee 32, 37073 Göttingen, Germany

[2]Lungenfachklinik Immenhausen, Robert-Koch-Straße 3, 34376 Immenhausen, Germany

[3]Clinics for Cardiology and Pulmonology, University Medical Center Göttingen, Robert-Koch-Straße 40, 37099 Göttingen, Germany

presence of exacerbations, or their counts, are often used as clinical endpoints, as treatments are supposed to delay or prevent the occurence of exacerbations. Negative binomial models are commonly advocated and used in the context of COPD exacerbations (Keene et al., 2007, 2008; Aaron et al., 2008; European Medicines Agency (EMA), 2012), but while the presence of overdispersion and the importance of its consideration appears to be undoubted, published evidence on its actual magnitude on the other hand is rather sparse. For example, Anzueto et al. (2009) use "*an over-dispersion estimate of 1.5*" and Calverley et al. (2009) use "*a correction for overdispersion of 2*", while the actual conventions used for quantifying overdispersion may be ambiguous. Also, the way in which results are quantified varies considerably across studies; many studies quote exacerbation rate estimates, some provide standard errors (or confidence intervals) in addition, and some studies again quote numbers (or proportions) of patients with and without an exacerbation during the study.

Differences in the reporting of study results commonly pose problems when combining information sources that aggregate data not in the same way. Off-the-shelf software may then not be suitable to deal with the peculiarities of a given problem. The generic case of combining data of a common format will here be referred to as *meta analysis* while the more general case of combining data in different formats or possibly originating from different experimental designs is commonly called *evidence synthesis* (Spiegelhalter et al., 2004). Our aim was to set up a model to infer both unknowns while incorporating the different sources of data in a coherent manner. Here we aim for a model that consistently based on an underlying negative binomial process, to which the different data sources are linked in individual ways. The eventual formulation of a coherent data likelihood then allows for consistent inference based on all available information.

In our case, both rate *and* overdispersion are important parameters determining the possible trial outcomes, while neither one is usually the figure of primary interest – the actual focus usually lies on quantities like *rate ratios*. If one knows the total number of exacerbations (or a rate estimate) and the number of zero-counts (exacerbation-free patients), one can immediately estimate rate and overdispersion jointly (e.g. via a maximum-likelihood approach). If only one of the figures is given, only one parameter may be estimated for the other being given or fixed. While we do not know any of the parameters precisely, we have some information on their probable order of magnitude from related studies. The idea now is to utilize a Bayesian framework (Sutton and Abrams, 2001; Spiegelhalter et al., 2004) to coherently incorporate the information from different sources. This way we are able incorporate exact likelihoods into the joint model, and to "borrow strength" among studies in order to make sense of the limited amount of information that may be contained in any single one.

The outline of the paper is as follows. In Section 2, the context of count data in modeling exacerbation counts in COPD is introduced. Section 3 describes Poisson and negative binomial models for count data in general and the problems arising with data encountered in studies investigating COPD. The proposed approach to data modeling is developed, including a hierarchical model for the meta-analysis. In Section 4, the model is applied to actual data, and model variations are explored in order to investigate the impact of considering additional data.

# 2 Motivating example: Meta-analysis in COPD

Chronic obstructive pulmonary disease (COPD) is a major cause of death and disability worldwide, and the burden of this disorder will continue to increase in the coming decades despite therapeutic advances (Decramer et al., 2012). Thus there is substantial need to improve symptomatic and prognostic burden in COPD. Randomized controlled trials (RCTs) investigating a primary endpoint are the gold standard in proving efficacy of one therapy over another or over placebo. Lung function such as forced expiratory volume in 1 second ($FEV_1$) is a physiological measure and a global marker of disease severity in COPD. However, $FEV_1$ and other measures of lung function correlate poorly with patient-related outcomes such as symptoms, exercise tolerance, quality of life, exacerbations and mortality. Mortality is a key endpoint, but two COPD trials failed to show significant effects on mortality, despite including about 6 000 COPD patients each (TORCH (Calverley et al., 2007), UPLIFT (Tashkin et al., 2008)). Exacerbations of COPD relate to mortality, impaired quality of life, lung function and health care costs. Furthermore, the chronic and progressive course of COPD is frequently aggravated by exacerbations (Decramer et al., 2012). Thus many phase 3 studies exploit COPD exacerbations as the primary endpoint (Cazzola et al., 2008). There are different definitions of exacerbations mostly categorized by severity and the use of health care resources. Although all definitions harbour some subjective aspect, reproducibility and validity within one study is generally high. Problems arise, if different studies are accumulated and compared to each other as done in pooled- or meta-analyses (Suissa, 2006), which are usually addressed by the inclusion of variance components reflecting the heterogeneity in the statistical model.

We are currently working on a meta-analysis comparing different treatments in COPD (Röver et al., in preparation). In order to find relevant studies, we performed a literature search using PubMed; in addition we considered studies that were cited in the meta-analyses by Puhan et al. (2009); Mills et al. (2011); Fleer et al. (2012); Karner et al. (2012); Dong et al. (2013). In the present analysis, we will focus on the subset of studies comparing *long-acting muscarinic antagonist (LAMA)* with placebo treatment.

# 3 Statistical model

## 3.1 Notation and basic properties

A Poisson model naturally arises as a model for event counts that are associated with a certain rate (Spiegelhalter et al., 2004, Ch. 2.6). The Poisson model is specified through a single parameter, the *rate* $\lambda > 0$, which is usually to be interpreted *per unit time*, i.e., as a *yearly* rate, for example. The number of events occuring over a duration $\delta$ then follows a Poisson distribution with mean $\delta\lambda$; the expected number as well as its variance are equal to $\delta\lambda$. The negative binomial distribution in addition possesses an overdispersion parameter $\varphi \geq 0$. For $\varphi = 0$, the model again simplifies to the Poissonian. With $\varphi > 0$ the distribution's expectation remains the same, but the variance increases to $\delta\lambda(1 + \varphi\delta\lambda)$ (Lawless, 1987; Hilbe, 2011). The negative binomial model arises as a Poisson mixture distribution, where, instead of being constant, the rate is

