Biometrics 000, 000–000 DOI: 000
000 0000

Predictively Consistent Prior Effective Sample Sizes

Beat Neuenschwander[1,*], Sebastian Weber[1], Heinz Schmidli[1], and Anthony O'Hagan[2]

[1]Novartis Pharma AG, Basel, Switzerland

[2]Sheffield University, UK

email: beat.neuenschwander@novartis.com

Summary: Determining the sample size of an experiment can be challenging, even more so when incorporating external information via a prior distribution. Such information is increasingly used to reduce the size of the control group in randomized clinical trials. Knowing the amount of prior information, expressed as an equivalent prior *effective sample size (ESS)*, clearly facilitates trial designs. Various methods to obtain a prior's *ESS* have been proposed recently. They have been justified by the fact that they give the standard *ESS* for one-parameter exponential families. However, despite being based on similar information-based metrics, they may lead to surprisingly different *ESS* for non-conjugate settings, which complicates many designs with prior information. We show that current methods fail a basic predictive consistency criterion, which requires the expected posterior–predictive *ESS* for a sample of size N to be the sum of the prior *ESS* and N. The *expected local-information-ratio ESS* is introduced and shown to be predictively consistent. It corrects the *ESS* of current methods, as shown for normally distributed data with a heavy-tailed Student-t prior and exponential data with a generalized Gamma prior. Finally, two applications are discussed: the prior *ESS* for the control group derived from historical data, and the posterior *ESS* for hierarchical subgroup analyses.

Key words: Co-data, Fisher information, historical data, meta-analytic-predictive prior distribution, prior predictive distribution

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Sample sizes are an integral part of clinical trial designs and usually follow from error rate (type-I, power) or precision requirements. Such sample size determinations are standard if no trial-external information is formally included in the analysis of the parameter of interest.

If trial-external information contributes to the inference, one would ideally want to quantify it via an equivalent *effective sample size (ESS)*. Yet this can be difficult. For example, if historical control data inform the prior distribution for the response rate of the control group in a randomized trial, the amount of prior information is not simply the number of historical control subjects. It must be less due to between-trial heterogeneity, which is unknown.

In health-care applications, additional data (or co-data, Neuenschwander, Roychoudhury, Schmidli (2016)) are increasingly valued. In addition to the above example, applications include medical device trials (FDA (2010)), non-inferiority trials with historical or even concurrent placebo (FDA (2010)), pediatric trials with adult data (FDA (2004), Goodman and Sladky (2005)), health-technology assessments (Dias, Welton, Sutton, Ades (2011), Spiegelhalter, Abrams, Myles (2004)), pharmacometrics (Demin, Hamren, Luttringer, Pillai et al. (2012), Nedelman, Bretz, Fisch, Georgieva et al. (2010)), and bridging studies.

In Section 2 we will review the standard *ESS* for the one-parameter exponential family, discuss current methods for non-conjugate settings (Malec (2001), Morita, Thall, Müller (2008), Neuenschwander, Capkun-Niggli, Spiegelhalter (2010), Pennello and Thompson (2008)), introduce the *expected local-information ratio* ESS_{ELIR} as an alternative, investigate the different *ESS* for two examples, and show that only ESS_{ELIR} is predictively consistent. In Section 3, prior and posterior ESS_{ELIR} will be discussed in the context of two recent phase II trials.

2. Methodology

In this Section, we aim to quantify the information for the parameter θ of a statistical model $f(Y|\theta)$, expressed as an equivalent *effective sample size (ESS)*. The information about θ is given probabilistically as a prior (or posterior) distribution $p(\theta)$. The discussion will be restricted to one-dimensional parameters.

2.1 *Effective sample sizes under conjugacy*

Prior effective sample sizes are well understood for conjugate one-parameter exponential families, such as: normal data with mean μ (and known variance s^2) and a normal prior with variance s^2/n_0 ($ESS = n_0$); binary data with response probability θ and a Beta(a, b) prior ($ESS = a+b$); and, Poisson data with mean (hazard) θ and a Gamma(a, b) prior ($ESS = b$).

These *ESS* can be motivated in various ways. First, in the updating rule from prior to posterior parameters, the sample size n appears explicitly, suggesting a corresponding prior *ESS*. For example, for Poisson data with a Gamma(a, b) prior, the second parameter of the posterior Gamma distribution is $b + n$, implying b as the prior *ESS*.

Second, the posterior mean is a weighted average of the prior mean and the standard parameter estimate, with weights proportional to the prior *ESS* and the sample size n. Again, for Poisson data, the prior mean and parameter estimate are a/b and $\sum Y_j/n$, and the posterior mean $(a+\sum Y_j)/(b+n)$ is the weighted average of the two, with weights proportional to b and n. Of note, for exponential data with mean μ and an inverse-Gamma(a, b) prior, however, the *ESS* for the weighted-mean approach is $a - 1$ ($a > 1$), slightly different from the above updating-rule *ESS*=a. Moreover, for exponential data with censoring, *ESS* refers to the effective number of events rather than number of observations.

2.2 *Variance- and precision-ratio methods*

A third, information-based justification is less well-known but will serve as a basis for approaches to *ESS* beyond conjugacy. It relates the variance (or precision) of the prior to the variance of an estimator Y_N for θ from a sample of size N. The *ESS* is then the N for which the two variances are the same. Since the variance of Y_N will usually depend on θ, the expected variance under $p(\theta)$ is taken instead, which leads to

$$ESS = \frac{N\, E_\theta\{\mathrm{Var}(Y_N|\theta)\}}{\mathrm{Var}(\theta)} \tag{1}$$

It can be shown that (1) gives the standard *ESS* for the main one-parameter exponential families. In the sequel, we will use a small modification, which will be needed for the *ESS* of Sections 2.3 and 2.4. Letting $i_F(\theta)$ and $i_F(Y_1;\theta)$ be the expected and observed Fisher information for one information unit,

$$i_F(\theta) = E_{Y_1|\theta}\left\{i_F(Y_1;\theta)\right\} = -E_{Y_1|\theta}\left\{\frac{d^2 \log f(Y_1|\theta)}{d\theta^2}\right\}$$

the variance-ratio and precision-ratio *ESS* are defined as

$$ESS_{VR} = \frac{E_\theta\{i_F^{-1}(\theta)\}}{\mathrm{Var}(\theta)}, \quad ESS_{PR} = \frac{\mathrm{Var}^{-1}(\theta)}{E_\theta\{i_F(\theta)\}} \tag{2}$$

ESS_{VR} and ESS_{PR} are equal or close to the standard *ESS* for the main one-parameter exponential families. For example, for Poisson data and a Gamma(a, b) prior, $\mathrm{Var}(\theta) = a/b^2$, $i_F(\theta) = 1/\theta$, $E\{i_F^{-1}(\theta)\} = a/b$, and $ESS_{VR} = b$. On the other hand, $E\{i_F(\theta)\} = b/(a-1)$, and $ESS_{PR} = b(a-1)/a$, which will be close to b except for very small a.

