Submitted to *Service Science*
manuscript SS-07-21-ART-0071.R2

Online Appendix for "Adaptive Design of Personalized Dose-Finding Clinical Trials"

Saeid Delshad
Department of Industrial Engineering, Clemson University, sdelsha@clemson.edu,

Amin Khademi
Department of Industrial Engineering, Clemson University, khademi@clemson.edu,

Appendix A: Proofs

A.1. Proof of Proposition 1

We use backward induction on n. For $n = N-1$, we can show $Q^{N-1}(s,z,x') \le V^N(s)$ as follows

$$\begin{aligned}
Q^{N-1}(s,z,x') &= \mathbb{E}\Big\{V^N\big(\eta(s^{N-1},z^{N-1},x^{N-1},\omega_y^N)\big)\Big|s^{N-1}=s,z^{N-1}=z,x^{N-1}=x'\Big\}\\
&= \mathbb{E}\Big\{\mathbb{E}_x\big\{\mathrm{Var}\big(g(\Theta,x)\big|\eta(s^{N-1},z^{N-1},x^{N-1},\omega_y^N)\big)\big\}\Big|.\Big\}\\
&\overset{\langle 1\rangle}{=} \mathbb{E}_x\mathbb{E}\Big\{\mathrm{Var}\big(g(\Theta,x)\big|\eta(s^{N-1},z^{N-1},x^{N-1},\omega_y^N)\big)\Big|.\Big\}\\
&= \mathbb{E}_x\mathbb{E}\Big\{\Big[\mathbb{E}\big(g^2(\Theta,x)\big|\eta(s^{N-1},z^{N-1},x^{N-1},\omega_y^N)\big)\Big]-\Big[\mathbb{E}\big(g(\Theta,x)\big|\eta(s^{N-1},z^{N-1},x^{N-1},\omega_y^N)\big)\Big]^2\Big|.\Big\}\\
&= \mathbb{E}_x\Big\{\mathbb{E}\Big\{\Big[\mathbb{E}\big(g^2(\Theta,x)\big|\eta(s^{N-1},z^{N-1},x^{N-1},\omega_y^N)\big)\Big]\Big|.\Big\}\\
&\quad -\mathbb{E}\Big\{\Big[\mathbb{E}\big(g(\Theta,x)\big|\eta(s^{N-1},z^{N-1},x^{N-1},\omega_y^N)\big)\Big]^2\Big|.\Big\}\Big\}\\
&\overset{\langle 2\rangle}{\le} \mathbb{E}_x\Big\{\mathbb{E}\Big\{\Big[\mathbb{E}\big(g^2(\Theta,x)\big|\eta(s^{N-1},z^{N-1},x^{N-1},\omega_y^N)\big)\Big]\Big|.\Big\}\\
&\quad -\Big[\mathbb{E}\Big\{\mathbb{E}\big(g(\Theta,x)\big|\eta(s^{N-1},z^{N-1},x^{N-1},\omega_y^N)\big)\Big|.\Big\}\Big]^2\Big\}\\
&\overset{\langle 3\rangle}{=} \mathbb{E}_x\Big\{\mathbb{E}\Big[g^2(\Theta,x)\Big|.\Big]-\Big(\mathbb{E}\Big[g(\Theta,x)\Big|.\Big]\Big)^2\Big\}\\
&= \mathbb{E}_x\big\{\mathrm{Var}\big(g(\Theta,x)|s^{N-1}=s,z^{N-1}=z,x^{N-1}=x'\big)\big\}=V^N(s),
\end{aligned}$$

where $\langle 1\rangle$ is true because the variance is non-negative and $g(\Theta,x)\in\mathcal{Z}$ has a bounded support, $\langle 2\rangle$ is the result of Jensen's inequality for convex functions, and $\langle 3\rangle$ is the result of tower property of conditional expectation as $\sigma\big(s,z,x'\big)\subseteq\sigma\big(\eta(s,z,x',\omega_y)\big)$ where $\sigma(\cdot)$ is the σ-algebra operator. Then assuming

$$Q^t(s,z,x')=\mathbb{E}\Big\{V^{t+1}\big(\eta(s^t,z^t,x^t,\omega_y^{t+1})\big)\Big|s^t=s,z^t=z,x^t=x'\Big\}\le V^{t+1}(s),\quad t=n+1,\dots,N-1,$$

all we need to do is to show that the induction hypothesis is true for n, which is done as follows

$$\begin{aligned}
Q^n(s,z,x') &= \mathbb{E}\Big\{V^{n+1}\big[\eta(s^n,z^n,x^n,\omega_y^{n+1})\big]\Big|s^n=s,z^n=z,x^n=x'\Big\}\\
&\overset{\langle 4\rangle}{=} \mathbb{E}\Big\{\min_{z'} Q^{n+1}\big[\eta(s^n,z^n,x^n,\omega_y^{n+1}),z',x^{n+1}\big]\Big|.\Big\}\\
&\overset{\langle 5\rangle}{\leq} \min_{z'}\mathbb{E}\Big\{Q^{n+1}\big[\eta(s^n,z^n,x^n,\omega_y^{n+1}),z',x^{n+1}\big]\Big|.\Big\}\\
&= \min_{z'}\mathbb{E}\Big\{\mathbb{E}\Big(V^{n+2}\big[\eta\big(\eta(s^n,z^n,x^n,\omega_y^{n+1}),z',x^{n+1},\omega_y^{n+2}\big)\big]\big|\eta(s^n,z^n,x^n,\omega_y^{n+1}),z',x^{n+1}\Big)\Big|.\Big\}\\
&\overset{\langle 6\rangle}{=} \min_{z'}\mathbb{E}\Big\{V^{n+2}\big[\eta\big(\eta(s^n,z^n,x^n,\omega_y^{n+1}),z',x^{n+1},\omega_y^{n+2}\big)\big]\Big|.\Big\}\\
&\overset{\langle 7\rangle}{=} \min_{z'}\mathbb{E}\Big\{V^{n+2}\big[\eta\big(\eta(s^n,z',x^{n+1},\omega_y^{n+2}),z^n,x^n,\omega_y^{n+1}\big)\big]\Big|.\Big\}\\
&\overset{\langle 8\rangle}{=} \min_{z'}\mathbb{E}\Big\{\mathbb{E}\Big(V^{n+2}\big[\eta\big(\eta(s^n,z',x^{n+1},\omega_y^{n+2}),z^n,x^n,\omega_y^{n+1}\big)\big]\Big|\eta(s^n,z',x^{n+1},\omega_y^{n+2}),z^n,x^n\Big)\Big|.\Big\}\\
&= \min_{z'}\mathbb{E}\Big\{Q^{n+1}\big[\eta(s^n,z',x^{n+1},\omega_y^{n+2}),z^n,x^n\big]\Big|s^n=s,z^n=z,x^n=x'\Big\},
\end{aligned}$$

where $\langle 4\rangle$ is true because $V^{n+1}(s)=\min_z Q^{n+1}(s,z,x')$, $\langle 5\rangle$ is correct due to Jensen's inequality for concave functions, $\langle 6\rangle$ and $\langle 8\rangle$ is the result of tower property of conditional expectation as $\sigma\big(s,z,x'\big)\subseteq\sigma\big(\eta(s,z,x',\omega_y)\big)$, and $\langle 7\rangle$ is true because changing the measurement order of z' and z^n does not change the final resulting distribution. Also, note that ω_y^{n+1} is independent of ω_y^{n+2}. Therefore, we have

$$\begin{aligned}
&Q^n(s,z,x')\leq \min_{z'}\mathbb{E}\Big\{Q^{n+1}\big[\eta(s^n,z',x^{n+1},\omega_y^{n+2}),z^n,x^n\big]\Big|s^n=s,z^n=z,x^n=x'\Big\}\\
&\overset{\langle 9\rangle}{\leq} \min_{z'}\mathbb{E}\Big\{V^{n+2}\big[\eta(s^n,z',x^{n+1},\omega_y^{n+2})\big]\Big|s^n=s,z^n=z,x^n=x'\Big\}=V^{n+1}(s),
\end{aligned}$$

where $\langle 9\rangle$ is true because of induction hypothesis.