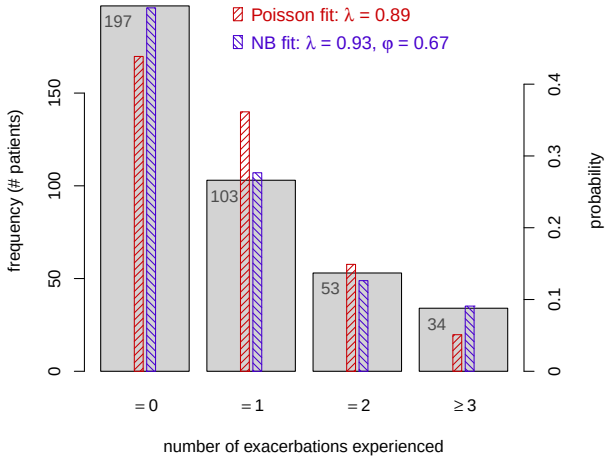


Figure 1: An example of actual exacerbation counts observed in placebo-treated patients over a duration of 48 weeks taken from Sethi et al. (2010). A negative binomial model fits the data better than a Poisson model as it allows for a greater fraction of extreme (i.e., zero or large) counts, and accordingly also fits better in the moderate range. 95% confidence intervals for the Poisson and negative binomial rates are [0.79, 0.99] and [0.80,1.07], respectively. The overdispersion's 95% confidence interval is [0.30, 1.04].

a Gamma distributed random variable with expectation $\lambda$ and variance $\lambda^2\varphi$, so that the rate's coefficient of variation is $\sqrt{\varphi}$; this way the negative binomial model is able to account for extra-Poissonian heterogeneity in the data.

## 3.2 Illustrating example

Figure 1 provides an illustration of the difference between Poisson and negative binomial models in comparison with actual count data. The bars represent the observed frequencies of exacerbation counts in an actual study (Sethi et al., 2010). The red and blue lines show the best-fitting Poisson and negative binomial models (fitted via maximum likelihood) for the given data. One can see that while the two rate parameters barely differ, the additional consideration of a non-zero overdispersion provides a much better fit to the data, as it allows to account for the increased number of "extreme" outcomes (zero or large exacerbation counts) that would be unlikely to occur under a plain Poissonian model. In the negative binomial model, the standard error associated with the rate estimate is larger than in the Poisson model by a factor of a third (0.068 instead of 0.051). Looking at the data, one can already see that the two commonly quoted summary statistics of (mean) rate and the number of exacerbation-free patients carry somewhat complementary pieces of information which in combination allow to infer both $\lambda$ and $\varphi$.

### 3.3 Parameter estimation and sufficient statistics

When we assume a Poisson model for the observed event counts $x_i$ in a group of $n$ patients that were all observed over durations $\delta_i$, then all that is required in order to infer the unkown rate $\lambda$ is the total count $t = x_1 + \ldots + x_n$ (and the total observation time $\delta_t = \delta_1 + \ldots + \delta_n$), as the total constitutes the *sufficient statistic* in this case. These numbers are rarely explicitly quoted, but being provided with a rate estimate $\hat{\lambda} = t/\delta_t$ (which is both the maximum-likelihood and moment estimator) and assuming a constant observation duration ($\delta_1 = \ldots = \delta_n = \delta$), one can again directly infer the total $t$. In the case of a negative binomial distribution on the other hand, there is no simple form for a sufficient statistic; ideally one would need the complete data (all individual counts $x_i$) in order to derive for example maximum-likelihood estimates (Lawless, 1987). Only for the case of constant durations $\delta_i = \delta$, the maximum-likelihood estimator for the rate is the same as in the Poisson case, and there is also a moment estimator for the overdispersion available, although its properties are questionable, as it may also turn out negative.

Another common alternative approach to modeling count data of the present kind is to consider *odds ratios*, i.e., to compare chances of an exacerbation (at least one event vs. no event) between treatment groups; this approach is used e.g. by Puhan et al. (2009); Fleer et al. (2012); Karner et al. (2012). However, if we assume an underlying Poisson or negative binomial process, the odds ratio only equals the rate ratio in the limit of small rates or small exposure durations. Otherwise odds ratios and rate ratios will in general differ, depending on the exposure duration and the amount of overdispersion present, and hence are not directly comparable (see Appendix A.1 for an explicit derivation).

### 3.4 Available data

In the studies that we will investigate in the following, the most commonly used figures for quantifying disease severity are the number of exacerbation-free patients, rate estimates, partly in conjunction with confidence intervals or standard errors, or the median time until the first exacerbation. Standard errors and confidence intervals can easily be converted into one another (Higgins and Green, 2011), and numbers of exacerbation-free patients are commonly derived from the total number of patients and the provided percentage. For a given rate estimate $\hat{\lambda}$, we assume that it results as the total count $t$ divided by the cumulative observation time $nd$ in order to infer the total number as $t = \hat{\lambda} n \delta$.

### 3.5 Marginal and joint modeling of total counts and zero-counts

Suppose that a total of $n$ patients are observed in a trial of duration $\delta$. Each patient's event count follows a negative binomial distribution with rate $\delta\lambda$ and overdispersion $\varphi$. Let $T$ be the total number of events observed; by the central limit theorem its marginal distribution is approximately normal with moments

$$\mathrm{E}[T] \;=\; n\delta\lambda \qquad \text{and} \qquad \mathrm{Var}(T) \;=\; n\delta\lambda\,(1 + \varphi\delta\lambda). \tag{1}$$

The marginal distribution of the number of "zero-counts", the number of event-free patients, $Z$, is Binomial with size $n$ and probability $\pi_0$ given by the

probability of a zero outcome for the negative binomial model,

$$\pi_0 \;=\; \begin{cases} (1+\varphi\delta\lambda)^{-\frac{1}{\varphi}} & \text{for } \varphi > 0 \\ \exp(-\delta\lambda) & \text{for } \varphi = 0. \end{cases} \tag{2}$$

$T$ and $Z$ are not stochastically independent, rather they are negatively correlated; i.e., when there is a large number of zeroes in the data, we expect a small total count. An approximate joint probability distribution of $T$ and $Z$ can be derived by extending the ideas above. We can split the joint probability density into the factors $p(t,z) = p(z) \times p(t|z)$. The conditional density $p(t|z)$ describes the probability distribution of the total count $T$ for which we know the number $Z=z$ of zeroes in the total sum of negative binomial samples. The total $T$ thus is constituted of $n-z$ summands that are drawn from a "truncated" negative binomial distribution excluding the zero outcome. We can again determine the first two moments of the truncated distribution and use the normal approximation as above; see Appendix A.2 for the explicit derivation. The marginal and joint distributions of $T$ and $Z$ then define the likelihood function for the data that is used in the following.