In the sequel, we will use ESS_{VR} and ESS_{PR} in (2) to represent the variance-ratio and precison-ratio methods. However, other variance-ratio methods have been suggested by Malec (2001), Neuenschwander et al. (2010), and Pennello and Thompson (2008). They obtain the *ESS* of a prior by relating its variance to the variance from an analysis for which the *ESS* is known. We do not include these proposals in the following comparisons because they are similar to ESS_{VR} and ESS_{PR}.

2.3 *The Morita-Thall-Müller (MTM) method*

Another, more involved information-based *ESS* has been suggested in the seminal paper by Morita et al. (2008). In addition to the Fisher information, it uses the information of the prior distribution $p(\theta)$

$$i(p(\theta)) = -\frac{d^2 \log p(\theta)}{d\theta^2} \tag{3}$$

and the information of an ϵ-information (large-variance) prior $p_0(\theta)$ with the same mean ($\overline{\theta}$) as $p(\theta)$

$$i(p_0(\theta)) = -\frac{d^2 \log p_0(\theta)}{d\theta^2}$$

The authors then define the *ESS* as the integer m that minimizes

$$|i(p_0(\overline{\theta})) + E_{Y_m}\{i_F(Y_m; \overline{\theta})\} - i(p(\overline{\theta}))| \tag{4}$$

Here, the expectation of Y_m is taken over the prior-predictive distribution under $p(\theta)$. (4) is the distance (evaluated at the prior mean $\overline{\theta}$) between the expected posterior information for a sample of size m based on the same-mean-large-variance prior $p_0(\theta)$ (the first two terms) and the information of the actual prior (third term).

The approach is noteworthy because it appears to be the first formal, metric-based approach to *ESS* that complies with the standard one-parameter exponential family *ESS*. Some points deserve attention:

(i) Evaluating the distance (4) at the mode may appear more natural. However, as the authors point out, only with the mean one obtains the one-parameter exponential family *ESS*.

(ii) The choice of the "same-mean-large-variance prior" $p_0(\theta)$ is not unique. Yet, since the prior $p_0(\theta)$ carries very little information, one would expect the consequences to be minor. For example, for Poisson data with hazard θ, conjugate Gamma(a, b) prior, and $p_0(\theta)$ chosen as log-normal(m_0, s_0^2), the following holds: $\overline{\theta} = a/b = \exp(m_0 + s_0^2/2)$, and for

$m_0 = \log(\overline{\theta}) - s_0^2/2$ and increasing s_0, $i(p_0(\overline{\theta})) \to -1.5/\overline{\theta}^2$. This implies $ESS_{MTM} = (a-1)/\overline{\theta} + 1.5/\overline{\theta} = b(1+0.5/a)$. The increase from the standard $ESS = b$ will be small except for very small a.

(iii) Restricting m to integers seems not important; one may minimize the distance (4) over continuous m and then round to integers. Setting (4) to zero and noting that $E_{Y_m}\{i_F(Y_m;\overline{\theta})\} = m \cdot E_{Y_1}\{i_F(Y_1;\overline{\theta})\}$, it follows that

$$ESS_{MTM} = \frac{i(p(\overline{\theta})) - i(p_0(\overline{\theta}))}{E_{Y_1}\{i_F(Y_1;\overline{\theta})\}} \tag{5}$$

Moreover, since $p_0(\theta)$ is not unique and really only needed to nudge the computation of the expected Bayesian posterior information with an "uninformative prior", a simplified version that ignores this prior could be used. Additionally approximating the expected posterior information by $m \cdot i_F(\theta)$ and using the prior mode $\widetilde{\theta}$ instead of the prior mean $\overline{\theta}$ leads to the *ESS* suggested by Pennello and Thompson (2008)

$$ESS_{MTM.P} = \frac{i(p(\widetilde{\theta}))}{i_F(\widetilde{\theta})} \tag{6}$$

which will usually be easier to compute than (5).

Finally, (5) and (6) appear similar to the precision-ratio *ESS* (2). That these similarities can be illusory will be shown in Section 2.5.

2.4 *The expected local-information-ratio (ELIR) ESS*

We propose yet another information-based *ESS*, which will be shown to be superior to current versions. The *expected local-information-ratio (ELIR)* method also uses the prior and Fisher information. However, instead of locally evaluating the respective information ratio at the mean (or mode), it is defined as the mean of the prior information to Fisher information ratio $r(\theta)$

$$ESS_{ELIR} = E_\theta\{r(\theta)\} = E_\theta\left\{\frac{i(p(\theta))}{i_F(\theta)}\right\} \tag{7}$$

First, and importantly, ESS_{ELIR} gives the well-known effective sample sizes for some standard one-parameter exponential families. In Table 1, the main quantities and ESS_{ELIR} are shown for the mean parameter as well as the natural parameter.

[Table 1 about here.]

For the natural parameter η, ESS_{ELIR} is the standard *ESS* without any boundary restriction on the parameters. Here, the information ratio $r(\eta) = i(p(\eta))/i_F(\eta)$ does not depend on the parameter. For the natural parameter, the sampling and prior distribution can be written as

$$f(y|\eta) = \exp\{y\eta - M(\eta)\}, \quad p(\eta) = \exp\{n_0 m_0 \eta - n_0 M(\eta)\}$$

Since $i_F(\eta) = d^2M(\eta)/d\eta^2$, it follows that $ESS_{ELIR} = n_0$. For example, for binary data with a Beta(a, b) prior for the mean μ, $\eta = \log\{\mu/(1-\mu)\}$, $M(\eta) = \log\{1 + \exp(\eta)\}$, and $n_0 = a+b$. For Poisson data with a Gamma prior for the mean μ, $\eta = \log(\mu)$, $M(\eta) = \exp(\eta)$, and $n_0 = b$.

