Accumulation Bias: How to handle it ALL-IN

Judith ter Schure

January 27, 2021

A Technical details

This blog post discusses approaches to meta-analysis that control type-I error averaged over study series size. This is called error control *surviving over time* in Ter Schure and Grünwald (2019)), as will become more clear in the technical details below.

A.1 Time: timing and chronology

Following notation from Ter Schure and Grünwald (2019), we denote the number of studies available on a certain topic by t. This number t can also indicate the *timing* of a meta-analysis, such that a meta-analysis can possibly occur at time $t = 1, 2, 3, \ldots$ up to some maximum number of studies T. The number of studies and the timing of a meta-analysis share the notion of chronology; of past, present and future studies. At $t = 3$, we have three studies available that we can possibly meta-analyse. The fact that a third study exists can depend on the result of the first and second, but can never depend on the result of a future fourth study. Analogously, our timing of a meta-analysis after three studies can depend on the results of those three studies, but never on future meta-analyses. Note that dependencies in time are *possible*, but not necessary, to apply the notation from the accumulation basis framework (Ter Schure and Grünwald, 2019). Simultaneous studies can also be described, in which case their existence cannot depend on each other. A "no dependency"-relation does require the simultaneous studies to be assigned an arbitrary chronology, but their order plays no further role than to express a set of studies as a series. In the example of this blog, however, the extreme *Gold Rush* scenario, we assume a very real chronology and deterministic dependency between all the studies in a series.

A.2 Extreme *Gold Rush* expressed in accumulation bias notation A(t)

$A(t)$ denotes the probability that t studies accumulate and are analysed together in a meta-analysis. $A(t)$ has two components, the first indicates whether the topic "survives" the $(t-1)^{\text{th}}$ study, in which case the maximum number of studies T is larger than $t-1$ ($T \geq t$ or $T > t-1$ captured by the survival function $S(t-1)$), and the second indicates whether we bother to meta-analyze the series at its size t (the event $\mathcal{A}^{(t)}$). In our extreme *Gold Rush* world we assume that only three-study series are synthesized in a meta-analysis, such that $\mathbf{P}[\mathcal{A}^{(t)}]$ is only 1 for $\mathcal{A}^{(3)}$ and always 0 for $\mathcal{A}^{(2)}$ and $\mathcal{A}^{(1)}$ (we do not perform any 2-study or 1-study meta-analyses). In general, $A(t)$ depends on $S(t)$ and $\mathcal{A}^{(t)}$ as follows:

$$A(t \mid z_1 \ldots z_t) = \mathbf{P}[\mathcal{A}^{(t)} \mid T \geq t, z_1 \ldots z_t] \cdot S(t-1 \mid z_1 \ldots z_{t-1}) \tag{1}$$

In this simplified version of the *Gold Rush* scenario $S(t)$ is always either 0 or 1 if the study results z_1 and z_2 are known:

$$
\begin{aligned}
S(1 \mid z_1) &= \begin{cases} 1, & \text{if} \quad z_1 = \mathbf{z_1^*}. \\ 0, & \text{otherwise.} \end{cases} \\
S(2 \mid z_1, z_2) &= \begin{cases} 1, & \text{if} \quad z_1 = \mathbf{z_1^*}, z_2 = \mathbf{z_2^*}. \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
\tag{2}
$$

$S(t)$ is a survival probability, because a series can only grow to three studies (it has survive the second study ($S(2)$)) if it has also grown to two studies (it has survived the first study ($S(1)$). In contrast to the deterministic extreme *Gold Rush* in this paper, Ter Schure and Grünwald (2019) describe a probabilistic version where the probability of a replication study is larger following a significant positive result, but not always zero following a nonsignificant one. In that case we specify *hazards* of stopping after observing a certain result, which are probabilities of stopping, given that the series accumulated so far. The survival probability is defined in terms of these hazards, following standard survival analysis notation.

