Appendix 1. Proof of formula (3)

Denote by k_h the number of historical studies each including n_h patients.

Denote by n_c the number of patients in the current single-arm trial.

Consider the following model:

$$Y_{ij} = \mu + u_i + \varepsilon_{ij}$$

where Y_{ij} is the value of the endpoint Y for patient $j = 1, ..., n_h$ in study $i = 1, ..., k_h$.

The random variables u_i and ε_{ij} are moreover assumed independent for all i, j with respective variance τ^2 and σ^2 for all i, j.

The mean value of the endpoint Y in study i is:

$$Y_{i.} = \frac{1}{n_h}\sum_{j=1}^{n_h}(\mu + u_i + \varepsilon_{ij}) = \frac{\sum_{j=1}^{n_h}\mu + \sum_{j=1}^{n_h}u_i + \sum_{j=1}^{n_h}\epsilon_{ij}}{n_h} = \mu + u_i + \frac{1}{n_h}\sum_{j=1}^{n_h}\epsilon_{ij}$$

and the variance of this mean value is:

$$Var(Y_{i.}) == Var(\mu + u_i + \frac{1}{n_h}\sum_{j=1}^{n_h}\varepsilon_{ij}) = \tau^2 + \frac{\sigma^2}{n_h} \text{ since } u_i \text{ and } \varepsilon_{ij} \text{ are independent.}$$

The variance of the average of the mean value of the endpoint over the k_h historical studies is therefore:

$$Var(\frac{1}{k_h}\sum_{i=1}^{k_h}Y_{i.}) = \frac{\tau^2 + \frac{\sigma^2}{n_h}}{k_h} = \frac{\tau^2}{k_h} + \frac{\sigma^2}{n_h k_h}$$

Similarly, the variance of the mean value of the endpoint in the current single-arm trial is $\tau^2 + \frac{\sigma^2}{n_c}$. Note that since τ^2 is a general between-study variance there is a contribution for both the variance of the control averaged over the k historical studies and the experimental results from single current study.

The variance of the treatment contrast is therefore by independence of the studies:

$$V_{historical} = \left(\frac{\tau^2}{k_h} + \frac{\sigma^2}{n_h k_h}\right) + \left(\tau^2 + \frac{\sigma^2}{n_c}\right)$$

$$V_{historical\ comparison} = \left(\frac{\tau^2}{k_h} + \frac{\sigma^2}{n_h k_h}\right) + \left(\tau^2 + \frac{\sigma^2}{n_c}\right)$$

Appendix 2. Methodology for selection of studies in the meta-analysis of the rate of morphologic complete remission under treatment with Azacitidine.

A systematic review was performed in accordance with PRISMA guidelines (36) to identify clinical trials and observational studies of patients diagnosed with AML and treated with first-line Azacitidine monotherapy, in either elderly patients or patients unsuitable for chemotherapy. Only articles in English were included. Search results (1986 items) were screened for clinical and observational studies, and two independent researchers scanned the titles/abstract for relevance. Exclusions were identified by non-original research, wrong study type, wrong indication, or no appearance of Azacitidine treatment. Included titles (184 items) where then further screened and exclusions were identified by no Azacitidine monotherapy treatment arm available, no first-line treatment, no report on complete response (CR), not unfit for chemotherapy, no reporting of separate outcomes for MDS or AML in the study population, sub-analysis of previous published studies, and studies with fewer than 5 patients in the AML group.

Full publications of eligible articles were obtained and checked by a third independent reviewer. A total of 19 out of the 20 papers were eligible for the meta-analysis for the estimate of CR rate (the only abstract remained was excluded as only data on survival were available).

Appendix 3. Comparison to a Bayesian approach

As said before, the MAP developed by Schmidli et al (15, 16) has become a standard Bayesian approach for the integration of historical controls. In this paragraph we will compare the preceding frequentist results to those obtained when applying the MAP to the AML data. We shall slightly modify the approach to reflect the fact that we decided to use the Normal approximation to the binomial in our frequentist approach.

We assume that the true (unknown) probability of response for trial $1, \dots, k$ amongst the set of historical trials providing control data is given by $\mu_{h,i} \sim N(M, \tau^2)$ where M, τ^2 are unknown parameters describing the distribution of the unobserved effects (assumed Normally distributed) from trial to trial. On the other hand, the true (unknown) probability of response for the current trial is given by $\mu_c \sim N(M_c, \tau^2)$. Here the subscript c indicates that this is the response in the *current* trial. The parameter τ^2 is as before and M_c is also assumed unknown and normally distributed. The parameter $\theta = M_c - M$ represents the effect of the new treatment compared to control and is the parameter of primary interest. At the next (lower) level, each historical observed proportion is modelled as if it were Normally distributed with $p_{h,i} \sim N(\mu_{h,i}, \sigma_{h,i}^2)$. Here $\cancel{\sigma_i^2}\ \sigma_{h,i}^2\ $ is assumed to be a true known value. This is not quite right but is a standard assumption in meta-analysis, whether frequentist or Bayesian. In fact $\sigma_{h,i}^2 = Var\left(p_{h,i}\right)$ and so in our implementation we have estimated it using equation **Error! Reference source not found.**(7). For the current trial, the observed proportion is modelled as $p_c \sim N(\mu_c, \sigma_c^2)$. There are two obvious ways to handle σ_c^2. The first is to calculate it as $\sigma_c^2 = p_c\left(1 - p_c\right)/n_c$, where n_c is the number of patients in the current trial and

plug this is as if it were a known value. The second is to replace it by $\mu_c \left(1-\mu_c\right)/n_c$, where

μ_c is, of course, an unknown parameter to be estimated under the model. We have chosen the

former simpler course. Finally, the specification is completed by giving very vague prior

Normal distributions $N(0,100)$ to M_c and M respectively and an inverse gamma distribution

$1/\Gamma(1000,1000)$ to τ^2. The analysis then proceeds by using $p_{h,i}, \sigma^2_{h,i}, p_c, \sigma^2_c$ as inputs.

Of course, this approach only works well for reasonably sized studies but we illustrate it here

just to permit a simple comparison. In practice the original approach of Schmidli et al may be

expected to be superior.

We kept the same example of an experimental single-arm trial with $p_c = 50/100$ of all

patients having complete remission, leading to $\sigma^2_c = 0.0025$. Over 99750 MCMC iterations,

the average of τ^2 was estimated as 0.0041 with a standard deviation of 0.0029. The treatment

effect θ was on average 0.3184 with a standard deviation of 0.0833 and a 95% credible

interval of $[0.1523; 0.4820]$, leading to a critical probability $P(\theta > 0/data)$=0.9993. We

would therefore conclude that the new experimental treatment is more effective than the

standard of care. Although the 95% credible interval was slightly larger than the 95%

confidence interval of [0.1723; 0.4655] this Bayesian approach seems to offer consistent

results with the frequentist one.