While the standard effective sample sizes are obtained for the natural parameter, some special cases arise for vague priors of the mean parameter μ (Table 1). For example, for binary data with a Beta(a,b) prior, $ESS = a + b$ is only obtained for $a > 1, b > 1$. If one of the parameters is less than 1, ESS_{ELIR} is not defined because the expectation of the local information ratio $r(\mu) = (a-1)(1-\mu)/\mu + (b-1)\mu/(1-\mu)$ does not exist; for the uniform distribution ($a = b = 1$), $ESS_{ELIR} = 0$; and, finally $a = 1, b > 1$ (or $a > 1, b = 1$) leads somewhat suprisingly to $ESS_{ELIR} = 1$, since (for the former) $ESS_{ELIR} = (b-1)E_\mu\{\mu/(1-\mu)\} = (b-1)a/(b-1) = a = 1$.

2.5 *Examples*

We now discuss two examples with non-conjugate prior distributions and show that the *ESS* for the methods discussed so far can differ considerably.

2.5.1 *Normal data with a Student-t prior.* We first assume normal data (with known variance s^2) and a Student-t prior with df degrees of freedom for the mean parameter θ. Such a heavy-tailed prior is robust in the sense that the prior influence decreases with increasing conflict between the data and the prior (O'Hagan (1979), O'Hagan and Pericchi (2012)). The Fisher and prior information are

$$i_F(\theta) = 1/s^2, \quad i(p(\theta)) = \frac{df+1}{df}\frac{1-\theta^2/df}{(1+\theta^2/df)^2}$$

Noting that the prior information for a t-prior with scale s_0 is $i(p(\theta))/s_0^2$, the variance-ratio and precision-ratio *ESS* are

$$ESS_{VR} = ESS_{PR} = (s/s_0)^2(df-2)/df \qquad (df > 2)$$

On the other hand, using a large-variance (in the limit improper) prior $p_0(\theta)$,

$$ESS_{MTM} = (s/s_0)^2(df+1)/df \qquad (df > 1)$$

and $ESS_{MTM.P} = ESS_{MTM}$. Finally, intergrating $i(p(\theta))$ over the prior distribution gives

$$ESS_{ELIR} = (s/s_0)^2(df+1)/(df+3)$$

Table 2 (upper part) shows that for small df, the interesting case if robustness is the aim, these *ESS* differ considerably. This is problematic when deciding on the size of an experiment that will incorporate prior information in the analysis. Of note, the above formulas show that with increasing degrees of freedom, *ESS* is increasing when using the inverse of the variance (ESS_{VR}, ESS_{PR}) but decreasing when using the curvature at the mean (ESS_{MTM}). For ESS_{ELIR}, which uses the expected curvature, *ESS* is increasing.

[Table 2 about here.]

2.5.2 *Exponential data with a generalized Gamma prior.* The second example assumes exponentially distributed data, for which the prior of the hazard parameter θ is a generalized

Gamma distribution with shape parameter a, scale parameter s, and family parameter f. Its density is

$$p(\theta) = \frac{f\theta^{a-1}\exp\{-(\theta/s)^f)\}}{s^a\Gamma(a/f)}$$

This three-parameter distribution offers more flexibility than the conjugate Gamma distribution and may thus be useful when representing prior information (for example three quartiles elicited from experts). It includes Gamma ($f = 1$) and Weibull ($f = a$) distributions as special cases. The Fisher and prior information are

$$i_F(\theta) = 1/\theta^2, \quad i(p(\theta)) = (a-1)/\theta^2 + f(f-1)\theta^{f-2}/s^f$$

The variance-ratio and precision ratio *ESS* are

$$ESS_{VR} = E(\theta^2)/\mathrm{Var}(\theta), \quad ESS_{PR} = E^{-1}(1/\theta^2)/\mathrm{Var}(\theta)$$

which follow from $E(\theta^r) = s^r\Gamma\{(a+r)/f\}/\Gamma(a/f)$; note that ESS_{PR} only exists for $a > 2$. Further, using a large-variance Gamma prior for $p_0(\theta)$,

$$ESS_{MTM} = a + f(f-1)\left[\Gamma\{(a+1)/f\}\right]^f/\{\Gamma(a/f)\}^f$$

Using the mode $\widetilde{\theta} = s\{(a-1)/f\}^{1/f}$, the simplified $ESS_{MTM.P}$ is

$$ESS_{MTM.P} = af - f$$

Finally, the expected local-information-ratio *ESS* is

$$ESS_{ELIR} = af - 1$$

Table 2 (lower part) shows the effective sample sizes for some parameter constellations (including Gamma, Weibull, and genuine generalized Gamma distributions), which have been grouped by equal ESS_{ELIR} values. The dilemma is the same as in the first example: the *ESS* can differ considerably, in particular with increasing amounts of prior information.

2.6 *The predictive consistency criterion*

So far we have discussed various approaches to *ESS*, which work well for conjugate settings but can differ considerably otherwise. To resolve this dilemma, more than fulfilling the exponential family criterion is needed. We require the *ESS* to meet the additional *predictive consistency criterion*:

> Predictive consistency: for a sample of size N, the expected posterior *ESS* must be the sum of the prior *ESS* and N.

For some of the examples of Section 2.5, Table 3 shows the difference between the expected posterior *ESS* and N, which should be the prior *ESS*. Only ESS_{ELIR} seems to be predictively consistent. It should be noted that the results are not exact: they represent the mean of 10000 simulations, each generating data of size N (10,100,1000) from the prior predictive distribution and then obtaining the respective posterior distribution and its *ESS*. Interestingly, for large N it seems that the difference of the expected posterior *ESS* and N converges to ESS_{ELIR} for all methods.

[Table 3 about here.]