In our exteme *Gold Rush* any meta-analysis has zero probability to occur except for the three-study meta-analysis ($\mathbf{P}[\mathcal{A}^{(3)}] = 1$ and for all other t $\mathbf{P}[\mathcal{A}^{(t)}] = 0$ independent of the observed results $z_1, z_2, \ldots z_t$), we find nonzero $A(t)$ only for the last two scenarios below:

$$
\begin{aligned}
A(2 \mid z_1^-) &= \mathbf{P}[\mathcal{A}^{(2)} \mid T \geq 2] \cdot S(1 \mid z_1^-) = 0 \cdot 0 = 0 \\
A(2 \mid \mathbf{z_1^*}, z_2^-) &= \mathbf{P}[\mathcal{A}^{(2)} \mid T \geq 2] \cdot S(1 \mid \mathbf{z_1^*}) = 0 \cdot 1 = 0 \\
A(2 \mid \mathbf{z_1^*}, \mathbf{z_2^*}) &= \mathbf{P}[\mathcal{A}^{(2)} \mid T \geq 2] \cdot S(1 \mid \mathbf{z_1^*}) = 0 \cdot 1 = 0 \\
A(3 \mid \mathbf{z_1^*}, \mathbf{z_2^*}, z_3^- \mid T \leq 3) &= \mathbf{P}[\mathcal{A}^{(3)} \mid T \geq 3] \cdot S(2 \mid \mathbf{z_1^*}, \mathbf{z_2^*}) = 1 \cdot 1 = 1 \\
A(3 \mid \mathbf{z_1^*}, \mathbf{z_2^*}, \mathbf{z_3^*}) &= \mathbf{P}[\mathcal{A}^{(3)} \mid T \geq 3] \cdot S(2 \mid \mathbf{z_1^*}, \mathbf{z_2^*}) = 1 \cdot 1 = 1
\end{aligned}
$$

A.3 Extreme *Gold Rush* conditional sampling distribution

The sampling distribution of $Z^{(t)}$ under accumulation bias is a distribution that conditions on having t studies available and analyzing them, which happens with probability $A(t)$ given the data.

Using notation from Ter Schure and Grünwald (2019) we express the accumulation of t studies as $T \geq t$, indicating that once we have t studies available, our maximum amount of studies T is at least t (it is either t or larger). $\mathcal{A}^{(t)}$ indicates the event that we perform a meta-analysis of the t studies available. We denote the conditional sampling distribution of a $z^{(t)}$-score by $f_0(z^{(t)} \mid \mathcal{A}^{(t)}, T \geq t)$, and obtain its expression by observing that $A(t)$ is a probability of a t-study meta-analysis conditioned on the data, and we need a probability of the data conditioned on the occurrence of a t-study meta-analysis; Bayes' rule transposes the conditional in the following expression:

$$
\begin{aligned}
f_0\left(z^{(t)} \,\middle|\, \mathcal{A}^{(t)}, T \geq t\right) &= \frac{f_0(z^{(t)}) \cdot \mathbf{P}_0\left[\mathcal{A}^{(t)}, T \geq t \,\middle|\, z^{(t)}\right]}{\mathbf{P}_0\left[\mathcal{A}^{(t)}, T \geq t\right]} \\
&= \frac{f_0(z^{(t)}) \cdot \overline{A}_0\left(t \,\middle|\, z^{(t)}\right)}{\overline{A}_0\left(t\right)},
\end{aligned} \tag{3}
$$

where we define:

$$
\begin{aligned}
\overline{A}_0\left(t \,\middle|\, z^{(t)}\right) &:= \mathbf{E}_0\left[A\left(t \,|\, Z_1, \ldots, Z_t\right) \,\middle|\, Z^{(t)} = z^{(t)}\right] \\
\overline{A}_0\left(t\right) &:= \mathbf{E}_0\left[A\left(t \,|\, Z_1, \ldots, Z_t\right)\right].
\end{aligned}
$$