## 3.6 The hierarchical model

In the following we set up a hierarchical model with study-level rate and overdispersion parameters, allowing for heterogeneity between studies in both parameters. Depending on the particular type of data provided, different studies then contribute different pieces of information on both parameters.

We have data on $k$ studies investigating both treatment and placebo. Each study $i$ has an associated study duration $\delta_i$ and is comprised of $\ell_i$ study arms. The $i$th study's $j$th arm corresponds to a treatment category indexed by $m_{ij}$ ($m_{ij} \in \{0,1\}$, where $m_{ij} = 0$ indicates a placebo group), and has a number $n_{ij}$ of patients associated. Some study arms provide a rate estimate $\hat{\lambda}_{ij}$ together with an associated standard error, some provide only their total exacerbation count $T_{ij}$, (corresponding to $\hat{\lambda}_{ij} n_{ij} \delta_i$ for some rate estimate $\hat{\lambda}_{ij}$), some provide the number $Z_{ij}$ of zero-counts (exacerbation-free patients), and some provide both $T_{ij}$ and $Z_{ij}$. Studies directly providing rate estimates and standard errors are considered via the common normal approximation to the likelihood (Spiegelhalter et al., 2004). Studies quoting one or both of $T$ and $Z$ are considered in the analysis via the likelihood function described in Section 3.5. While one might use the information on rates and standard errors within a negative binomial model (which might in fact be considered an overall more consistent approach), here we used the normal approximation in order to demonstrate the gain compared to the most basic "classical" approach.

The hierarchical model is parameterized as follows. The rate for the $j$th arm of the $i$th study is simply given by $\lambda_i$ if it is a placebo group (i.e., if $m_{ij} = 0$), or otherwise by $\lambda_i \times \vartheta \times \psi_{ij}$. The parameter of primary interest is $\vartheta$, the treatment effect, and $\psi_{ij}$ is a random effect accounting for particularities in treatment, dosing, etc. differing between studies or study arms. The overdispersion in the $i$th study is given by $\varphi_i$.

The prior for each trial's "placebo" rate $\lambda_i$ is normal on the log-scale (i.e., log-normal):

$$\log(\lambda_i) \;\sim\; \text{Normal}\big(\mu_\lambda, \sigma_\lambda^2\big). \tag{3}$$

The corresponding hyperparameters ($\mu_\lambda$, $\sigma_\lambda^2$) are assigned vague uniform prior distributions on the log scale:

$$\mu_\lambda \sim \text{Unif}\big(\log(10^{-2}), \log(10^{2})\big), \quad \log(\sigma_\lambda) \sim \text{Unif}\big(\log(10^{-3}), \log(10)\big). \quad (4)$$

Similarly, each trial's overdispersion $\varphi_i$ also has a normal prior on the log-scale, i.e.,

$$\log(\varphi_i) \sim \text{Normal}\big(\mu_\varphi, \sigma_\varphi^2\big), \quad (5)$$

again with similar priors for the hyperparameters ($\mu_\varphi$, $\sigma_\varphi^2$):

$$\mu_\varphi \sim \text{Unif}\big(\log(10^{-4}), \log(10^{4})\big), \quad \log(\sigma_\varphi) \sim \text{Unif}\big(\log(10^{-3}), \log(10)\big). \quad (6)$$

For the treatment effect we use a normal prior that is centered around a neutral effect (i.e., $\vartheta = 1$) and loosely confined to within the range $[1/10, 10]$ with $\approx 90\%$ probability, i.e.

$$\log(\vartheta) \sim \text{Normal}\big(0, \log(4)^2\big). \quad (7)$$

For each arm's random effect we again use a vague and also heavy-tailed Student-$t$-prior with 3 degrees of freedom centered around $\psi_{ij} = 1$, i.e., no effect:

$$\log(\psi_{ij}) \sim \text{t}_3\big(0, \sigma_\psi^2\big). \quad (8)$$

The heavy-tailed prior here is meant to allow for individual odd studies and to bound their influence on the overall result. The scale parameter $\sigma_\psi$ again is an unknown with a uniform prior on the log scale:

$$\log(\sigma_\psi) \sim \text{Unif}\big(\log(10^{-3}), \log(10)\big). \quad (9)$$

The parameters accounting for heterogeneity ($\sigma_\lambda$, $\sigma_\varphi$ and $\sigma_\psi$) here are modelled on their logarithmic scales. In a related setting, similar priors (uniform for the logarithmic heterogeneity) have been investigated among a range of other uninformative or weakly informative choices, which were all found to yield comparable inferences as long as the number of considered studies was not too low (Lambert et al., 2005). The priors used here are truncated versions of the (otherwise improper) Jeffreys priors for location and scale parameters (Jaynes, 1968; Jeffreys, 1946). The study-specific rate and overdispersion parameters ($\lambda_i$ and $\phi_i$) and the arm-specific random effect ($\psi_{ij}$) are for now treated as a priori independent.

# 4 Application to COPD example

## 4.1 Investigated studies

Our literature search resulted in 24 placebo-controlled studies investigating LAMA treatment and providing information on exacerbations. The resulting numbers of studies providing total exacerbation counts and/or the number of zero-counts (exacerbation-free patients) are shown in Table 1. Based on the different types of information provided, we defined 3 subsets of the data, namely: (A) the studies providing rate estimates and standard errors, (B) the above studies, plus the ones giving both total counts *and* zero-counts, and (C) the above studies, plus the ones giving either total counts *or* zero-counts. The (overlapping) data subsets (A), (B) and (C) consist of 4, 12, and 24 studies, respectively. All data are shown in Table 2.