In fact, it can be shown that ESS_{ELIR} is predictively consistent for any planned sample size N. The proof is as follows: let Y_N be the predictive data of size N with posterior distribution $p(\theta|Y_N)$, for which the posterior ESS_{ELIR} is

$$E_{\theta|Y_N}\left\{\frac{i(p(\theta)) - d^2 \log f(Y_N|\theta)/d\theta^2}{i_F(\theta)}\right\}$$

The expected posterior effective sample size under the prior predictive distribution is then

$$\begin{aligned}
&E_{Y_N}\left[E_{\theta|Y_N}\left\{\frac{i(p(\theta)) - d^2 \log f(Y_N|\theta)/d\theta^2}{i_F(\theta)}\right\}\right] \\
&= E_{\theta}\left[E_{Y_N|\theta}\left\{\frac{i(p(\theta)) - d^2 \log f(Y_N|\theta)/d\theta^2}{i_F(\theta)}\right\}\right] \\
&= E_{\theta}\left\{\frac{i(p(\theta)) + N i_F(\theta)}{i_F(\theta)}\right\} = ESS_{ELIR} + N
\end{aligned}$$

2.7 *Computations*

Computing ESS_{ELIR} (7) of a prior analytically was possible in the examples of Section 2.5. For priors derived from historical data, obtaining ESS_{ELIR} analytically will usually not be possible, except for special cases with known (or assumed) variance components in the hierarchical model or a known power parameter for power priors (Pocock (1976), Chen and Ibrahim (2000), Neuenschwander, Branson, and Spiegelhalter (2009)).

If the prior *ESS* cannot be computed analytically, approximations can be used. First, if the prior is parametric, the information $i(p(\theta))$ may be available analytically but the expectation (7) may require numerical integration or simulations from $p(\theta)$ to obtain the empirical mean as an estimate of ESS_{ELIR}.

A second approximation will be needed if $p(\theta)$ is not directly available. For example, $p(\theta)$ may be a large simulation sample (typically from an MCMC analysis); see Section 3 for two such applications. While inconvenient, this does not pose serious problems because $p(\theta)$ can be approximated by a mixture of standard distributions (e.g., normal, Beta, Gamma) to any degree of accuracy (Dallal and Hall (1983), Diaconis and Ylvisaker (1984)), and the respective information $i(p(\theta))$ follows from the second derivatives of the log-mixture distribution (see Appendix).

In this context, it should be noted that the *ESS* of a mixture distribution is not the respective weighted average of the component-wise *ESS*, not even for the conjugate cases of Section 2.1. For example, for normal data with known variance $s^2 = 100$ and a mixture prior for the mean θ, $p(\theta) = 0.5 \times N(-2, 2^2) + 0.5 \times N(2, 2^2)$, the weighted *ESS* is $0.5 \times 100/4 + 0.5 \times 100/4 = 25$. The other methods give $ESS_{VR} = ESS_{PR} = 100/\text{Var}(\theta) = 100/8 = 12.5$, $ESS_{MTM} = 0$ (because the prior curvature at the mean is 0 for this special mixture distribution), whereas the predictively consistent *ESS* is $ESS_{ELIR} = 13.7$.

3. Applications

3.1 *Prior ESS for a proof-of-concept trial using historical control data*

We now discuss a recent randomized proof-of-concept phase II trial in which the prior for the parameter in the control group was informed by historical data. Proof-of-concept trials aim to provide initial evidence of efficacy for a new treatment. They increasingly use Bayesian approaches for design and analysis (Fisch, Jones, Jones, Kerman et al. (2015)). Baeten, Baraliakos, Braun, Sieper et al. (2013) describe such a trial where patients with ankylosing spondylitis, a chronic inflammatory disease, were randomized to the monoclonal antibody *secukinumab* (n=24) or to placebo (n=6). The Bayesian primary analysis leveraged historical placebo data, which allowed the investigators to allocate fewer patients to placebo. This reduced costs and trial duration and also facilitated recruitment.

The primary efficacy endpoint was binary (response at week six). Eight historical randomized placebo-controlled clinical trials provided data on the placebo response rate (Table 4). The authors used the meta-analytic-predictive (MAP) approach (Spiegelhalter et al. (2004), Neuenschwander et al. (2010), Schmidli, Gsteiger, Roychoudhury, O'Hagan et al. (2014)) to quantify the historical placebo information.

The number of responders in the placebo group of the j-th historical trial is $r_j|\pi_j \sim Bin(\pi_j, n_j), \quad j = 1, \ldots, 8$, where π_j is the true placebo response rate and n_j the number of patients in the placebo group. Table 4 shows the fairly heterogeneous historical placebo data, with observed response rates in the range of 12% (trial 7) to 37% (trial 3).

[Table 4 about here.]

Denoting the placebo response rate in the new trial by $\pi_\star$ and using the log-odds transformation, $\theta = \log\{\pi/(1-\pi)\}$, the simplest MAP approach assumes exchangeable parameters, $\theta_\star, \theta_1, \ldots, \theta_8 \mid \mu, \tau \sim N(\mu, \tau^2)$. Here, the between-trial standard deviation τ characterizes

between-trial heterogeneity, that is, the extent to which the trial parameters can deviate from the mean μ.

The Bayesian MAP analysis requires priors for the parameters μ and τ. For μ, a vague prior is typically used. However, more care is needed for τ, in particular for few historical trials. For example, the still popular uniform priors with large upper bound will essentially disregard the historical data because they put too much probability mass on unrealistically large between-trial standard deviations. Here, we use a half-normal prior (Spiegelhalter et al. (2004)), which puts most of its probability mass on realistic between-trial heterogeneities ($\tau < 2$ on the log-odds scale). In the following, we use a $N(0, 10^2)$ prior for μ and a half-normal prior with scale 1 for τ. Alternatively, t-priors for μ and half-Cauchy or half-t priors for τ could be used (Gelman (2006), Polson and Scott (2012)).

Markov chain Monte Carlo (MCMC) have been used to simulate from the posterior distribution $p_{MAP}(\pi_\star) = p(\pi_\star \mid r_1, \ldots, r_8)$, which is summarized in Figure 1 and Table 4. It should be noted that MCMC provides only a simulation sample but no analytic solution for the MAP prior. This complicates both the calculation of the prior *ESS* and the Bayesian analysis at the end of the trial. Both issues can be addressed by approximations to the MAP prior. Baeten et al. (2013) used a single Beta distribution, but mixture distributions will usually provide much better approximations (Schmidli et al. (2014), Weber (2019), Weber et al. (2019)). Here, a mixture of two Beta distributions (Table 4) already provides a very good fit; Figure 1 displays the MAP prior (density plot) and the two-component mixture approximation.