A.2. Proof of Corollary 1

The belief about the doses are independent, i.e., $\mathrm{Cov}(\Theta_z,\Theta_{z'})=0,\forall z,z'\in\mathcal{Z}$, and we also have $x^n\Theta_z|\mathcal{F}^n\sim\mathcal{N}(x^n\mu_z^n,x^n\Sigma_z^n(x^n)^\top)$, where $x^n\Sigma_z^n(x^n)^\top=0$; hence, by updating equations, measuring dose z does not change our belief, i.e., $\eta(s^n,z,x^n,\omega_y^{n+1})=s^n$. Therefore, following Theorem 1 for all $z'\neq z$ we have

$$Q^n(s^n,z,x^n)=\mathbb{E}\Big\{V^{n+1}\big[\eta(s^n,z,x^n,\omega_y^{n+1})\big]\Big|s^n,z,x^n\Big\}=V^{n+1}(s^n)\geq Q^n(s^n,z',x^n).$$

A.3. Proof of Corollary 2

By Theorem 1 we have $Q^n(s,z,x')\leq V^{n+1}(s),\forall x',z$, and therefore, $\min_z Q^n(s,z,x')\leq V^{n+1}(s)$. On the other hand, $V^n(s,x')=\min_z Q^n(s,z,x')$; hence

$$\begin{aligned}
&V^n(s,x')\leq V^{n+1}(s),\quad \forall x'\in\mathcal{X}\\
&\Longrightarrow \mathbb{E}_x[V^n(s,x')]\leq V^{n+1}(s)\\
&\Longrightarrow V^n(s)\leq V^{n+1}(s).
\end{aligned}$$

A.4. Proof of Proposition 2

Given $\mathcal{F}^{n+}$, which includes the patient covariate x^n, sampling from dose $z^n=z$ leads to the posterior predictive distribution $y_z^{n+1}\sim\mathcal{N}(x^n\mu_z^n,\sigma_z^2+x^n\Sigma_z^n(x^n)^\top)$, i.e., defining a standard normal random variable

$\mathfrak{Z}^{n+1}$, we have $y_z^{n+1} = x^n\mu_z^n + \sqrt{\sigma_z^2 + x^n\Sigma_z^n(x^n)^\top}\mathfrak{Z}^{n+1}$. Given the Bayesian updating equations we have

$$\begin{aligned}
\mu_z^{n+1} &= \Sigma_z^{n+1}(\Sigma_z^n)^{-1}\mu_z^n + \sigma_z^{-2}\Sigma_z^{n+1}(x^n)^\top\left[x^n\mu_z^n + \sqrt{\sigma_z^2 + x^n\Sigma_z^n(x^n)^\top}\mathfrak{Z}^{n+1}\right]\\
&= \Sigma_z^{n+1}(\Sigma_z^n)^{-1}\mu_z^n + \sigma_z^{-2}\Sigma_z^{n+1}(x^n)^\top x^n\mu_z^n + \sigma_z^{-2}\Sigma_z^{n+1}(x^n)^\top\sqrt{\sigma_z^2 + x^n\Sigma_z^n(x^n)^\top}\mathfrak{Z}^{n+1}\\
&= \Sigma_z^{n+1}\left[(\Sigma_z^n)^{-1} + \sigma_z^{-2}(x^n)^\top x^n\right]\mu_z^n + \sigma_z^{-2}\Sigma_z^{n+1}(x^n)^\top\sqrt{\sigma_z^2 + x^n\Sigma_z^n(x^n)^\top}\mathfrak{Z}^{n+1}\\
&= \Sigma_z^{n+1}\left[\Sigma_z^{n+1}\right]^{-1}\mu_z^n + \sigma_z^{-2}\Sigma_z^{n+1}(x^n)^\top\sqrt{\sigma_z^2 + x^n\Sigma_z^n(x^n)^\top}\mathfrak{Z}^{n+1}\\
&= \mu_z^n + \left[\sigma_z^{-2}\Sigma_z^{n+1}(x^n)^\top\sqrt{\sigma_z^2 + x^n\Sigma_z^n(x^n)^\top}\right]\mathfrak{Z}^{n+1}.
\end{aligned}$$

Then, we have for the change in the covariance of dose $z \in \mathcal{Z}$ assuming that we sample from it for a patient with covariate vector x^n

$$\begin{aligned}
\tilde{\Sigma}_z^n &= \mathrm{Cov}(\mu_z^{n+1}|\mathcal{F}^{n+}, z^n = z)\\
&= \mathrm{Cov}(\mu_z^{n+1} - \mu_z^n|\mathcal{F}^{n+}, z^n = z)\\
&= \left[\sigma_z^{-2}\Sigma_z^{n+1}(x^n)^\top\sqrt{\sigma_z^2 + x^n\Sigma_z^n(x^n)^\top}\right]\left[\sigma_z^{-2}\Sigma_z^{n+1}(x^n)^\top\sqrt{\sigma_z^2 + x^n\Sigma_z^n(x^n)^\top}\right]^\top\\
&= \left[\sigma_z^{-2} + \sigma_z^{-4}x^n\Sigma_z^n(x^n)^\top\right]\Sigma_z^{n+1}(x^n)^\top x^n\Sigma_z^{n+1}.
\end{aligned}$$

For a total of two alternative doses with independent posterior beliefs, i.e., $\mathcal{Z} = \{z_1, z_2\}$, and for any patient with covariate vector $x \in \mathcal{X}$, we have

$$\begin{aligned}
&\Theta_{z_1}|\mathcal{F}^n \sim \mathcal{N}(\mu_{z_1}^n, \Sigma_{z_1}^n), \quad y_{z_1}^{n+1}|(\Theta_{z_1}, x) \sim \mathcal{N}(x\Theta_{z_1}, \sigma_{z_1}^2),\\
&\Theta_{z_2}|\mathcal{F}^n \sim \mathcal{N}(\mu_{z_2}^n, \Sigma_{z_2}^n), \quad y_{z_2}^{n+1}|(\Theta_{z_2}, x) \sim \mathcal{N}(x\Theta_{z_2}, \sigma_{z_2}^2).
\end{aligned}$$

Also, given $z_1 < z_2$, ED_L^x is as follows

$$\mathrm{ED}_L^x = \begin{cases} z_1 & \text{if} \quad x\Theta_{z_1} \geq Lx\Theta_{z_2},\\ z_2 & \text{if} \quad x\Theta_{z_1} < Lx\Theta_{z_2}. \end{cases}$$

Hence, the expectation and the variance of the target dose $\mathrm{ED}_L^x = g(\Theta, x)$ at epoch n is given by

$$\begin{aligned}
\mathbb{E}_n\left[g(\Theta,x)\right] &= z_1\mathbb{P}_n(x\Theta_{z_1} \geq Lx\Theta_{z_2}) + z_2\mathbb{P}_n(x\Theta_{z_1} < Lx\Theta_{z_2}),\\
\mathrm{Var}_n\left[g(\Theta,x)\right] &= z_1^2\mathbb{P}_n(x\Theta_{z_1} \geq Lx\Theta_{z_2}) + z_2^2\mathbb{P}_n(x\Theta_{z_1} < Lx\Theta_{z_2})\\
&\quad - \left[z_1\mathbb{P}_n(x\Theta_{z_1} \geq Lx\Theta_{z_2}) + z_2\mathbb{P}_n(x\Theta_{z_1} < Lx\Theta_{z_2})\right]^2\\
&= \left[z_2 - z_1\right]^2\left[\mathbb{P}_n(x\Theta_{z_1} < Lx\Theta_{z_2})\right]\left[1 - \mathbb{P}_n(x\Theta_{z_1} < Lx\Theta_{z_2})\right].
\end{aligned}$$

Note that since the beliefs about the doses are assumed to be independent, the mean and covariance matrix of any dose that is not sampled will not change. Hence, at decision epoch n, given sampling dose z_1 and z_2 for a patient with covariate vector $x^n = x'$ we have

$$\begin{aligned}
&(x\Theta_{z_1} - Lx\Theta_{z_2})|(\mathcal{F}^n, z^n = z_1, x^n = x') \sim \mathcal{N}(x\mu_{z_1}^n - Lx\mu_{z_2}^n, x\Sigma_{z_1}^{n+1}x^\top + L^2x\Sigma_{z_2}^n x^\top),\\
&(x\Theta_{z_1} - Lx\Theta_{z_2})|(\mathcal{F}^n, z^n = z_2, x^n = x') \sim \mathcal{N}(x\mu_{z_1}^n - Lx\mu_{z_2}^n, x\Sigma_{z_1}^n x^\top + L^2x\Sigma_{z_2}^{n+1}x^\top).
\end{aligned}$$

Then, by denoting $\Phi(\cdot)$ as the standard normal cumulative distribution function, for the expected variance of the target dose at epoch $n+1$ given sampling from dose z_1 at epoch n, we have

$$\begin{aligned}
&\mathbb{E}\Big\{\mathbb{E}_x\big(\mathrm{Var}_{n+1}\left[g(\Theta,x)\right]\big)\Big|z^n = z_1, x^n = x'\Big\}\\
&= \mathbb{E}\Big\{\left[z_2 - z_1\right]^2\sum_{x\in\mathcal{X}}\mathcal{P}_x\left[\mathbb{P}_{n+1}(x\Theta_{z_1} < Lx\Theta_{z_2})\right]\left[1 - \mathbb{P}_{n+1}(x\Theta_{z_1} < Lx\Theta_{z_2})\right]\Big|z^n = z_1, x^n = x'\Big\}\\
&= \left[z_2 - z_1\right]^2\sum_{x\in\mathcal{X}}\mathcal{P}_x\Phi\left(\frac{Lx\mu_{z_2}^n - x\mu_{z_1}^n}{\sqrt{x\Sigma_{z_1}^{n+1}x^\top + L^2x\Sigma_{z_2}^n x^\top}}\right)\left[1 - \Phi\left(\frac{Lx\mu_{z_2}^n - x\mu_{z_1}^n}{\sqrt{x\Sigma_{z_1}^{n+1}x^\top + L^2x\Sigma_{z_2}^n x^\top}}\right)\right],
\end{aligned}$$

and given sampling from dose 2 at epoch n, we have

$$\begin{aligned}
&\mathbb{E}\Big\{\mathbb{E}_x\big(\mathrm{Var}_{n+1}\big[g(\Theta,x)\big]\big)\Big|z^n=z_2,x^n=x'\Big\}\\
&=\mathbb{E}\Big\{\big[z_2-z_1\big]^2\sum_{x\in\mathcal{X}}\mathcal{P}_x\big[\mathbb{P}_{n+1}(x\Theta_{z_1}<Lx\Theta_{z_2})\big]\big[1-\mathbb{P}_{n+1}(x\Theta_{z_1}<Lx\Theta_{z_2})\big]\Big|z^n=z_2,x^n=x'\Big\}\\
&=\big[z_2-z_1\big]^2\sum_{x\in\mathcal{X}}\mathcal{P}_x\Phi\bigg(\frac{Lx\mu^n_{z_2}-x\mu^n_{z_1}}{\sqrt{x\Sigma^n_{z_1}x^\top+L^2x\Sigma^{n+1}_{z_2}x^\top}}\bigg)\bigg[1-\Phi\bigg(\frac{Lx\mu^n_{z_2}-x\mu^n_{z_1}}{\sqrt{x\Sigma^n_{z_1}x^\top+L^2x\Sigma^{n+1}_{z_2}x^\top}}\bigg)\bigg].
\end{aligned}$$

DOL's allocation function can be calculated using the closed form formula given above. In fact, DOL allocates to the dose which results in the smallest expected (over covariates) variance of the target dose at the next epoch. Therefore, for two allowable doses implementing DOL becomes easy and fast. Note that this closed form depends on the distribution of the covariates $\mathcal{P}_x$ which is assumed to be known.