Table 1: Four of the studies found provide rate estimates along with standard errors. The below table shows the numbers of studies *without standard errors*, but providing a total count (or rate) or the number of zero-counts. Another 8 studies provide both, and also considering studies only giving one of the two, we can include another 12 additional studies in the analysis.

| | | *total count provided?* | |
|---|---|---|---|
| | | No | Yes |
| *zero count* | No | 0 | 3 |
| *provided?* | Yes | 9 | 8 |

## 4.2 Implementation

We utilized the *JAGS* software (Plummer, 2003) in conjunction with the "rjags" package (Plummer, 2013) in order to carry out stochastic (Monte Carlo) integration of the unknown parameters' posterior distribution (Gelman et al., 1997). In order to ensure convergence of the algorithm, we ran four MCMC chains in parallel, starting from different overdispersed initial values. Chains were run for $10^7$ iterations, where the first 10% of samples were discarded as burn-in. Correlation between subsequent samples was reduced by using only every 100th sample. Convergence of parallel chains to the same mode was checked using the (multivariate) potential scale reduction factor (Brooks and Gelman, 1998), which was well below 0.1% in all cases. In the following, we show estimates, marginal posterior probability density functions etc. based on these Monte Carlo samples. The densities shown in figures 2 and 3 are based on kernel density estimates derived from the samples.

## 4.3 Results

We applied the above model to the complete data, and its subsets as described in Section 4.1. Figure 2 shows the marginal posterior distribution of the parameter of primary interest, the treatment effect $\vartheta$ for all three scenarios. One can see that the posterior distributions based on different data subsets are consistent with each other, and that with the inclusion of additional studies the evidence in favour of a substantial treatment effect is increased.

While only approximately comparable (see Section 3.1), the effect on the annualized exacerbation rate is of the same order of magnitude as the odds ratio estimates found in related meta-analyses (Puhan et al. (2009): 0.71, Fleer et al. (2012): 0.76, Karner et al. (2012): 0.78). Performing a simple random-effects meta-analysis (Viechtbauer, 2010) on the (logarithmic) rate ratios of subset (A) yields a combined estimate for the treatment effect of 0.69 and a 95% confidence interval of [0.52, 0.93], which is in line with the results from the Bayesian analysis yielding a median of 0.82 and credibility interval [0.56, 0.92]. Interestingly, the estimates differ by half a standard error, while the confidence / credibility limits still coincide. With the consideration of 8 additional studies reporting both a rate estimate and the proportion of event-free patients, the estimate changes to 0.73 [0.63, 0.83], and the inclusion of 12 studies providing one of the two figures narrows the estimate further down to 0.73 [0.65, 0.80].

The posterior distribution of the heterogeneity parameter $\sigma_\psi$ (right panel of

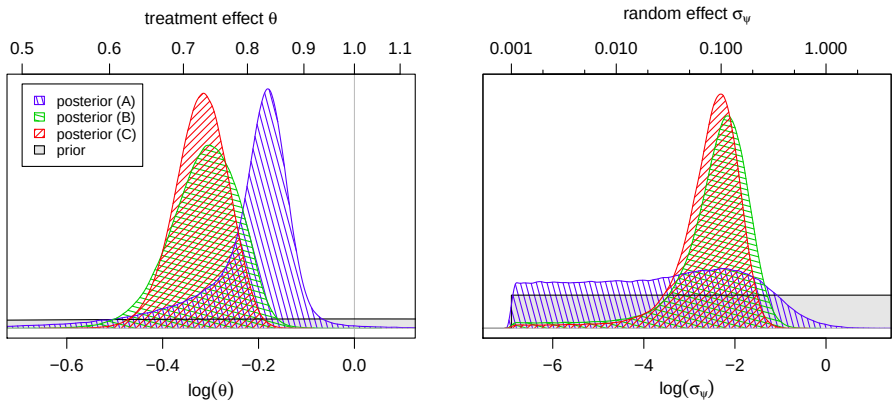


Figure 2: Marginal posterior probability density functions (kernel density estimates) based on different amounts of data. The prior density is also shown in comparison.

Figure 2) shows that the observed data allow to constrain the heterogeneity's magnitude to an order of $\approx 0.1$ or below. While using only data subset (A) we do not gain much information beyond that on this parameter, the inclusion of additional data allows to constrain the heterogeneity to be of a magnitude larger than, say $\approx 0.01$.

Figure 3 illustrates the information we gain about the rates and the amounts of overdispersion encountered in the analyzed studies. Again, considering different subsets of the studies, we gain more certainty about the four hyperparameters $(\mu_\lambda, \sigma_\lambda)$ and $(\mu_\varphi, \sigma_\varphi)$ describing mean and variability in rate and overdispersion among different studies with the inclusion of additional studies. While the resulting posterior distributions differ, they are consistent and exhibit significant overlap. Most notably, in scenario A where we only consider studies that provide rate estimates along with standard errors, we do not gain any information on the overdispersion $\varphi$, so that here the posterior equals the prior. The right panel shows the posterior predictive distribution of rate and overdispersion $(\lambda^*, \varphi^*)$ of a "new" study, illustrating what we have learned from the present set of studies about probable characteristics of an additional study from the same population; this would be of interest e.g. when planning a future study.

We also investigated whether we could find any particular reason for the differences in effect estimates based on different data subsets. While the results do not look inconsistent, it is interesting that the analyses based on subsets (B) and (C) tend to indicate a stronger treatment effect. One notable feature of studies in subset (A) is that these tend to be fairly long studies in comparison.

# 5 Discussion

In the context of meta-analysis, one is commonly faced with the problem of coherently reconciling data from different sources within a single model in order to infer quantities of interest. Here we were able to utilize data in the form of rate

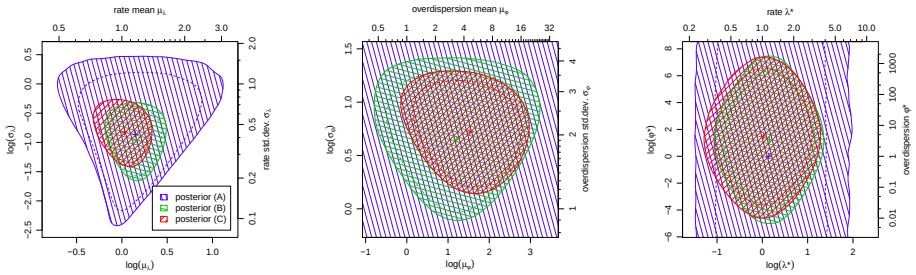


Figure 3: The left two panels show credibility areas for the four "population" rate and overdispersion hyperparameters $(\mu_\lambda, \sigma_\lambda)$ and $(\mu_\varphi, \sigma_\varphi)$. Solid lines enclose 95%, and dashed lines 90% probability; the crosses indicate median values. The right panel illustrates the resulting posterior predictive distribution for the parameters in a "new" study $(\lambda^*, \varphi^*)$.