[Figure 1 about here.]

For the design of the new trial, which aims for a smaller placebo group by leveraging the historical information, knowing the *ESS* is important. For the two- and three-component mixture approximation, which give a very similar fit, the ESS_{ELIR} are 36 and 38, consid-

erably larger than the *ESS=25* from a single Beta approximation. Of note, the predictively inconsistent ESS_{VR} is 26 for all approximations, whereas the ESS_{MTM} for the single Beta and the two mixture approximations are 26, 57, and 91, respectively.

3.2 *Posterior ESS for hierarchical subgroup analyses*

The aim of this section is to use effective sample sizes to quantify the gain of information for hierarchical subgroup analyses. Hierarchical models enable sharing information across similar but non-overlapping subgroups, which can be particularly helpful for small subgroups.

We use the phase II trial by Chugh, Wathen, Maki, Benjamin et al. (2009) who assessed the effect of *imatinib* in ten histological subtypes of sarcoma. The data are shown in Table 5: 179 patients were available for analysis, with sample sizes ranging from 2 to 29 for the ten subtypes. Observed response rates for clinical benefit response (CBR) varied from 0% for subtypes 2 and 9 to 24% for subtype 5.

The trial design (Thall, Wathen, Bekele, Champlin et al. (2003)) was based on a standard hierarchical model, which exploits the anticipated similarity of responses for the 10 subtypes. Robust extensions of this standard model have been discussed by Leon-Nevelo, Bekele, Müller, Quintana et al. (2012) and Neuenschwander, Wandel, Roychoudhury, Bailey (2015). In the sequel, ESS_{ELIR} for each subgroup will be given for the full exchangeability model and three robust extensions.

For binomial data, $r_j|\pi_j \sim Bin(\pi_j, n_j)$, a convenient hierarchical model assumes a normal random-effects distribution for the subtype-specific log-odds parameters $\theta_j = \log(\pi_j/(1-\pi_j)), j = 1 \ldots 10$, i.e., $\theta_j|\mu, \tau \sim N(\mu, \tau^2)$. The following prior distributions will be used: a vague normal distribution with mean 0 and standard deviation 2 for μ, and a half-normal distribution with scale parameter 1 for τ; the 95%-interval of the latter is (0.03,2.24), which covers very small to very large between-subtype heterogeneity on the log-odds scale. In addition, we use three robust mixture models, assuming the first component as $\theta_j|\mu, \tau \sim$

$N(\mu, \tau^2)$ (with priors as above) and the second component as $\theta_j \sim N(0, 2^2)$ for all subgroups. For each subgroup, mixture weights will be 0.9/0.1, 0.75/0.25, or 0.5/0.5. The four models will be denoted by (HM-100) for the full exchangeability model and HM-90, HM-75, HM-50 for the three mixture models.

Here, we are interested in the *ESS* relative to parallel binomial experiments for each subtype, for which the Fisher information is $i_F(\theta_j) = \exp(\theta_j)/\{1 + \exp(\theta_j)\}^2$. The MCMC posterior distributions $p(\theta_j|r_1, \ldots, r_{10})$ have been approximated by a mixture of four Beta distributions.

For the four models, Table 5 shows ESS_{ELIR} for the response rate in each subgroup. Of course, the information gain relative to the subgroup sample sizes n_j is the largest for the full exchangeability model (HM-100), with *ESS* between 54 and 78. On the other hand, the *ESS* for the robust mixture extensions can be considerably smaller (in particular for model HM-50) but are still much larger than the subgroup sample sizes n_j.

The results show that even under robust borrowing hierarchical model analyses for subgroups can lead to substantive information gains compared to stratified analyses. Finally, it should be noted that for the full exchangeability model, the posterior mixture weights increase for all subtypes, which justifies the full exchangeability design used in the actual trial (Thall et al. (2003)).

[Table 5 about here.]

4. Discussion

Modern drug and health-care development tend towards better use of the evidence, which involves using multiple data sources via meta-analytic methods (21st Century Cures Act (2015), European Commission (2014), European Medicines Agency (2013, 2018), FDA (2004, 2013, 2018)).

In this regard, the effective sample size *(ESS)* of trial-external information, which contributes to the inference in the actual trial, is an important metric. Various methods to obtain the ESS have been discussed recently. They are similar in that they relate the available information (formally the prior or posterior precision) to the Fisher information. Yet the methods can give surprisingly different *ESS*. We have shown that the expected local-information-ratio ESS_{ELIR} addresses the limitations of current methods. Importantly, it is predictively consistent and thus correctly quantifies the amount of information as an equivalent number of observations.

Our focus has been on one-dimensional parameters. Clearly, many applied problems involve more than one parameter, for which effective sample sizes of individual parameters (or even parameter vectors) are of interest; for ESS_{MTM} in such settings, see Morita, Thall, Müller (2012) and Thall, Herrick, Nguyen, Venier et al. (2014). Finding predictively consistent *ESS* for such cases requires further research.

Acknowledgements

We would like to thank the two reviewers and the associate editor for their excellent comments, which greatly helped us to improve the manuscript.

References

21st Century Cures Act. May 19, 2015. http://docs.house.gov/meetings/IF/IF00/20150519/103516/BILLS-1146ih.pdf.

Baeten D, Baraliakos X, Braun J, Sieper J, Emery P, van der Heijde D, McInnes I, van Laar J, Landewe R, Wordsworth P, Wollenhaupt J, Kellner H, Paramarta J, Wei J, Brachat A, Bek S, Laurent D, Li Y, Wang Y, Bertolino A, Gsteiger S, Wright AM and Hueber W. Anti-interleukin-17A monoclonal antibody secukinumab in ankylosing

spondylitis: a randomized, double-blind, placebo-controlled trial. *The Lancet* 2013; 382: 1705–1713.

Chen MH and Ibrahim JG. Power prior distributions for regression models. *Statistical Science*, 15(1):46–60, 2000.

Chugh R, Wathen K, Maki R, Benjamin R, Patel R, Myers P, Priebat D, Reinke D, Thomas D, Keohan M, Samuels B, and Baker L. Phase II multicenter trial of imatinib in 10 histologic subtypes of sarcoma using a Bayesian hierarchical statistical model. *Journal of Clinical Oncology* 2009; 27(19): 3148–3153.