A.5. Proof of Proposition 3

By considering only one patient type (homogeneous patients), this problem reduces to *impersonalized* dose-finding problem with independent beliefs about the alternative doses. Therefore, we can use a bandit model instead of a linear bandit model to find the unknown parameter vector $\theta=[\theta_1,\ldots,\theta_Z]^\top$ for the set of doses $\mathcal{Z}$. Given z and θ_z, the sampled response y^{n+1}_z has a normal distribution as $(y^{n+1}_z|z,\theta_z)\sim\mathcal{N}(\theta_z,\sigma^2_z),\forall z,n$. In this setting, denoting $\mathcal{F}^n$ as the filtration created by the set $\{\mu^0,\Sigma^0,z^0,y^1_{z^0},\ldots,z^{n-1},y^n_{z^{n-1}}\}$, the state of the trial at epoch n is defined as $s^n=(\mu^n,\Sigma^n)$ in which $\mu^n=[\mu^n_z:z\in\mathcal{Z}]^\top$ where $\mu^n_z=\mathbb{E}_n(\theta_z)$ is the mean of θ_z given $\mathcal{F}^n$ and $\Sigma^n=[\Sigma^n_z:z\in\mathcal{Z}]^\top$ where $\Sigma^n_z=\mathrm{Var}_n(\theta_z)$ is the variance of θ_z given $\mathcal{F}^n$. Consequently, equations (1), and (9) are replaced respectively with equations (1), (2), and (3) for any alternative dose $z\in\mathcal{Z}$ sampled at epoch $n=0,\ldots,N-1$.

$$y^{n+1}_z=\theta_z+\epsilon^{n+1}_z. \tag{1}$$

$$\begin{aligned}
\Sigma^{n+1}_z&=\big[(\Sigma^n_z)^{-1}+\sigma_z^{-2}\big]^{-1},\\
\mu^{n+1}_z&=\Sigma^{n+1}_z(\Sigma^n_z)^{-1}\mu^n_z+\sigma_z^{-2}\Sigma^{n+1}_z y^{n+1}_z.
\end{aligned} \tag{2}$$

$$\tilde{\Sigma}^n_z=\mathrm{Var}_n(\mu^{n+1}_z|z^n=z)=\mathrm{Var}_n(\mu^{n+1}_z-\mu^n_z|z^n=z)=\Sigma^n_z-\Sigma^{n+1}_z=\Sigma^n_z-\big[(\Sigma^n_z)^{-1}+\sigma_z^{-2}\big]^{-1}. \tag{3}$$

Note that in this setting, $\tilde{\Sigma}^n_z$ is the change in the variance of dose z given measuring it at epoch n. Since the beliefs about the doses are assumed to be independent, the mean and variance of any dose $z'\neq z$, which is not sampled, will not change. By the above formulation, the proposed policies can be simply modified and used for the impersonalized version of dose-finding problem where we look for a single target dose for all patients. To that end, for the impersonalized dose-finding problem with a total of two alternative doses as $\mathcal{Z}=\{z_1,z_2\}$ with $z_1<z_2$ given independent posterior beliefs we have

$$\begin{aligned}
\theta_{z_1}|\mathcal{F}^n&\sim\mathcal{N}(\mu^n_{z_1},\Sigma^n_{z_1}),\quad y^{n+1}_{z_1}|\theta_{z_1}\sim\mathcal{N}(\theta_{z_1},\sigma^2_{z_1}),\\
\theta_{z_2}|\mathcal{F}^n&\sim\mathcal{N}(\mu^n_{z_2},\Sigma^n_{z_2}),\quad y^{n+1}_{z_2}|\theta_{z_2}\sim\mathcal{N}(\theta_{z_2},\sigma^2_{z_2}).
\end{aligned}$$

Also, the impersonalized $(L\times 100)\%$ effectiveness target dose denoted as ED_L is as follows

$$\mathrm{ED}_L=\begin{cases}z_1 & \text{if}\quad \theta_{z_1}\geq L\theta_{z_2},\\ z_2 & \text{if}\quad \theta_{z_1}<L\theta_{z_2}.\end{cases}$$

Hence, the expectation and the variance of the target dose $\mathrm{ED}_L=g(\theta)$ at epoch n is given by

$$\mathbb{E}_n\big[g(\theta)\big]=z_1\mathbb{P}_n(\theta_{z_1}\geq L\theta_{z_2})+z_2\mathbb{P}_n(\theta_{z_1}<L\theta_{z_2}),$$

$$\begin{aligned}\mathrm{Var}_n\big[g(\theta)\big] &= z_1^2\mathbb{P}_n(\theta_{z_1} \geq L\theta_{z_2}) + z_2^2\mathbb{P}_n(\theta_{z_1} < L\theta_{z_2}) - \big[z_1\mathbb{P}_n(\theta_{z_1} \geq L\theta_{z_2}) + z_2\mathbb{P}_n(\theta_{z_1} < L\theta_{z_2})\big]^2 \\ &= \big[z_2 - z_1\big]^2\big[\mathbb{P}_n(\theta_{z_1} < L\theta_{z_2})\big]\big[1 - \mathbb{P}_n(\theta_{z_1} < L\theta_{z_2})\big].\end{aligned}$$

Also, given sampling z_1 we have $(\theta_{z_1} - L\theta_{z_2})|(\mathcal{F}^n, z^n = z_1) \sim \mathcal{N}(\mu^n_{z_1} - L\mu^n_{z_2}, \Sigma^{n+1}_{z_1} + L^2\Sigma^n_{z_2})$, and given sampling z_2 we have $(\theta_{z_1} - L\theta_{z_2})|(\mathcal{F}^n, z^n = z_2) \sim \mathcal{N}(\mu^n_{z_1} - L\mu^n_{z_2}, \Sigma^n_{z_1} + L^2\Sigma^{n+1}_{z_2})$. Then, for the expected variance of the target dose at epoch $n+1$ given sampling from dose z_1 and z_2, respectively we have

$$\begin{aligned}\mathbb{E}\Big\{\mathrm{Var}_{n+1}\big[g(\theta)\big]\Big|z^n = z_1\Big\} &= \mathbb{E}\Big\{\big[z_2 - z_1\big]^2\big[\mathbb{P}_{n+1}(\theta_{z_1} < L\theta_{z_2})\big]\big[1 - \mathbb{P}_{n+1}(\theta_{z_1} < L\theta_{z_2})\big]\Big|z^n = z_1\Big\} \\ &= \big[z_2 - z_1\big]^2\Phi\bigg(\frac{L\mu^n_{z_2} - \mu^n_{z_1}}{\sqrt{\Sigma^{n+1}_{z_1} + L^2\Sigma^n_{z_2}}}\bigg)\bigg[1 - \Phi\bigg(\frac{L\mu^n_{z_2} - \mu^n_{z_1}}{\sqrt{\Sigma^{n+1}_{z_1} + L^2\Sigma^n_{z_2}}}\bigg)\bigg],\end{aligned}$$

$$\begin{aligned}\mathbb{E}\Big\{\mathrm{Var}_{n+1}\big[g(\theta)\big]\Big|z^n = z_2\Big\} &= \mathbb{E}\Big\{\big[z_2 - z_1\big]^2\big[\mathbb{P}_{n+1}(\theta_{z_1} < L\theta_{z_2})\big]\big[1 - \mathbb{P}_{n+1}(\theta_{z_1} < L\theta_{z_2})\big]\Big|z^n = z_2\Big\} \\ &= \big[z_2 - z_1\big]^2\Phi\bigg(\frac{L\mu^n_{z_2} - \mu^n_{z_1}}{\sqrt{\Sigma^n_{z_1} + L^2\Sigma^{n+1}_{z_2}}}\bigg)\bigg[1 - \Phi\bigg(\frac{L\mu^n_{z_2} - \mu^n_{z_1}}{\sqrt{\Sigma^n_{z_1} + L^2\Sigma^{n+1}_{z_2}}}\bigg)\bigg],\end{aligned}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Let $\mathcal{A}_{z_1} = \dfrac{L\mu^n_{z_2} - \mu^n_{z_1}}{\sqrt{\Sigma^{n+1}_{z_1} + L^2\Sigma^n_{z_2}}}$ and $\mathcal{A}_{z_2} = \dfrac{L\mu^n_{z_2} - \mu^n_{z_1}}{\sqrt{\Sigma^n_{z_1} + L^2\Sigma^{n+1}_{z_2}}}$. Note that if $\mu^n_{z_1} > L\mu^n_{z_2}$ then $\Phi(\cdot) < 0.5$ which results in $\Phi(\cdot)[1-\Phi(\cdot)]$ being increasing, hence

$$\begin{cases}\mu^n_{z_1} > L\mu^n_{z_2} \\ \Sigma^{n+1}_{z_1} + L^2\Sigma^n_{z_2} < \Sigma^n_{z_1} + L^2\Sigma^{n+1}_{z_2}\end{cases} \implies \Phi(\mathcal{A}_{z_1})[1-\Phi(\mathcal{A}_{z_1})] < \Phi(\mathcal{A}_{z_2})[1-\Phi(\mathcal{A}_{z_2})],$$

$$\begin{cases}\mu^n_{z_1} > L\mu^n_{z_2} \\ \Sigma^{n+1}_{z_1} + L^2\Sigma^n_{z_2} > \Sigma^n_{z_1} + L^2\Sigma^{n+1}_{z_2}\end{cases} \implies \Phi(\mathcal{A}_{z_1})[1-\Phi(\mathcal{A}_{z_1})] > \Phi(\mathcal{A}_{z_2})[1-\Phi(\mathcal{A}_{z_2})].$$