estimates and odds ratios, and, everything being based on an underlying negative binomial model, differing follow-up times are not an issue, overdispersion is addressed, and inference may be concentrated directly on the parameter of primary interest, the treatment effect on the event rate. The use of a Bayesian framework easily allowed for a flexible specification of correlation structures, including random effects at different levels. The joint likelihood formulation here allowed for the coherent incorporation of information in terms of quoted event rates or proportions of event-free patients. This rendered those studies providing both figures most valuable, but also enabled the inclusion of a considerable number of studies providing only one of the two. The utilization of additional data sources improved parameter estimates and reinforced the results' validity.

The approach may in future be extended further by providing an interface for data in terms of estimated survival times (e.g., median time to first event), which would here need to be implemented via the corresponding distribution of survival times (Exponential distribution in case of a Poisson, and a Lomax distribution in case of a negative binomial model (Siri et al., 2012)). While by now we incorporated studies providing rates and standard error via a conventional normal approximation to the likelihood, this could as well be implemented via another (approximate) negative binomial likelihood, e.g., by assuming that standard errors are based on the emprical standard deviation, in which case this figure again carries information on the amount of overdispersion. For the moment we assumed independence between rate and overdispersion parameters, but introducing correlation between the two may also constitute a plausible and useful extension, adding robustness to the approach. Bivariate meta-analyses of the rate and overdispersion estimates for placebo control groups and active treatments, however, do not indicate the presence of correlation between the two parameters in our example data. Similarly, one may also question the assumption of equal overdispersion across arms within a single study. For our example, we checked the assumption by performing a meta-analysis based on placebo- and treatment-arm overdispersions, which does not provide evidence against the assumption of a common overdispersion parameter across groups. However, the chances of resolving such issues based on the given data may be

small, such investigations may be more realistically addressed based on individual patient data. Another further extension of the presented method would be to utilize information on potential missingness of data that may be deducible from the given set of studies (e.g. Duval and Tweedie, 2000; Copas, 2013; Copas et al., 2014). Up to now the type of measure reported (rate and/or zero counts) is assumed to be independent of the particular trial's outcome. We have already seen that the way of reporting may be connected with other features like study duration (see Sec. 4.3). It would be interesting to consider additional information (e.g. from study protocols) in order to test for potential correlations and biases here. The general problem remains to meet the two-fold challenge of first *finding* all relevant data and then also *utilizing* it in a coherent mannner. Modeling would of course be much simplified if data were available not in the form of study-level summary statistics, but in terms of detailed individual-patient data. While the fully Bayesian approach using MCMC methods yields accurate results, it is computationally rather costly, so it may be worth investigating alternative approaches (like empirical Bayes quasi-likelihood methods) in comparison. It would have been of interest to validate the approach with external individual patient data, however, we did not have access to this kind of data and cannot pursue this line of investigation. As suggested by a referee, a leave-one-out cross validation would be an excellent alternative validation, however, due to the associated computational costs, this is currently not feasible; we will further pursue this matter, possibly also by implementing checks via predictive distributions, along with the investigation of computational speed-ups. It should also be relatively easy to carry over the approximations used here to related models, for example to other generalizations of the Poisson distribution like the zero-inflated Poisson distribution.

**Acknowledgement**

Part of the work was funded by a unrestricted contribution (Oskar und Helene Medizinpreis 2010) to SA.

**Conflict of interest**

The authors have declared no conflict of interest.

# Appendix

## A.1 Rate ratios and odds ratios

Consider the odds for the occurrence of at least one event in the Poisson model. We have: $\mathrm{P}(X = 0) = \exp(-d\lambda)$ and $\mathrm{P}(X \geq 1) = 1 - \exp(-\delta\lambda)$. Assuming a multiplicative treatment effect $\vartheta < 1$ on the rate, the odds ratio is

$$\frac{1 - \exp(-\delta\vartheta\lambda)}{\exp(-\delta\vartheta\lambda)} \times \frac{\exp(-\delta\lambda)}{1 - \exp(-\delta\lambda)} \quad = \quad \frac{\exp(\delta\vartheta\lambda) - 1}{\exp(\delta\lambda) - 1} \tag{10}$$

which is $\approx \vartheta$ for small $\delta\lambda$, and smaller otherwise, i.e., the effect on odds ratio is larger than the effect on rate, and the effect will appear more pronounced for longer studies.

In the negative binomial model, we have: $\mathrm{P}(X=0) = (1+\varphi\delta\lambda)^{-\frac{1}{\varphi}}$ and $\mathrm{P}(X \geq 1) = 1-(1+\varphi\delta\lambda)^{-\frac{1}{\varphi}}$. Again assuming a treatment effect $\vartheta < 1$, the odds ratio is

$$\frac{1-(1+\varphi\delta\vartheta\lambda)^{-\frac{1}{\varphi}}}{(1+\varphi\delta\vartheta\lambda)^{-\frac{1}{\varphi}}} \times \frac{(1+\varphi\delta\lambda)^{-\frac{1}{\varphi}}}{1-(1+\varphi\delta\lambda)^{-\frac{1}{\varphi}}} \quad = \quad \frac{(1+\varphi\delta\vartheta\lambda)^{\frac{1}{\varphi}}-1}{(1+\varphi\delta\lambda)^{\frac{1}{\varphi}}-1} \tag{11}$$

which is $\approx \vartheta$ for small $\delta\lambda$, $< \vartheta$ for $\varphi < 1$, $= \vartheta$ for $\varphi = 1$ and $> \vartheta$ for $\varphi > 1$, i.e., the effect on the odds ratio again is approximately the same as the effect on $\theta$ for small event rate $\lambda$ or exposure duration $\delta$, and whether the effect in terms of odds ratio is smaller or greater than in terms of $\vartheta$ otherwise depends on the amount of overdispersion $\varphi$.