For the extreme *Gold Rush* scenario of this paper, and the sampling distribution of a three-study series illustrated in Figure 2, $\overline{A}_0\left(3\right)$ can be calculated as follows:

$$
\begin{aligned}
\overline{A}_0\left(3\right) &= \mathbf{E}_0[A(3 \mid Z_1, Z_2, Z_3)] \\
&= A(3 \mid \mathbf{z_1^*}, \mathbf{z_2^*}, z_3^-) \cdot \mathbf{P}_0[\mathbf{z_1^*}, \mathbf{z_2^*}, z_3^-] + A(3 \mid \mathbf{z_1^*}, \mathbf{z_2^*}, \mathbf{z_3^*}) \cdot \mathbf{P}_0[\mathbf{z_1^*}, \mathbf{z_2^*}, \mathbf{z_3^*}] \\
&= 1 \cdot \mathbf{P}_0[\mathbf{z_1^*}, \mathbf{z_2^*}, z_3^-] + 1 \cdot \mathbf{P}_0[\mathbf{z_1^*}, \mathbf{z_2^*}, \mathbf{z_3^*}] \\
&= 1 \cdot \alpha \cdot \alpha \cdot (1 - \alpha) + 1 \cdot \alpha \cdot \alpha \cdot \alpha \\
&= \frac{1}{1600} \quad (\text{for } \alpha = 2.5\%)
\end{aligned} \tag{4}
$$

The only three-study series that have nonzero $A(t)$ are $A(3 \mid \mathbf{z_1^*}, \mathbf{z_2^*}, z_3^-)$ and $A(3 \mid \mathbf{z_1^*}, \mathbf{z_2^*}, \mathbf{z_3^*})$, such that only these have to be enumerated in expectation $\overline{A}_03)$. $\overline{A}_0\left(3 \,\middle|\, z^{(t)}\right)$ can be obtained by considering all the possible combinations of Z_1, Z_2, Z_3 that could be summarized into a specific $z^{(t)}$ and taking into account their probabilities under the null hypothesis.

The value 1/1600 explains the statement in the beginning of the code in Figure 1 that 1600 first studies are needed for each sample of a three-study series.

A.4 A(t) behaves like a survival probability

Table 1 is an extension of the table in the blogpost and shows that even though $\overline{A}_0(t)$ indicates the null hypothesis probability of accumulating t studies and meta-analyzing them, it cannot in itself tell us how often the research effort is terminated at exactly those t studies. This is caused by the fact that $A(t)$ is partly a survival probability and can be illustrated by adding a column of $\overline{A}_0(t)$ values to our table that does not add up to one.

A.5 The martingale underlying the table

Table 1 is slightly modified in comparison to the blog to introduce more formal notation for the *Gold Rush* stopping rule. Here we show the specific martingale underlying this table and how Doob's Optional Stopping Theorem explains the relation between the values in the bottom row of the table.

Table 1. Possible study series under extreme *Gold Rush* accumulation bias.

	τ		$\mathbf{N}^*(\tau)$	$\overline{\mathbf{A}}_\mathbf{0}(\tau)$	$\mathbf{P}_0$	$\mathbf{N}^*(\tau) \cdot \mathbf{P}_0$	$\tau \cdot \mathbf{P}_0$
	1	z_1^-	0	1	$1-\alpha$	0	$1-\alpha$
	2	$\mathbf{z_1^*}, z_2^-$	1	α	$\alpha(1-\alpha)$	$\alpha(1-\alpha)$	$2\alpha(1-\alpha)$
	3	$\mathbf{z_1^*}, \mathbf{z_2^*}, z_3^-$	2	α^2	$\alpha^2(1-\alpha)$	$2\alpha^2(1-\alpha)$	$3\alpha^2(1-\alpha)$
$T=3$	3	$\mathbf{z_1^*}, \mathbf{z_2^*}, \mathbf{z_3^*}$	3	α^3	α^3	$3\alpha^3$	$3\alpha^3$
$\sum$					1	$\alpha+\alpha^2+\alpha^3$	$1+\alpha+\alpha^2$