However, if $\mu^n_{z_1} < L\mu^n_{z_2}$ then $\Phi(\cdot) > 0.5$ which results in $\Phi(\cdot)[1-\Phi(\cdot)]$ being decreasing, hence

$$\begin{cases}\mu^n_{z_1} < L\mu^n_{z_2} \\ \Sigma^{n+1}_{z_1} + L^2\Sigma^n_{z_2} < \Sigma^n_{z_1} + L^2\Sigma^{n+1}_{z_2}\end{cases} \implies \Phi(\mathcal{A}_{z_1})[1-\Phi(\mathcal{A}_{z_1})] < \Phi(\mathcal{A}_{z_2})[1-\Phi(\mathcal{A}_{z_2})],$$

$$\begin{cases}\mu^n_{z_1} < L\mu^n_{z_2} \\ \Sigma^{n+1}_{z_1} + L^2\Sigma^n_{z_2} > \Sigma^n_{z_1} + L^2\Sigma^{n+1}_{z_2}\end{cases} \implies \Phi(\mathcal{A}_{z_1})[1-\Phi(\mathcal{A}_{z_1})] > \Phi(\mathcal{A}_{z_2})[1-\Phi(\mathcal{A}_{z_2})].$$

Therefore, the condition on $\mu^n_1 > L\mu^n_2$ is irrelevant and generally we have

$$\begin{aligned}\Sigma^{n+1}_{z_1} + L^2\Sigma^n_{z_2} < \Sigma^n_{z_1} + L^2\Sigma^{n+1}_{z_2} &\implies \big[z_2 - z_1\big]^2\Phi(\mathcal{A}_{z_1})[1-\Phi(\mathcal{A}_{z_1})] < \big[z_2 - z_1\big]^2\Phi(\mathcal{A}_{z_2})[1-\Phi(\mathcal{A}_{z_2})] \\ &\implies z^n_{DOL} = z_1, \\ \Sigma^{n+1}_{z_1} + L^2\Sigma^n_{z_2} > \Sigma^n_{z_1} + L^2\Sigma^{n+1}_{z_2} &\implies \big[z_2 - z_1\big]^2\Phi(\mathcal{A}_{z_1})[1-\Phi(\mathcal{A}_{z_1})] > \big[z_2 - z_1\big]^2\Phi(\mathcal{A}_{z_2})[1-\Phi(\mathcal{A}_{z_2})] \\ &\implies z^n_{DOL} = z_2.\end{aligned}$$

Now, by equation (3), we know $\Sigma^{n+1}_{z_1} = \big[(\Sigma^n_{z_1})^{-1} + \sigma^{-2}_{z_1}\big]^{-1}$ and $\Sigma^{n+1}_{z_2} = \big[(\Sigma^n_{z_2})^{-1} + \sigma^{-2}_{z_2}\big]^{-1}$. By plugging these values and simplifying the above inequalities, we obtain the following result for DOL in the impersonalized trial with two doses

$$\tilde{\Sigma}^n_{z_1} > L^2\tilde{\Sigma}^n_{z_2} \implies z^n_{DOL} = z_1,$$

$$\tilde{\Sigma}^n_{z_1} < L^2\tilde{\Sigma}^n_{z_2} \implies z^n_{DOL} = z_2,$$

meaning that for $L = 1$, DOL allocates sampling to the alternative dose which reduces the posterior variance the most.

A.6. Proof of Proposition 4

In the impersonalized dose-finding problem, we consider a DM who seeks to minimize the expected variance of $\mathrm{ED}_L = g(\theta)$ at the end of the decision process. Therefore, the expected variance of ED_L at the end of the trial given initial state s^0 under any given policy π is $l_\pi^N(s^0) = \mathbb{E}_\pi\{\mathrm{Var}(g(\theta)|\mathcal{F}^N)|s^0\}$, where $\mathbb{E}_\pi(\cdot)$ indicates the expectation taken with respect to probability measure induced by policy π. Thus, having an initial prior s^0, the DM solves for

$$V^0(s^0) = \inf_{\pi\in\Pi} l_\pi^N(s^0).$$

Let s^n be the state of the system at each decision epoch $n \in \{0,1,\ldots,N-1\}$, then we have $A(s^n) := \mathcal{Z}$ as the action space and $z^n \in A(s^n)$ as the action taken in this epoch. Let $V^n(s^n)$ denote the value function at epoch n, which is a unique solution to the following Bellman equations

$$\begin{aligned} V^n(s^n) &= \min_{z^n\in A(s^n)} \Big\{\mathbb{E}\big\{V^{n+1}(s^{n+1})\big|s^n,z^n\big\}\Big\}, \quad n=0,1,\ldots,N-1,\\ V^N(s^N) &= \mathrm{Var}\big(g(\theta)|s^N\big). \end{aligned}$$

Also, redefine $\eta(\cdot)$ to be the state transition function with respect to equation (2), i.e., $s^{n+1} = \eta(s^n,z^n,\omega_y^{n+1})$, where ω_y^{n+1} represents the randomness in the sample response $y_{z^n}^{n+1}$.

Frazier et al. (2008) argued in Theorem 7.1 of that paper for the KG policy that: "If the KG persistence property holds on a covering $\{\mathcal{S}^n\}_{n=k}^N$ of the future from k for some $k \in \{0,\ldots,N-1\}$, then $V^{KG,k}(s) = V^k(s)$ for every $s \in \mathcal{S}^k$".

Given the above-mentioned value and transition functions for the impersonalized dose-finding trial, similar to Theorem 7.1 of Frazier et al. (2008) for the standard KG, we can show that, the DOL persistence property on a covering $\{\mathcal{S}^n\}_{n=k}^N$ of the future from k for some $k \in \{0,\ldots,N-1\}$ results in DOL optimality, i.e., $V^{DOL,k}(s) = V^k(s)$ for every $s \in \mathcal{S}^k$. Since, the above-mentioned setup for the impersonalized dose-finding problem is similar to the setup of Frazier et al. (2008) (with the exception of the objective function), the argument to prove this theorem for DOL closely follows the steps of the proof of Theorem 7.1 of Frazier et al. (2008). Therefore, we omit the proof and refer the readers to Frazier et al. (2008) to see how persistence property results in optimality for one-step look-ahead policies.

Yet to prove DOL optimality for two doses, we need to show that given the allocation rule stated in Proposition 3, the DOL persistence property holds true. To that end, let $n \in \{0,\ldots,N-1\}$ and $s^n = (\mu^n,\Sigma^n) \in \mathcal{S}$. In case $\tilde{\Sigma}_{z_1}^n > L^2\tilde{\Sigma}_{z_2}^n$, we have $z_{DOL}^n(s^n) = z_1$. In this case, given sampling dose z_2, the variance of $s^{n+1} = \eta(s^n,z_2,\omega_y^{n+1})$ is as $\Sigma_{z_1}^{n+1} = \Sigma_{z_1}^n$ and $\Sigma_{z_2}^{n+1} = \big[(\Sigma_{z_2}^n)^{-1} + \sigma_{z_2}^{-2}\big]^{-1}$, which results in $\Sigma_{z_2}^{n+1} \le \Sigma_{z_2}^n$ and subsequently $\tilde{\Sigma}_{z_2}^{n+1} \le \tilde{\Sigma}_{z_2}^n$. This is while $\tilde{\Sigma}_{z_1}^{n+1} = \tilde{\Sigma}_{z_1}^n$. Hence, at epoch $n+1$ we will have $\tilde{\Sigma}_{z_1}^{n+1} > L^2\tilde{\Sigma}_{z_2}^{n+1}$ which means $z_{DOL}^{n+1}(s^{n+1}) = z_1$, i.e., the DOL allocation remains at dose z_1 almost surely if not sampled. Also, in case $\tilde{\Sigma}_{z_1}^n < L^2\tilde{\Sigma}_{z_2}^n$, we have $z_{DOL}^n(s^n) = z_2$. In this case, given sampling dose z_1, the variance of $s^{n+1} = \eta(s^n,z_1,\omega_y^{n+1})$ is as $\Sigma_{z_1}^{n+1} = \big[(\Sigma_{z_1}^n)^{-1} + \sigma_{z_1}^{-2}\big]^{-1}$ and $\Sigma_{z_2}^{n+1} = \Sigma_{z_2}^n$, which results in $\Sigma_{z_1}^{n+1} \le \Sigma_{z_1}^n$ and subsequently $\tilde{\Sigma}_{z_1}^{n+1} \le \tilde{\Sigma}_{z_1}^n$. This is while $\tilde{\Sigma}_{z_2}^{n+1} = \tilde{\Sigma}_{z_2}^n$. Hence, at epoch $n+1$ we will have $\tilde{\Sigma}_{z_1}^{n+1} < L^2\tilde{\Sigma}_{z_2}^{n+1}$ which means $z_{DOL}^{n+1}(s^{n+1}) = z_2$, i.e., the DOL allocation remains at dose z_2 almost surely if not sampled. Therefore, in general for any arbitrary $s^n = s$ and $z \ne z_{DOL}^n(s)$, we have $z_{DOL}^{n+1}\big(\eta(s,z,\omega_y^{n+1})\big) = z_{DOL}^n(s)$, i.e., DOL persistence property holds. Hence, DOL is optimal for the impersonalized dose-finding trial with two doses.