## A.2 Marginal and joint distributions of total and zero counts

Suppose we have negative binomial random variables $X_1, \ldots, X_n$ with rate $\lambda$, overdispersion $\varphi$ and a common corresponding exposure duration $\delta$. The (approximate) marginal distributions of the total count $T$ and the number of zero-counts $Z$ are described in Section 3.5. The *joint* distribution of $T$ and $Z$ may be derived by noting that the density function may be factorised as $p(t,z) = p(z) \times p(t|z)$. The *conditional* distribution of $T|Z$ is essentially that of a sum of $n-z$ draws from a "truncated" negative binomial distribution only taking positive values, i.e., the conditional distribution of a negative binomial draw $X$ given that $X \neq 0$. Instead of the moments given in (1), the "truncated" distribution instead has the expectation

$$\mathrm{E}[X|X>0] \;=\; \tfrac{1}{1-\mathrm{P}(X=0)} \sum_{j=1}^{\infty} j \times \mathrm{P}(X=j) \;=\; \tfrac{\mathrm{E}[X]}{1-\mathrm{P}(X=0)} \tag{12}$$

which for the general negative binomial case yields:

$$\mathrm{E}[X|X>0] \;=\; \begin{cases} \frac{\delta\lambda}{1-(\frac{1}{1+\varphi\delta\lambda})^{1/\varphi}} & \text{for } \varphi > 0 \\ \frac{\delta\lambda}{1-\exp(-\delta\lambda)} & \text{for } \varphi = 0. \end{cases} \tag{13}$$

The variance is

$$\begin{aligned}
\mathrm{Var}(X|X>0) \;&=\; \sum_{j=1}^{\infty} \big(j - \mathrm{E}[X|X>0]\big)^2 \times \mathrm{P}(X=j|X>0) \\
&=\; \big(\mathrm{E}[X] - \mathrm{E}[X|X>0]\big)^2 + \sum_{j=1}^{\infty} \big(j - \mathrm{E}[X]\big)^2 \mathrm{P}(X=j|X>0) \\
&=\; \big(\mathrm{E}[X] - \mathrm{E}[X|X>0]\big)^2 + \frac{1}{1-\mathrm{P}(X=0)} \sum_{j=1}^{\infty} \big(j - \mathrm{E}[X]\big)^2 \mathrm{P}(X=j) \\
&=\; \big(\mathrm{E}[X] - \mathrm{E}[X|X>0]\big)^2 \\
&\quad + \frac{1}{1-\mathrm{P}(X=0)} \Big\{ \sum_{j=0}^{\infty} \big(j - \mathrm{E}[X]\big)^2 \mathrm{P}(X=j) - \mathrm{E}[X]^2 \mathrm{P}(X=0) \Big\} \\
&=\; \frac{\mathrm{Var}(X) - \mathrm{E}[X]^2 \mathrm{P}(X=0)}{1-\mathrm{P}(X=0)} + \big(\mathrm{E}[X] - \mathrm{E}[X|X>0]\big)^2
\end{aligned}$$

$$= \quad \frac{\mathrm{Var}(X) - \mathrm{E}[X]^2\,\mathrm{P}(X{=}0)}{\mathrm{P}(X{>}0)} + \mathrm{E}[X]^2\Big(\frac{\mathrm{P}(X{=}0)}{\mathrm{P}(X{>}0)}\Big)^2$$

which for the negative binomial case ($\varphi > 0$) yields

$$\mathrm{Var}(X|X{>}0) \;=\; \frac{\delta\lambda}{1-(1+\varphi\delta\lambda)^{-\frac{1}{\varphi}}} + (\delta\lambda)^2\,\frac{2(1+\varphi\delta\lambda)^{-\frac{2}{\varphi}} - (1+\varphi)(1+\varphi\delta\lambda)^{-\frac{1}{\varphi}} + \varphi}{\big(1-(1+\varphi\delta\lambda)^{-\frac{1}{\varphi}}\big)^2} \tag{14}$$

and for the Poisson case ($\varphi = 0$)

$$\mathrm{Var}(X|X{>}0) \;=\; \delta\lambda + \frac{(\delta\lambda)^2}{\big(\exp(\delta\lambda)-1\big)^2} + \frac{\delta\lambda - (\delta\lambda)^2}{\exp(\delta\lambda)-1}. \tag{15}$$

Again assuming a normal approximation we then get an (approximate) joint density of $T$ and $Z$

$$\begin{aligned} p(t,z) &= p(z) \times p(t|z) \\ &= \binom{n}{z}\,\pi_0^z\,(1-\pi_0)^{n-z} \times \frac{1}{\sqrt{2\pi(n{-}z)\sigma^2}}\,\exp\Big(-\tfrac{1}{2}\tfrac{(t-(n-z)\theta)^2}{(n-z)\,\sigma^2}\Big) \end{aligned}$$

where

$$\begin{aligned} \pi_0 &= \begin{cases} (1+\varphi\delta\lambda)^{-\frac{1}{\varphi}} & \text{for } \varphi > 0 \\ \exp(-\delta\lambda) & \text{for } \varphi = 0 \end{cases} \\ \theta &= \tfrac{\delta\lambda}{1-\pi_0} \\ \sigma^2 &= \begin{cases} \theta + (\delta\lambda)^2\,\frac{2\pi_0^2-(1+\varphi)\pi_0+\varphi}{(1-\pi_0)^2} & \text{for } \varphi > 0 \\ \delta\lambda + \frac{(\delta\lambda)^2}{(\frac{1}{\pi_0}-1)^2} + \frac{\delta\lambda-(d\lambda)^2}{\frac{1}{\pi_0}-1} & \text{for } \varphi = 0. \end{cases} \end{aligned}$$

# References


S. D. Aaron, K. L. Vandemheen, D. Fergusson, F. Maltais, J. Bourbeau, R. Goldstein, M. Balter, D. O'Donnell, A. McIvor, S. Sharma, G. Bishop, J. Anthony, R. Cowie, S. Field, A. Hirsch, P. Hernandez, R. Rivington, J. Road, V. Hoffstein, R. Hodder, D. Maciniuk, D. McCormack, G. Fox, G. Cox, H. B. Prins, G. Forg, D. Bleskie, S. Doucette, I. Mayers, K. Chapman, N. Zamel, and M. FitzGerald. Counting, analysing and reporting exacerbations of COPD in randomized controlled trials. *Thorax*, 63(2):122–128, 2008.