Dallal S, Hall W. Approximating priors by mixtures of natural conjugate priors. *Journal of the Royal Statistical Society Serires B* 1983; 45: 278-286.

Demin I, Hamren B, Luttringer O, Pillai G, and Jung T. Longitudinal model-based meta-analysis in rheumatoid arthritis: an application toward model-based drug development. *Clinical Pharmacology and Therapeutics* 2012; 92(3): 251–261.

P. Diaconis and D. Ylvisaker. Quantifying prior opinion. *Bayesian Statistics (Proceedings of the Second Valencia International Meeting)* 1984; 2:133–148.

Dias S, Welton N, Sutton A, and Ades A. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials, 2011.

European Comission. Innovative Medicines initiative 2: Europe's fast track to better medicines 2014.

European Medicines Agency. Reflection paper on the use of extrapolation in the development of medicines for paediatrics 2018. https://www.ema.europa.eu/documents/scientific-guideline/adopted-reflection-paper-use-extrapolation-development-medicines-paediatrics-revision-1_en.pdf [Online; accessed 20-December 2018].

European Medicines Agency. Concept paper on extrapolation of efficacy and safety

in medicine development 2013. https://www.ema.europa.eu/documents/scientific-guideline/concept-paper-extrapolation-efficacy-safety-medicine-development_en.pdf [Online; accessed 20-December 2018].

Fisch R, Jones I, Jones J, Kerman J, Rosenkranz GK and Schmidli H. Bayesian design of proof-of-concept trials. *Therapeutic Innovation & Regulatory Science* 2015; 49(1): 155-162.

Food and Drug Administration (FDA). Innovation/Stagnation - Challenge and Opportunity on the Critical Path to New Medical Products 2004. [Online; accessed 6-September-2015].

Food and Drug Administration (FDA). Guidance for Industry: Non-Inferiority Clinical Trials 2016. https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf [Online; accessed 20-December-2018].

Food and Drug Administration (FDA). Guidance for the use of Bayesian statistics in medical device clinical trials 2010. www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/ GuidanceDocuments/ ucm071072.htm (accessed Jan 2014).

Food and Drug Administration (FDA). Paving the Way for Personalized Medicine: FDA's Role in a New Era of Medical Product Development 2013.

Food and Drug Administration (FDA). Guidance for Industry and Food and Drug Administration Staff: Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices 2016. https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments [Online; accessed 20-December-2018].

Food and Drug Administration (FDA). Framework for FDA's Real-World Evidence Program 2018.

Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian*

Analysis 2006; 1(3): 515–534.

Goodman S, Sladky J. A Bayesian approach to randomized controlled trials in children utilizing information from adults: the case of Guillain-Barre syndrome. *Clinical Trials* 2005; 2(10): 305–310.

Leon-Novelo LG, Bekele BN, Müller P, Quintana F, Wathen K. Borrowing strength with nonexchageable priors over subpopulations. *Biometrics* 2012; 68:550-558.

Malec D. A closer look at combining data among a small number of binomial experiments. *Statistics in Medicine* 2001; 12: 1811–1824.

Morita S, Thall PF and Müller P. Determining the effective sample size of a parametric prior. *Biometrics* 2008; 64: 595–602.

Morita S, Thall PF, and Müller P. Prior effective sample size in conditionally independent hierarchical models. *Bayesian Analysis* 2012; 7(3):561–614.

Nedelman J, Bretz F, Fisch R, Georgieva A, Hsu C, Kahn J, Kawai L, Lowe P, Maca J, Pinheiro J, Rossini A, Schmidli H, Steimer J, and Yu J. Modeling and simulation in clinical drug development. In: *Pharmaceutical Sciences Encyclopedia* 2010; 43: 1–29.

Neuenschwander B, Branson M, and Spiegelhalter DJ. A note on the power prior. *Statistics in Medicine*, 28:3562–3566, 2009.

Neuenschwander B, Capkun-Niggli G, Branson M, and Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. *Clinical Trials* 2010; 7: 5–18.

Neuenschwander B, Wandel S, Roychoudhury S, and Bailey S. Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical Statistics* 2015; 15: 123-134.

Neuenschwander B, Roychoudhury S, and Schmidli H. On the use of co-data in clinical trials. *Statistics in Biopharmaceutical Research* 2016; 8(3): 1-10.

O'Hagan, A. On outlier rejection phenomena in Bayes inference. *Journal of the Royal*

Statistical Society, Series B 1979; **41,** 358–367.

O'Hagan A and Pericchi L. Bayesian heavy-tailed models and conflict resolution: a review. *Brazilian Journal of Probability and Statistics* 2012; 26: 372–401.

Pennello G and Thompson L. Experience with reviewing Bayesian medical device trials. *Journal of Biopharmaceutical Statistics* 2008; 18(1): 81–115.

Pocock, SJ. The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, 29(3):175–188, 1976.

Polson NG and Scott JG. On the Half-Cauchy prior for a global scale parameter. *Bayesian Analysis* 2012; 7: 887–902.

SAS Institute. *SAS user guide: Statistics. The FMM procedure.* Cary, NC: SAS Institute Inc., 2014.

Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan O, Spiegelhalter DJ and Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* 2014; 70: 1023–32.

Spiegelhalter DJ, Abrams KR and Myles JP. *Bayesian Approaches to Clinical trials and Health-Care Evaluation.* 2004; Chichester: Wiley.

Thall PF, Herrick RC, Nguyen HQ, Venier JJ, and Norris JC. Effective sample size for computing prior hyperparameters in Bayesian phase I-II dose-finding. *Clinical Trials* 2014; 11(6):657–666.

Thall PF, Wathen J, Bekele B, Champlin R, Baker L, and Benjamin R. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Journal of Clinical Oncology* 2003; 22: 763–780.

S. Weber. *RBesT: R Bayesian Evidence Synthesis Tools*, 2019. R package version 1.4-0.

Weber S, Li Y, Seaman J III, Kakizume T, and Schmidli H. Applying meta-analytic predictive priors with the R Bayesian evidence synthesis tools. *pre-print* 2019. arXiv:1907.00603

Supporting Information

The on-line material is available at Biometrics website on Wiley Online Library. It contains ESS_{ELIR} functions for mixtures of normal and Beta distributions and code to reproduce the results of Table 3 and the two applications. Note that the R-package *RBesT* (Weber (2019), Weber et al. (2019)) is required.