A.7. Proof of Corollary 3

Recall that we denoted n_z as the number of samples that have been allocated to dose z until epoch n. Without any loss of generality assume that we are starting the trial with a non-informative initial prior for both alternative doses, i.e., $\Sigma^0_{z_1} = \Sigma^0_{z_2} = +\infty$. Therefore, we have for the variance of each dose at epoch n, $\Sigma^n_{z_1} = \dfrac{\sigma^2_{z_1}}{n_{z_1}}$ and $\Sigma^n_{z_2} = \dfrac{\sigma^2_{z_2}}{n_{z_2}}$. Based on Proposition 3 the allocation rule of DOL for the impersonalized dose-finding problem with two alternatives doses as $\mathcal{Z} = \{z_1, z_2\}$ with $z_1 < z_2$ given independent posterior beliefs about the alternative doses is

$$z^n_{DOL}(s^n) = \begin{cases} z_1 & \text{if} \quad \tilde{\Sigma}^n_{z_1} > L^2 \tilde{\Sigma}^n_{z_2} \implies \Sigma^n_{z_1} - \big[(\Sigma^n_{z_1})^{-1} + \sigma^{-2}_{z_1}\big]^{-1} > L^2 \Big\{ \Sigma^n_{z_2} - \big[(\Sigma^n_{z_2})^{-1} + \sigma^{-2}_{z_2}\big]^{-1} \Big\}, \\ z_2 & \text{if} \quad \tilde{\Sigma}^n_{z_1} < L^2 \tilde{\Sigma}^n_{z_2} \implies \Sigma^n_{z_1} - \big[(\Sigma^n_{z_1})^{-1} + \sigma^{-2}_{z_1}\big]^{-1} < L^2 \Big\{ \Sigma^n_{z_2} - \big[(\Sigma^n_{z_2})^{-1} + \sigma^{-2}_{z_2}\big]^{-1} \Big\}. \end{cases}$$

Given the above allocation rule, we define two functions $v(\xi) := \dfrac{\sigma^2_{z_1}}{\xi} - \big[\dfrac{\xi}{\sigma^2_{z_1}} + \dfrac{1}{\sigma^2_{z_1}}\big]^{-1}$ and $\bar{v}^\kappa := L^2 \Big\{ \dfrac{\sigma^2_{z_2}}{\kappa} - \big[\dfrac{\kappa}{\sigma^2_{z_2}} + \dfrac{1}{\sigma^2_{z_2}}\big]^{-1} \Big\}$ corresponding to two sides of the given inequality. Note that both of these functions are decreasing with respect to ξ and κ. Then, the statement $v(\xi) = \bar{v}^\kappa$ means that by the time we take ξ samples from dose z_1, we have also taken κ samples from dose z_2. Thus

$$v(\xi) = \bar{v}^\kappa \implies \frac{\sigma^2_{z_1}}{\xi} - \big[\frac{\xi}{\sigma^2_{z_1}} + \frac{1}{\sigma^2_{z_1}}\big]^{-1} = L^2 \Big\{ \frac{\sigma^2_{z_2}}{\kappa} - \big[\frac{\kappa}{\sigma^2_{z_2}} + \frac{1}{\sigma^2_{z_2}}\big]^{-1} \Big\} \implies \frac{\xi(\xi+1)}{\kappa(\kappa+1)} = \frac{1}{L^2}\Big(\frac{\sigma^2_{z_1}}{\sigma^2_{z_2}}\Big),$$

which implies that $\dfrac{n_{z_1}}{n_{z_2}} \to \dfrac{1}{L^2}\Big(\dfrac{\sigma_{z_1}}{\sigma_{z_2}}\Big)$ as the total number of samples N goes to infinity.

Appendix B: Algorithms

B.1. DOL Algorithms

Algorithm 1 DOL allocation procedure given correlated beliefs about the alternative doses

Given $s^n = (\mu^n, \Sigma^n)$ and $x^n = (x_1^n, x_2^n, \ldots, x_K^n) \sim \mathcal{P}_x$ at epoch $n \in \{0, \ldots, N-1\}$,

for each dose $z \in \mathcal{Z}$ **do**

  for each $m = 1 : M$ **do**

    Sample $\hat{\Theta}^{(m)} \sim \mathcal{N}(\mu^n, \Sigma^n)$.

    Sample $y_z^{(m)} | \hat{\Theta}^{(m)} \sim \mathcal{N}\big(\langle \delta_z^{x^n}, \hat{\Theta}^{(m)} \rangle, \sigma_z^2\big)$.

    Using x^n and $y_z^{(m)}$, obtain $(\mu^{n+1(m)}, \Sigma^{n+1(m)})$ according to updating equation (3), generating filtration $\mathcal{F}_z^{n+1(m)}$ by the set $\{\mu^0, \Sigma^0, x^0, z^0, y^1, \ldots, x^n, z, y_z^{(m)}\}$.

    for each $c = 1 : C$ **do**

      Sample $\hat{\Theta}^{(m,c)} \big| \mathcal{F}_z^{n+1(m)} \sim \mathcal{N}(\mu^{n+1(m)}, \Sigma^{n+1(m)})$.

      for each $x \in \mathcal{X}$ **do**

        Find the sample target dose $\mathrm{ED}_L^{x\,(m,c)} \big| \mathcal{F}_z^{n+1(m)} = g\big(\hat{\Theta}^{(m,c)} \big| \mathcal{F}_z^{n+1(m)}, x\big)$.

      end for

    end for

    for each $x \in \mathcal{X}$ **do**

      Evaluate the sample variance $U_{z,x}^{(m)} = \mathrm{Var}\big(\mathrm{ED}_L^{x\,(m,c)} \big| \mathcal{F}_z^{n+1(m)}\big)$ over C replications.

    end for

    Evaluate $U_z^{(m)} = \sum_x \mathcal{P}_x U_{z,x}^{(m)}$.

  end for

  Evaluate $U_z = \sum_m \dfrac{U_z^{(m)}}{M}$.

end for

Allocate sampling to the dose $z_{DOL}^n = \arg\min_{z \in \mathcal{Z}} U_z$ and obtain belief s^{n+1} using equation (3).

Algorithm 2 DOL allocation procedure given independent beliefs about the alternative doses

Given $s^n = (\mu^n, \Sigma^n)$ and $x^n = (x_1^n, x_2^n, \ldots, x_K^n) \sim \mathcal{P}_x$ at epoch $n \in \{0, \ldots, N-1\}$,

for each dose $z \in \mathcal{Z}$ **do**

  for each $m = 1 : M$ **do**

    Sample $\hat{\Theta}_z^{(m)} \sim \mathcal{N}(\mu_z^n, \Sigma_z^n)$.

    Sample $y_z^{(m)} | \hat{\Theta}_z^{(m)} \sim \mathcal{N}(x^n \hat{\Theta}_z^{(m)}, \sigma_z^2)$.

    Using x^n and $y_z^{(m)}$, obtain $(\mu_z^{n+1^{(m)}}, \Sigma_z^{n+1^{(m)}})$ according to updating equation (8), generating filtration $\mathcal{F}_z^{n+1^{(m)}}$ by the set $\{\mu^0, \Sigma^0, x^0, z^0, y^1, \ldots, x^n, z, y_z^{(m)}\}$.

    for each $c = 1 : C$ **do**

      Create $\hat{\Theta}^{(m,c)} \big| \mathcal{F}_z^{n+1^{(m)}} = [\hat{\Theta}_z^{(m,c)} : z \in \mathcal{Z}]$ by sampling $\hat{\Theta}_z^{(m,c)} \sim \mathcal{N}(\mu_z^{n+1^{(m)}}, \Sigma_z^{n+1^{(m)}})$ for each $z \in \mathcal{Z}$.

      for each $x \in \mathcal{X}$ **do**

        Find the sample target dose $\mathrm{ED}_L^{x\,(m,c)} \big| \mathcal{F}_z^{n+1^{(m)}} = g\big(\hat{\Theta}^{(m,c)} \big| \mathcal{F}_z^{n+1^{(m)}}, x\big)$.

      end for

    end for

    for each $x \in \mathcal{X}$ **do**

      Evaluate the sample variance $U_{z,x}^{(m)} = \mathrm{Var}\big(\mathrm{ED}_L^{x\,(m,c)} \big| \mathcal{F}_z^{n+1^{(m)}}\big)$ over C replications.

    end for

    Evaluate $U_z^{(m)} = \sum_x \mathcal{P}_x U_{z,x}^{(m)}$.

  end for

  Evaluate $U_z = \sum_m \dfrac{U_z^{(m)}}{M}$.

end for

Allocate sampling to the dose $z_{DOL}^n = \arg\min_{z \in \mathcal{Z}} U_z$ and obtain belief s^{n+1} using equation (3).