A. Anzueto, G. T. Ferguson, G. Feldman, K. Chinsky, A. Seibert, A. Emmett, K. Knobil, D. O'Dell, C. Kalberg, and G. Crater. Effect of fluticasone propionate/salmeterol (250/50) on COPD exacerbations and impact on patient outcomes. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 6(5): 320–329, October 2009.

E. D. Bateman, D. Tashkin, N. Siafakas, L. Dahl, R. aand Towse, D. Massey, D. Pavia, and N. S. Zhong. A one-year trial of tiotropium Respimat plus usual therapy on COPD patients. *Respiratory Medicine*, 104(10):1460–1472, October 2010.

S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7 (4):434–455, December 1998.

P. M. A. Calverley, J. A. Anderson, B. Celli, G. T. Ferguson, C. Jenkins, P. W. Jones, J. C. Yates, and J. Vestbo. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *The New England Journal of Medicine*, 356(8):775–789, February 2007.

P. M. A. Calverley, K. F. Rabe, U.-M. Goehring, S. Kristiansen, L. M. Fabbri, and F. J. Martinez. Roflumilast in symptomatic chronic obstructive pulmonary disease: two randomized clinical trials. *The Lancet*, 374(9691): 685–694, August 2009.

M. Cazzola, W. MacNee, F. J. Martinez, K. F. Rabe, L. G. Franciosi, P. J. Barnes, V. Brusasco, P. S. Burge, P. M. A. Calverley, B. R. Celli, P. W. Jones, D. A. Mahler, B. Make, M. Miravitlles, C. P. Page, P. Palange, D. Parr, M. Pistolesi, S. I. Rennard, M. P. Rutten-van Mölken, R. Stockley, S. D. Sullivan, J. A. Wedzicha, and E. F. Wouters. Outcomes for COPD pharmacological trials: from lung function to biomarkers. *European Respiratory Journal*, 31(2):416–469, February 2008.

J. Copas, K. Dwan, J. Kirkham, and P. Williamson. A model-based correction for outcome reporting bias in meta-analysis. *Biostatistics*, 15(2):370–383, April 2014.

J. B. Copas. A likelihood-based sensitivity analysis for publication bias in meta-analysis. *Journal of the Royal Statistical Society C*, 62(1):47–66, January 2013.

M. Decramer, W. Janssens, and M. Miravitlles. Chronic obstructive pulmonary disease. *The Lancet*, 379(9823):1341–1351, April 2012.

G. C. Donaldson, T. A. Seemungal, J. R. Hurst, and J. A. Wedzicha. Detrended fluctuation analysis of peak expiratory flow and exacerbation frequency in COPD. *European Respiratory Journal*, 40(5):1123–1129, November 2012.

Y.-H. Dong, Lin H.-H., W.-Y. Shau, Y.-C. Wu, C.-H. Chang, and M.-S. Lai. Comparative safety of inhaled medications in patients with chronic obstructive pulmonary disease: systematic review and mixed treatment comparison meta-analysis of randomised controlled trials. *Thorax*, 68(1):48–56, January 2013.

S. Duval and R. Tweedie. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56 (2):455–463, June 2000.

European Medicines Agency (EMA). Guideline on clinical investigation of medicinal products in the treatment of chronic obstructive pulmonary disease (COPD), June 2012. EMA/CHMP/483572/2012.

D. Fleer, M. F. Kerekes, C. Kiefer, K. Quitmann, S. Sturtz, V. Vervölgyi, S. Waffenschmidt, and B. Wieseler. Tiotropiumbromid bei COPD. IQWiG-Berichte Nr. 137, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln, June 2012.

A. Gelman, J. B. Carlin, H. Stern, and D. B. Rubin. *Bayesian data analysis*. Chapman & Hall / CRC, Boca Raton, 1997.

J. P. T. Higgins and S. Green. *Cochrane handbook for systematic reviews and interventions*. The Cochrane Collaboration, March 2011. Version 5.1.0.

J. Hilbe. *Negative Binomial Regression*. Cambridge University Press, 2nd edition, 2011.

JAGS. JAGS version 3.4.0, September 2013. URL `http://mcmc-jags.sourceforge.net/`.

E. T. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, SEC-4(3):227–241, September 1968.

H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London, Series A*, 186(1007):453–461, September 1946.

C. Karner, J. Chong, and P. Poole. Tiotropium versus placebo for chronic obstructive pulmonary disease. *Cochrane Database of Systematic Reviews*, 7: CD009285, 2012.

O. N. Keene, M. R. K. Jones, P. W. Lane, and J. Anderson. Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharmaceutical statistics*, 6(2):89–97, April 2007.

O. N. Keene, P. M. A. Calverley, P. W. Jones, J. Vestbo, and J. A. Anderson. Statistical analysis of exacerbation rates in COPD: TRISTAN and ISOLDE revisited. *European Respiratory Journal*, 32(1):17–24, July 2008.

P. C. Lambert, A. J. Sutton, P. R. Burton, K. R. Abrams, and D. R. Jones. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, 24(15): 2401–2428, July 2005.

J. F. Lawless. Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, 15(3):209–225, September 1987.

E. J. Mills, E. Druyts, I. Ghement, and M. A. Puhan. Pharmacotherapies for chronic obstructive pulmonary disease: a multiple treatment comparison meta-analysis. *Clinical Epidemiology*, 3(1):107–129, March 2011.

M. Plummer. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, and A. Zeileis, editors, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. Vienna, Austria, March 2003.

Martyn Plummer. *rjags: Bayesian graphical models using MCMC*, 2013. URL `http://CRAN.R-project.org/package=rjags`. R package version 3-10.

D. J. Powrie, T. M. A. Wilkinson, G. C. Donaldson, P. Jones, K. Scrine, K. Viel, S. Kesten, and J. A. Wedzicha. Effect of tiotropium on sputum and serum inflammatory markers and exacerbations in COPD. *European Respiratory Journal*, 30(3):472–478, September 2007.