Appendix

Information $i(p(\theta))$ of a mixture distribution

If the prior (or posterior distribution) is a mixture distribution with K components, $p(\theta) = \sum_{j=1}^{K} w_j p_j(\theta)$, its information is

$$\begin{aligned} i(p(\theta)) &= -d_\theta^2 \log p(\theta) \\ &= \frac{1}{p^2(\theta)} \left[\sum_{j=1}^{K} w_j p_j(\theta) d_\theta \log p_j(\theta) \right]^2 \\ &\quad -\frac{1}{p(\theta)} \sum_{j=1}^{K} w_j p_j(\theta) \left[\{ d_\theta \log p_j(\theta) \}^2 + d_\theta^2 \log p_j(\theta) \right] \end{aligned}$$

Here, d_θ and d_θ^2 denote the first and second derivative, respectively.

Fitting mixture distributions

Various procedures are available for fitting mixture distributions, for example the *mixfit* and *automixfit* functions in the R-package *RBesT* (Weber (2019), Weber et al. (2019)), or *SAS PROC FMM* (2014). The former has been used to fit the prior and posterior distributions of the applications in Section 3.

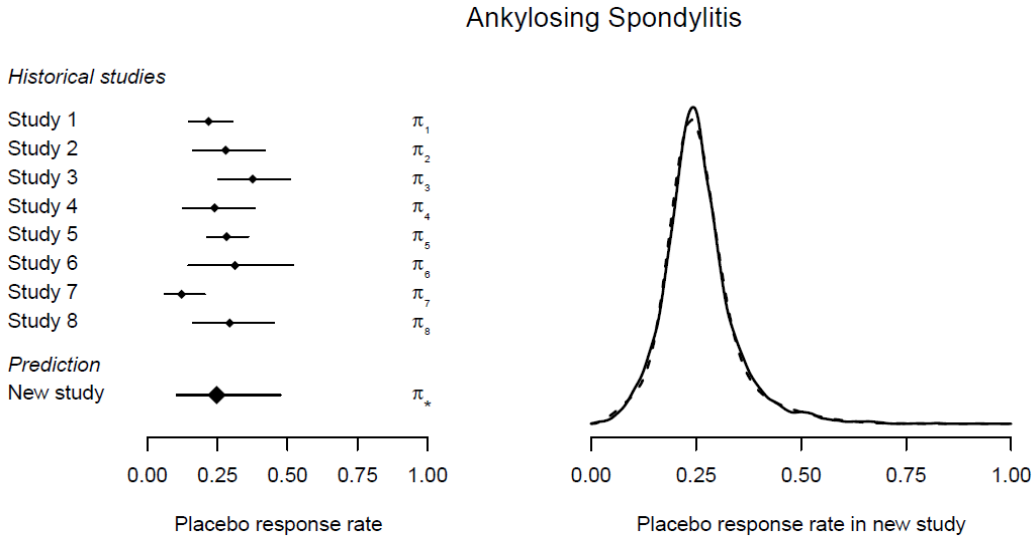

Figure 1. Median and 95%-intervals for event rates of historical ankylosing spondylitis trials and MAP event rate for new trial (left panel), and MAP prior density (solid line) with two-component Beta mixture approximation (dashed line) (right panel).

Table 1

Prior information $i(p(\mu))$ and $i(p(\eta))$, Fisher unit information $i_F(\mu)$ and $i_F(\eta)$, local-information ratio $r(\mu)$ and $r(\eta)$, and expected local-information-ratio ESS_{ELIR} for some one-parameter exponential families: μ and η are the mean and natural parameter, respectively.

parameter	$i(p(\mu)), i(p(\eta))$	$i_F(\mu), i_F(\eta)$	$r(\mu), r(\eta)$	ESS_{ELIR}
	Normal (σ known): $Y \sim N(\mu, \sigma^2)$, $\mu \sim N(m_0, s_0^2)$			
μ	s_0^{-2}	σ^{-2}	σ^2/s_0^2	σ^2/s_0^2
	Binomial: $Y \sim \mathsf{Bin}(\mu, 1)$, $\mu \sim \mathsf{Beta}(a, b)$, $\eta = \mathsf{logit}(\mu)$			
μ	$(a-1)/\mu^2 + (b-1)/(1-\mu)^2$	$1/\mu + 1/(1-\mu)$	$(a-1)(1-\mu)/\mu + (b-1)\mu/(1-\mu)$	$a + b\,(a, b > 1)$
η	$(a+b)\exp(\eta)/(1+\exp(\eta))^2$	$\exp(\eta)/(1+\exp(\eta))^2$	$a+b$	$a+b$
	Poisson: $Y \sim \mathsf{Pois}(\mu)$, $\mu \sim \mathsf{Ga}(a, b)$, $\eta = \log(\mu)$			
μ	$(a-1)/\mu^2$	$1/\mu$	$(a-1)/\mu$	$b\,(a > 1)$
η	$b\exp(\eta)$	$\exp(\eta)$	b	b
	Exponential: $Y \sim \mathsf{Exp}(1/\mu)$, $\mu \sim \mathsf{Inv\text{-}Ga}(a, b)$, $\eta = 1/\mu$			
μ	$-(a+1)/\mu^2 + 2b/\mu^3$	$1/\mu^2$	$-(a+1) + 2b/\mu$	$a-1$
η	$(a-1)/\eta^2$	$1/\eta^2$	$a-1$	$a-1$
	Chi-square: $s_d^2 \sim \mathsf{Ga}(d/2, d/(2\sigma^2))$ $\sigma^2 \sim \mathsf{Inv\text{-}Ga}(a, b)$, $\eta = 1/\sigma^2$			
σ^2	$-(a+1)/\sigma^4 + 2b/\sigma^6$	$d/(2\sigma^4)$	$-2(a+1)/d + 4b/(2\sigma^2)$	$2(a-1)/d$
η	$(a-1)/\eta^2$	$d/(2\eta^2)$	$2(a-1)/d$	$2(a-1)/d$

Table 2

Prior ESS for various methods: for normal data (known $s^2 = 100$) with Student-t(df) prior, and for exponential data with generalized Gamma(a,s=1,f) prior.