B.2. PAS Algorithms

Algorithm 3 PAS allocation procedure given correlated beliefs about the alternative doses

Given $s^n = (\mu^n, \Sigma^n)$ and $x^n = (x_1^n, x_2^n, \ldots, x_K^n) \sim \mathcal{P}_x$ at epoch $n \in \{0, \ldots, N-1\}$,

for each dose $z \in \mathcal{Z}$ **do**

Sample $\hat{y}_z^n \sim \mathcal{N}\Big(\big\langle \delta_z^{x^n}, \mu^n \big\rangle, [\delta_z^{x^n}]^\top \Sigma^n \delta_z^{x^n}\Big)$.

end for

Set $\hat{y}^n = [\hat{y}_z^n : z \in \mathcal{Z}]$ and find the sample target dose $\mathrm{ED}_L^{x^n} \big| \hat{y}^n$.

Allocate sampling to the dose $z_{PAS}^n = \mathrm{ED}_L^{x^n} \big| \hat{y}^n$ and obtain belief s^{n+1} using equation (3).

Algorithm 4 PAS allocation procedure given independent beliefs about the alternative doses

Given $s^n = (\mu^n, \Sigma^n)$ and $x^n = (x_1^n, x_2^n, \ldots, x_K^n) \sim \mathcal{P}_x$ at epoch $n \in \{0, \ldots, N-1\}$,

for each dose $z \in \mathcal{Z}$ **do**

Sample $\hat{y}_z^n \sim \mathcal{N}(x^n \mu_z^n, x^n \Sigma_z^n (x^n)^\top)$.

end for

Set $\hat{y}^n = [\hat{y}_z^n : z \in \mathcal{Z}]$ and find the sample target dose $\mathrm{ED}_L^{x^n} \big| \hat{y}^n$.

Allocate sampling to the dose $z_{PAS}^n = \mathrm{ED}_L^{x^n} \big| \hat{y}^n$ and obtain belief s^{n+1} using equation (3).

B.3. PPAS Algorithms

Algorithm 5 PPAS allocation procedure given correlated beliefs about the alternative doses

Given $s^n = (\mu^n, \Sigma^n)$ and $x^n = (x_1^n, x_2^n, \ldots, x_K^n) \sim \mathcal{P}_x$ at epoch $n \in \{0, \ldots, N-1\}$,

for each dose $z \in \mathcal{Z}$ **do**

Sample $\tilde{y}_z^n \sim \mathcal{N}\Big(\big\langle \delta_z^{x^n}, \mu^n \big\rangle, [\delta_z^{x^n}]^\top \Sigma^n \delta_z^{x^n} + \sigma_z^2\Big)$.

end for

Set $\tilde{y}^n = [\tilde{y}_z^n : z \in \mathcal{Z}]$ and find the sample target dose $\mathrm{ED}_L^{x^n} \big| \tilde{y}^n$.

Allocate sampling to the dose $z_{PPAS}^n = \mathrm{ED}_L^{x^n} \big| \tilde{y}^n$ and obtain belief s^{n+1} using equation (3).

Algorithm 6 PPAS allocation procedure given independent beliefs about the alternative doses

Given $s^n = (\mu^n, \Sigma^n)$ and $x^n = (x_1^n, x_2^n, \ldots, x_K^n) \sim \mathcal{P}_x$ at epoch $n \in \{0, \ldots, N-1\}$,

for each dose $z \in \mathcal{Z}$ **do**

Sample $\tilde{y}_z^n \sim \mathcal{N}(x^n \mu_z^n, x^n \Sigma_z^n (x^n)^\top + \sigma_z^2)$.

end for

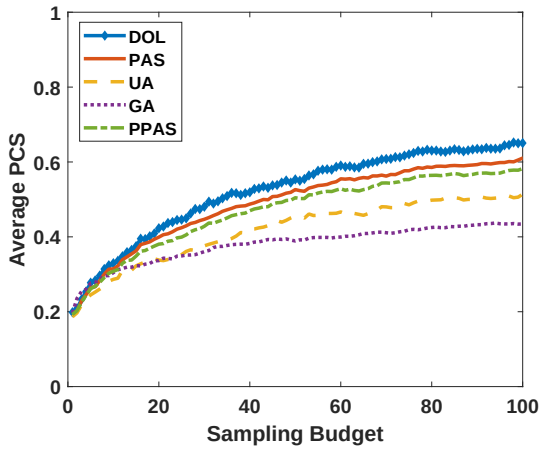
Set $\tilde{y}^n = [\tilde{y}_z^n : z \in \mathcal{Z}]$ and find the sample target dose $\mathrm{ED}_L^{x^n} \big| \tilde{y}^n$.

Allocate sampling to the dose $z_{PPAS}^n = \mathrm{ED}_L^{x^n} \big| \tilde{y}^n$ and obtain belief s^{n+1} using equation (3).

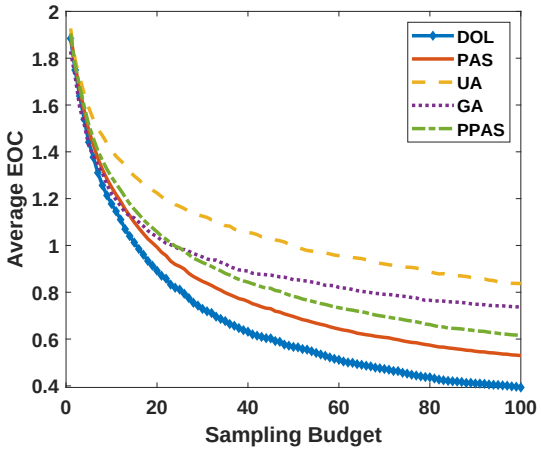
Appendix C: Additional Numerical Results

C.1. Performance of DOL vs the benchmarks in the base experiment of Section 7

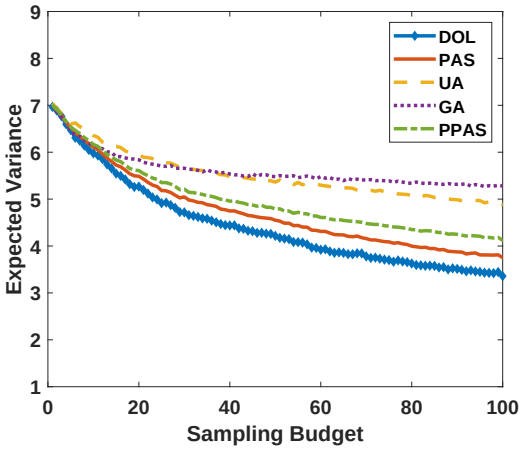
Below are the PCS, EOC, EVar and average allocation plots for the first (base) experiment in Section 7 presented in Figure 6. As we can see, PCS and EOC plots are consistent with EVar results, i.e., DOL outperforms other benchmarks in terms of PCS and EOC as well. Also, in plot 6-d we can see the exploration-exploitation behavior of DOL in comparison to other the benchmark policies in which DOL turned out to be more explorative than the posterior adaptive policies.