M. A Puhan, L. M. Bachmann, J. Kleijnen, G. ter Riet, and A. G. Kessels. Inhaled drugs to reduce exacerbations in patients with chronic obstructive pulmonary disease: a network meta-analysis. *BMC Medicine*, 7(2), January 2009.

S. Sethi, P. W. Jones, M. S. Theron, M. Miravitlles, E. Rubinstein, J. A. Wedzicha, and R. Wilson. Pulsed moxifloxacin for the prevention of exacerbations of chronic obstructive pulmonary disease: a randomized controlled trial. *Respiratory Research*, 11:10, January 2010.

P. Siri, E. Henninger, and M. P. Sormani. A parametric model fitting time to first event for overdispersed data: application to time to relapse in multiple sclerosis. *Lifetime Data Analysis*, 18(2):139–156, 2012.

D. J. Spiegelhalter, K. R. Abrams, and J. P. Myles, editors. *Bayesian approaches to clinical trials and health-care evaluation*. Wiley & Sons, 2004.

S. Suissa. Statistical treatment of exacerbations in therapeutic trials of chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 173(8):842–846, April 2006.

A. J. Sutton and K. R. Abrams. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, 10(4):277–303, August 2001.

D. P. Tashkin, B. Celli, S. Senn, D. Burkhart, S. Kesten, S. Menjoge, and M. Decramer. A 4-year trial of tiotropium in chronic obstructive pulmonary disease. *The New England Journal of Medicine*, 359(15):1543–1554, October 2008.

W. Viechtbauer. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), August 2010.

Table 2: The example data underlying the analyses. Each study comprises a placebo arm (P) and one or two LAMA treatment arms (L). Some studies report exacerbation rates along with standard errors (group A). Other studies in addition report both the number of exacerbation-free patients (zero-counts) and the total exacerbation count (group B), and yet a larger set reports at least one of the two (group C). When rates are quoted, the total may be derived from the study duration and number of included patients.

| first author (year) | group | arm | patients | duration (yr) | rate | std. err. | total | zeroes |
|---|---|---|---|---|---|---|---|---|
| Ambrosino (2008) | B, C | P | 106 | 0.4808 | | | 26 | 85 |
| | | L | 103 | 0.4808 | | | 19 | 89 |
| Bateman (2010a) | B, C | P | 653 | 1.0000 | 1.91 | | 1247 | 365 |
| | | L | 670 | 1.0000 | 0.93 | | 623 | 421 |
| | | L | 667 | 1.0000 | 1.02 | | 680 | 421 |
| Bateman (2010b) | A, B, C | P | 1953 | 0.9231 | 0.87 | 0.051 | 1568 | 1112 |
| | | L | 1939 | 0.9231 | 0.69 | 0.0561 | 1235 | 1256 |
| Beeh (2006) | C | P | 403 | 0.2308 | | | | 323 |
| | | L | 1236 | 0.2308 | | | | 1056 |
| Brusasco (2003) | C | P | 400 | 0.5000 | 1.49 | | 298 | |
| | | L | 402 | 0.5000 | 1.07 | | 215 | |
| Casaburi (2002) | B, C | P | 371 | 1.0000 | 0.95 | | 352 | 215 |
| | | L | 550 | 1.0000 | 0.76 | | 418 | 352 |
| Chan (2007) | B, C | P | 305 | 0.9230 | 0.92 | | 259 | 180 |
| | | L | 608 | 0.9230 | 0.88 | | 494 | 340 |
| Cooper (2010) | C | P | 225 | 1.8460 | 0.54 | | 194 | |
| | | L | 239 | 1.8460 | 0.51 | | 208 | |
| Donaldson (2012) | A, B, C | P | 52 | 1.0000 | 2.129 | 0.3206 | 111 | |
| | | L | 48 | 1.0000 | 0.9855 | 0.2283 | 47 | |
| Donohue (2002) | C | P | 201 | 0.5000 | | | | 109 |
| | | L | 209 | 0.5000 | | | | 132 |
| Donohue (2010) | C | P | 418 | 0.5000 | 0.72 | | 150 | |
| | | L | 415 | 0.5000 | 0.53 | | 110 | |
| D'Urzo (2011) | C | P | 260 | 0.5000 | | | | 197 |
| | | L | 532 | 0.5000 | | | | 439 |
| Dusser (2006) | B, C | P | 510 | 1.0000 | 1.69 | | 862 | 238 |
| | | L | 500 | 1.0000 | 1.1 | | 550 | 238 |
| Freeman (2007) | B, C | P | 162 | 0.2308 | | | 44 | 160 |
| | | L | 182 | 0.2308 | | | 23 | 181 |
| Johansson (2008) | C | P | 117 | 0.2308 | | | | 113 |
| | | L | 107 | 0.2308 | | | | 105 |
| Moita (2008) | C | P | 160 | 0.2500 | | | | 156 |
| | | L | 144 | 0.2500 | | | | 140 |
| Niewoehner (2005) | B, C | P | 915 | 0.5000 | 1.05 | | 480 | 620 |
| | | L | 914 | 0.5000 | 0.85 | | 388 | 659 |
| Powrie (2007) | A, B, C | P | 73 | 1.0000 | 2.46 | 0.4471 | 134 | 26 |
| | | L | 69 | 1.0000 | 1.17 | 0.2709 | 60 | 39 |
| Rio (2007) | C | P | 125 | 0.2308 | | | | 116 |
| | | L | 123 | 0.2308 | | | | 113 |
| Tashkin (2008) | A, B, C | P | 3006 | 4.0000 | 0.85 | 0.02 | 10220 | 955 |
| | | L | 2986 | 4.0000 | 0.73 | 0.02 | 8719 | 986 |
| Tonnel (2008) | B, C | P | 288 | 0.7500 | 1.83 | | 395 | 158 |
| | | L | 266 | 0.7500 | 1.05 | | 209 | 165 |
| Troosters (2011) | C | P | 219 | 0.2308 | | | | 194 |
| | | L | 238 | 0.2308 | | | | 228 |
| Verkindre (2006) | C | P | 46 | 0.2308 | | | | 43 |
| | | L | 45 | 0.2308 | | | | 45 |
| Voshaar (2008) | C | P | 181 | 0.2308 | | | | 160 |
| | | L | 180 | 0.2308 | | | | 155 |
| | | L | 180 | 0.2308 | | | | 164 |