ESS for normal data with Student-t prior

df	VR	PR	MTM	MTM.P	ELIR
2	—	—	150	150	60
3	33	33	133	133	67
4	50	50	125	125	71
5	60	60	120	120	75
10	80	80	110	110	85
50	96	96	102	102	96

ESS for exponential data with generalized-Gamma prior (a,s=1,f)

distribution	a	f	VR	PR	MTM	MTM.P	ELIR
Gamma	9.00	1.00	10.0	6.2	9.0	8.0	8.0
Weibull	3.00	3.00	8.6	3.5	7.3	6.0	8.0
gen-Gamma	2.54	3.54	7.9	2.3	6.4	5.4	8.0
Gamma	25.00	1.00	26	22	25	24	24
Weibull	5.00	5.00	20	15	18	20	24
gen-Gamma	4.52	5.52	19	14	16	19	24
Gamma	49.00	1.00	50	46	49	48	48
Weibull	7.00	7.00	36	32	33	42	48
gen-Gamma	6.52	7.52	35	30	31	41	48
Gamma	81.00	1.00	82	78	81	80	80
Weibull	9.00	9.00	58	53	53	72	80
gen-Gamma	8.51	9.51	55	51	50	71	80
Gamma	121.00	1.00	122	118	121	120	120
Weibull	11.00	11.00	84	79	77	110	120
gen-Gamma	10.51	11.51	81	76	74	109	120
Gamma	169.00	1.00	170	166	169	168	168
Weibull	13.00	13.00	115	110	106	156	168
gen-Gamma	12.51	13.51	111	107	102	155	168

Table 3

Prior ESS and expected posterior ESS − N for planned sample sizes N = 10, 100, and 1000: for normal data (variance=100) with Student-t prior and exponential data with Weibull prior (see Section 2.5)

$Y|\theta \sim N(\theta, 10^2), \quad \theta \sim$ Student-t(df)

		prior ESS	(expected posterior ESS)–N		
	method		N=10	N=100	N=1000
df=2	VR	—	36	54	59
	MTM	150	88	78	62
	MTM.P	150	137	85	62
	ELIR	**60**	**60**	**60**	**60**
df=3	VR	33	48	62	67
	MTM	133	110	83	70
	MTM.P	133	125	87	70
	ELIR	**67**	**67**	**67**	**68**
df=4	VR	50	57	68	70
	MTM	125	112	88	73
	MTM.P	125	119	90	73
	ELIR	**71**	**72**	**72**	**71**
df=5	VR	60	63	72	75
	MTM	120	112	89	77
	MTM.P	120	115	91	77
	ELIR	**75**	**75**	**75**	**75**
df=10	VR	80	80	83	85
	MTM	110	107	94	86
	MTM.P	110	107	95	86
	ELIR	**85**	**85**	**85**	**85**
df=50	VR	96	96	96	96
	MTM	102	101	99	97
	MTM.P	102	99	99	97
	ELIR	**96**	**96**	**96**	**96**

$Y|\theta \sim$ Exp$(\theta), \quad \theta \sim$ Weibull(a, s)

		prior ESS	(expected posterior ESS)–N		
	method		N=10	N=100	N=1000
a=3	VR	8.6	9.6	10	10
	PR	3.6	5.5	6.0	6.2
	MTM	7.3	8.2	8.8	9.9
	MTM.P	6	6.8	7.7	7.9
	ELIR	**8**	**8.0**	**7.9**	**8.0**
a=5	VR	20	23	26	26
	PR	15	19	22	22
	MTM	18	20	24	24
	MTM.P	20	20	22	23
	ELIR	**24**	**24**	**24**	**24**
a=7	VR	36	41	49	50
	PR	32	37	45	46
	MTM	33	37	44	48
	MTM.P	42	41	44	47
	ELIR	**48**	**48**	**48**	**48**
a=9	VR	58	64	77	84
	PR	53	59	73	80
	MTM	53	57	69	79
	MTM.P	72	69	71	78
	ELIR	**80**	**80**	**80**	**81**
a=11	VR	84	91	111	123
	PR	79	86	107	119
	MTM	77	82	99	117
	MTM.P	110	107	105	116
	ELIR	**120**	**120**	**120**	**121**
a=13	VR	115	123	148	168
	PR	110	118	144	164
	MTM	106	111	133	158
	MTM.P	156	152	146	158
	ELIR	**168**	**168**	**167**	**166**

Table 4

Historical ankylosing spondylitis data, summaries and mixture approximations of meta-analytic-predictive(MAP) prior, and ESS_{ELIR}

Historical data

23/107 (21%), 12/44 (27%), 19/51 (37%), 9/39 (23%), 39/139 (28%), 6/20 (30%), 9/78 (12%), 10/35 (29%)

MAP prior

mean	sd	2.5%-97.5%
0.26	0.084	0.11-0.46

Approximations (single Beta and mixtures)

	w_1 (a_1, b_1)	w_2 (a_2, b_2)	w_3 (a_3, b_3)	mean	sd	2.5%-97.5%	ESS_{ELIR}
Beta	1.00 (6.8,19.7)			0.26	0.083	0.11-0.43	26
2-comp Beta	0.66 (16.7,51.1)	0.34 (3.4,9.0)		0.26	0.084	0.11-0.47	36
3-comp Beta	0.62 (6.0,17.7)	0.34 (36.0,110)	0.04 (2.5,4.1)	0.26	0.085	0.11-0.45	38

Table 5

Sarcoma subtype data (number of responders/patients) and ESS_{ELIR} for each subtype from hierarchical model analyses assuming full exchangeability (HM-100) or robust mixtures with 90, 75, or 50% weight for exchangeability (HM-90, HM-75, HM-50).

			HM-100	HM-90	HM-75	HM-50
Subtype	r/n	(%)		ESS		
1. Angiosarcoma	2/15	(13)	65	60	50	36
2. Ewing	0/13	(0)	57	46	36	24
3. Fibrosarcoma	1/12	(8)	60	54	44	31
4. Leiomyosarcoma	6/28	(21)	78	72	64	47
5. Liposarcoma	7/29	(24)	76	68	59	43
6. MFH	3/29	(10)	74	69	59	46
7. Osteosarcoma	5/26	(19)	79	73	64	47
8. MPNST	1/5	(20)	57	48	39	23
9. Rhabdomysarcoma	0/2	(0)	54	43	33	18
10. Synovial	2/20	(15)	72	66	57	42