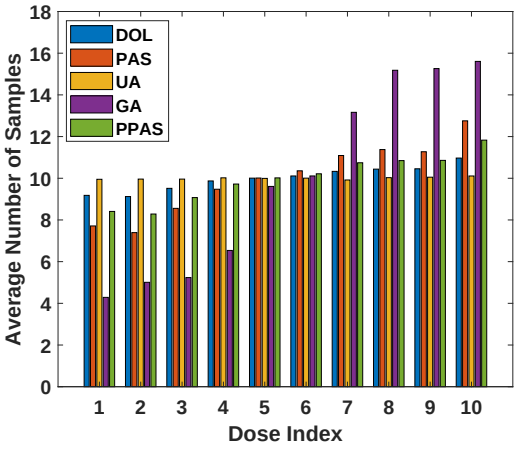
(a) Average PCS over both patient types



(b) Average EOC over both patient types



(c) EVar of the target dose over both patient types

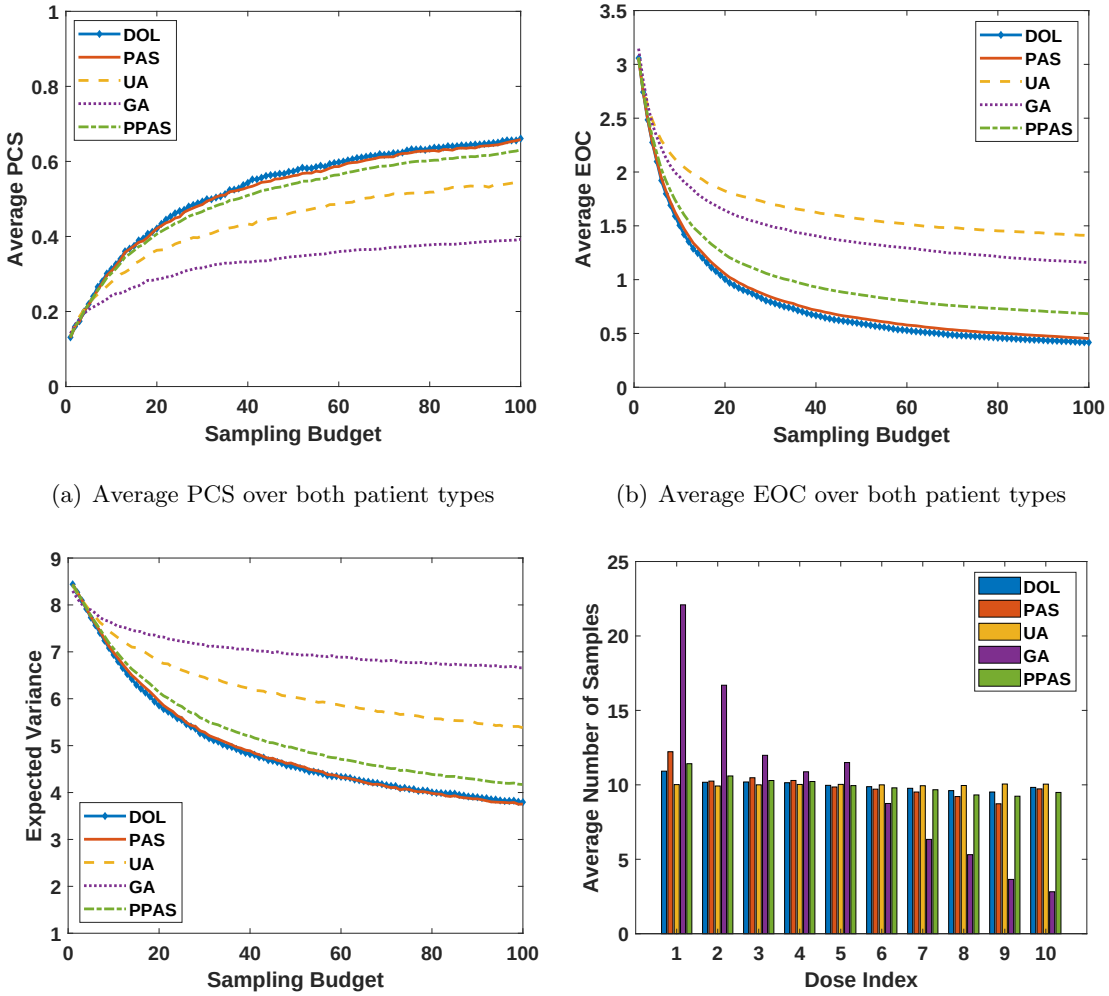


(d) Average number of allocations to each dose

Figure 6 Performance measures averaged over both patient types for the base experiment with equal sampling standard deviations of $\sigma_z = 3$ over all doses $z \in \mathcal{Z}$ considering fully correlated increasing (enthusiastic) prior belief.

C.2. Sensitivity analysis with respect to the prior

Here, we repeat the experiment in Section 7 with flat (reference/non-informative) and skeptical priors. Note that in the skeptical prior we assume a decreasing initial dose-response belief. The results are presented in Figures 7 and 8 respectively. We can see in these scenarios that DOL still performs competitively in terms of different performance measures, but the performance of PAS is closer to DOL compared to the results we had for the experiment with the enthusiastic prior.



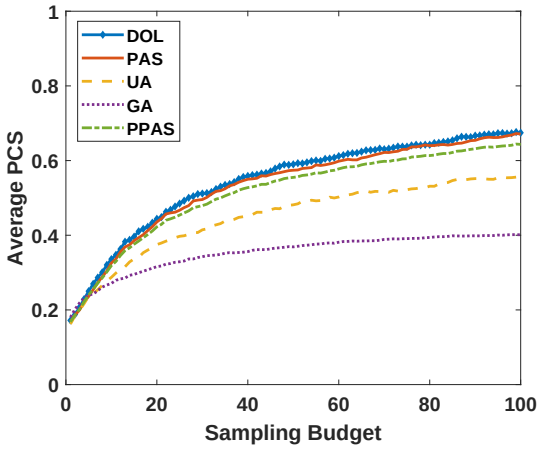
(a) Average PCS over both patient types

(b) Average EOC over both patient types

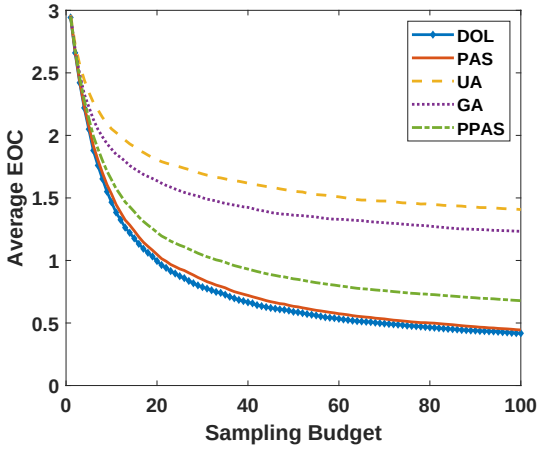
(c) EVar of the target dose over both patient types

(d) Average number of allocations to each dose

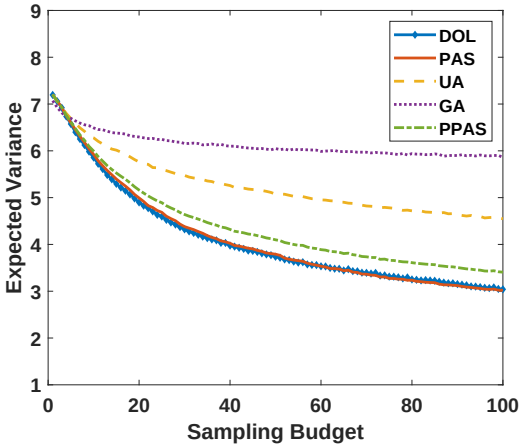
Figure 7 Performance measures on average over both patient types for the experiment with equal sampling standard deviations of $\sigma_z = 3$ over all doses $z \in \mathcal{Z}$ considering flat (non-informative) prior.



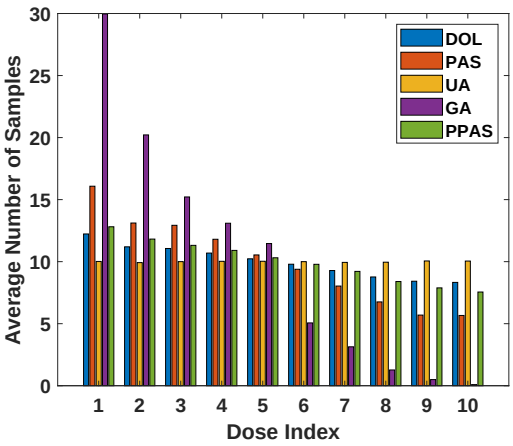
(a) Average PCS over both patient types



(b) Average EOC over both patient types



(c) EVar of the target dose over both patient types

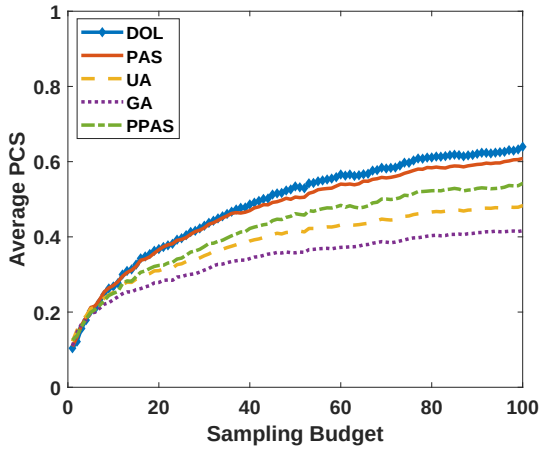


(d) Average number of allocations to each dose

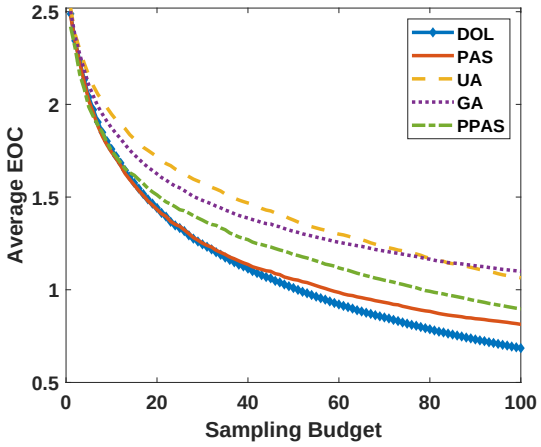
Figure 8 Performance measures on average over both patient types for the experiment with equal sampling standard deviations of $\sigma_z = 3$ over all doses $z \in \mathcal{Z}$ considering decreasing (skeptical) prior belief.

C.3. Sensitivity analysis of misspecified prior

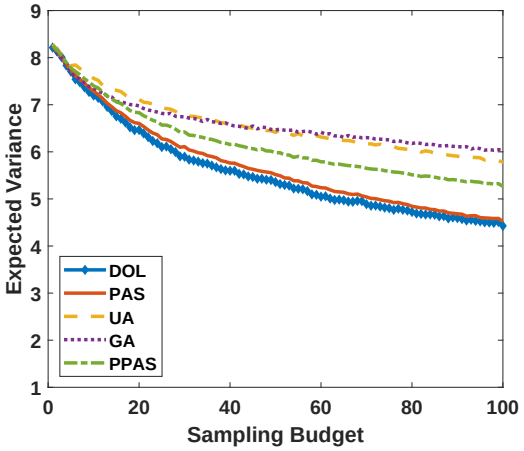
Here, we repeat the experiment in Section 7 with only one difference. Here, instead of starting the trial from an increasing prior mean response, we consider a flat prior mean response generated by assuming an initial prior mean vector $\mu^0 = [\mu_1^0, \mu_2^0, \ldots, \mu_Z^0]^\top$ with $\mu_z^0 = [0.5, 0.5]^\top, \forall z \in \mathcal{Z}$. Note that we start from the exact same prior covariance matrix and use the same problem instances (true dose-response curves) generated for the experiment in Section 7 to control the experiment. We can see the drop in performance especially at the beginning of the trial due to the fact that the problem instances are generated from the increasing prior described in Section 7. However, after a few samples, the posterior belief quickly converges to the true parameter instances controlling the simulation, and the PCS at the end of the trial ($N = 100$) is close to the performance we got on average for the experiment in Section 7.