We assume that each individual study Z-score is independently sampled from a standard normal distribution with mean zero, such that the probability of obtaining a significant and positive result ($\mathbf{z}^*$ if $z \geq z_\alpha$) is α. Using $\mathbf{1}_{\mathbf{z}^*}(z_i)$ for the indicator function that indicates whether z_i is significant and positive, the martingale $\{M_1, M_2, M_3, \ldots\}$ underlying Table 1 is defined as follows:

$$M_t = \sum_{i=1}^{t} \mathbf{1}_{\mathbf{z}^*}(z_i) - t\alpha.$$

$\{M_1, M_2, M_3, \ldots\}$ is a martingale since

$$\begin{aligned} \mathbf{E}_0\left[M_t - M_{t-1}\right] &= \mathbf{E}_0\left[\sum_{i=1}^{t} \mathbf{1}_{\mathbf{z}^*}(z_i) - t\alpha - \left(\sum_{i=1}^{t-1} \mathbf{1}_{\mathbf{z}^*}(z_i) - (t-1)\alpha\right)\right] \\ &= \mathbf{E}_0\left[\mathbf{1}_{\mathbf{z}^*}(z_t) - \alpha\right] = \alpha - \alpha = 0. \end{aligned} \tag{5}$$

We denote the number of significant positive studies in a series of size t by $N^*(t)$ in Table 1 and express this number in terms of M_t:

$$N^*(t) = \sum_{i=1}^{t} \mathbf{1}_{\mathbf{z}^*}(z_i) = M_t + t\alpha.$$

The *Gold Rush* stopping rule implies that we only stop accumulating studies at series size $t = \tau$ if we find the first nonsignificant study (z_τ^-, where $\tau = \min_t\{\mathbf{1}_{\mathbf{z}^*}(z_t) = 0\}$) or if we arrive at the maximum series size $t = T$. This stopping rule implies that we stop accumulating studies at either series size τ or at size T, whichever comes first. So we stop at $\tau \wedge T$. We express the expected number of studies that is significant and positive in terms of the expectation of the martingale under this stopping rule:

$$\mathbf{E}_0[N^*(\tau \wedge T)] = \mathbf{E}_0[M_{\tau \wedge T}] - \mathbf{E}_0[\tau \wedge T]\alpha,$$

and since $\tau \wedge T$ is always finite, by Doob's Optional Stopping theorem we have:

$$\mathbf{E}_0[M_{\tau\wedge T}] = \mathbf{E}_0[M_1] = \mathbf{E}_0\left[\sum_{i=1}^{1} \mathbf{1}_{\mathbf{z}^*}(z_i) - 1\alpha\right] = \mathbf{E}_0[\mathbf{1}_{\mathbf{z}^*}(z_i)] - \alpha = \alpha - \alpha = 0$$

$$\begin{gathered} \text{such that} \\ \mathbf{E}_0[N^*(\tau \wedge T)] = \mathbf{E}_0[\tau \wedge T]\alpha \\ \text{and} \\ \frac{\mathbf{E}_0[N^*(\tau \wedge T)]}{\mathbf{E}_0[\tau \wedge T]} = \alpha \end{gathered} \tag{6}$$

This is shown for the *Gold Rush* stopping rule and $T = 3$ by Table 1, but holds for any stopping rule and finite T.

References

Judith ter Schure and Peter Grünwald. Accumulation Bias in meta-analysis: the need to consider time in error control [version 1; peer review: 2 approved]. *F1000Research*, 8:962, June 2019. ISSN 2046-1402. doi: 10.12688/f1000research.19375.1. URL https://f1000research.com/articles/8-962/v1.