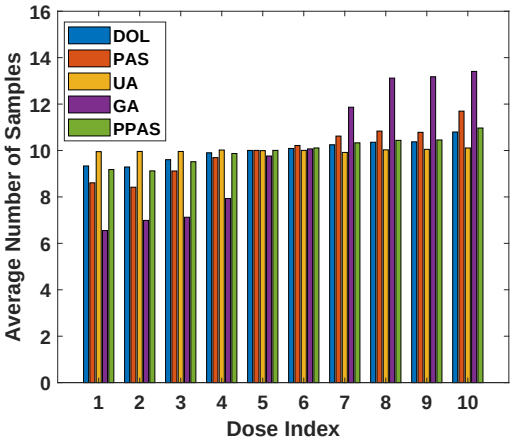
(a) Average PCS over both patient types



(b) Average EOC over both patient types



(c) EVar of the target dose over both patient types

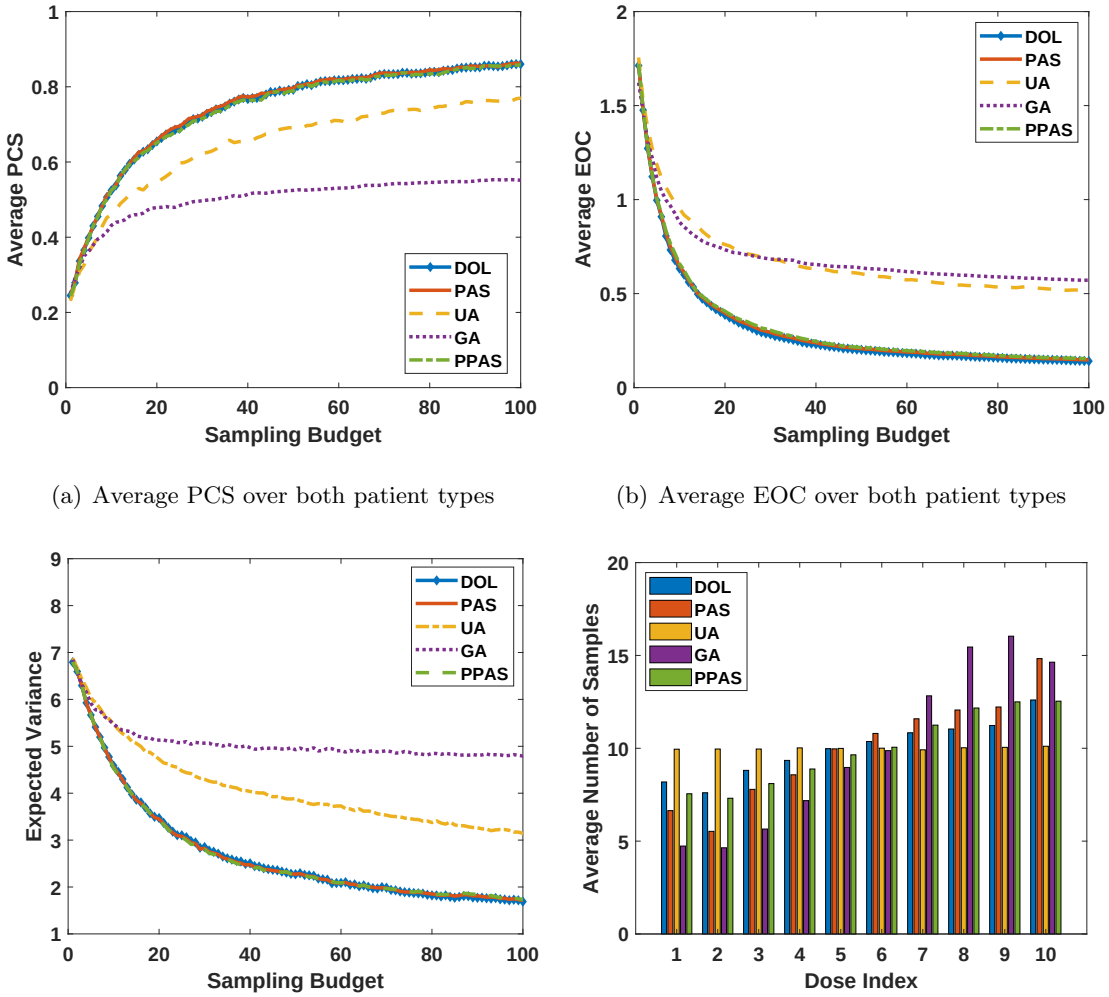


(d) Average number of allocations to each dose

Figure 9 Performance measures averaged over both patient types for the experiment with misspecified (flat) initial prior mean responses and increasing problem instances (true dose-response curves).

C.4. Sensitivity analysis with respect to sampling/measurement variances

For this experiment, we repeat the experiment in Section 7 with only one change: we change the amount of sampling standard deviations to $\sigma_z = 1$ and $\sigma_z = 5$ over all doses $z \in \mathcal{Z}$. Figures 10 and 11 respectively present the results. DOL performs competitively versus the benchmarks in high and low noise conditions. However, considering the higher computational cost, DOL seems to be a better choice over PAS and PPAS for trials with moderate sampling errors, where the sampling/measurement variances are neither too large nor too small.



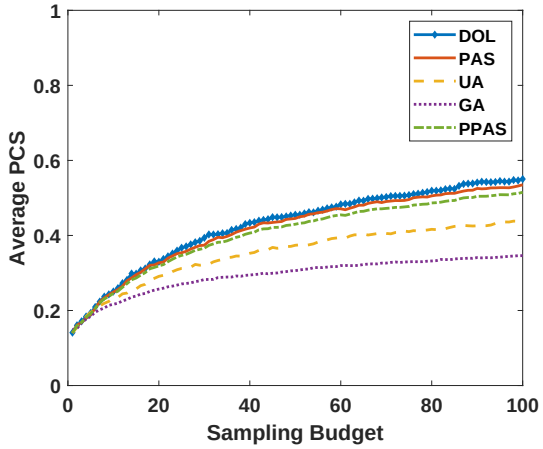
(a) Average PCS over both patient types

(b) Average EOC over both patient types

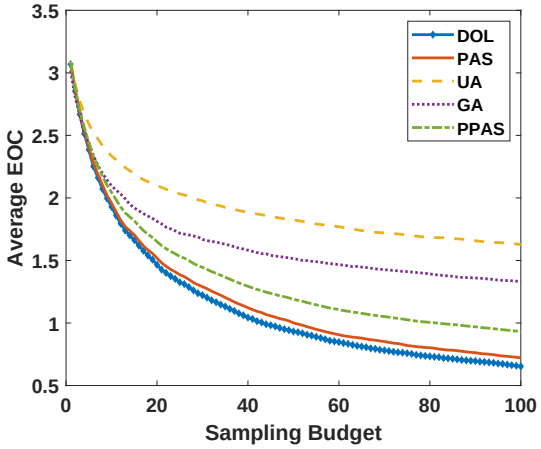
(c) EVar of the target dose over both patient types

(d) Average number of allocations to each dose

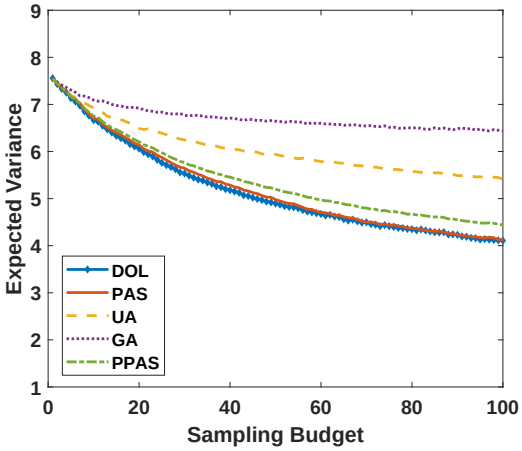
Figure 10 Performance measures on average over both patient types for the experiment with equal sampling standard deviations of $\sigma_z^2 = 1$ over all doses $z \in \mathcal{Z}$.



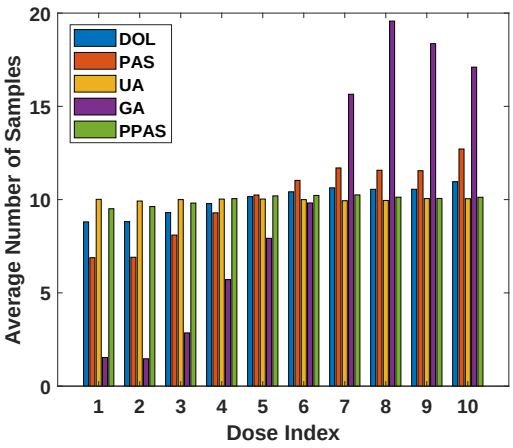
(a) Average PCS over both patient types



(b) Average EOC over both patient types



(c) EVar of the target dose over both patient types

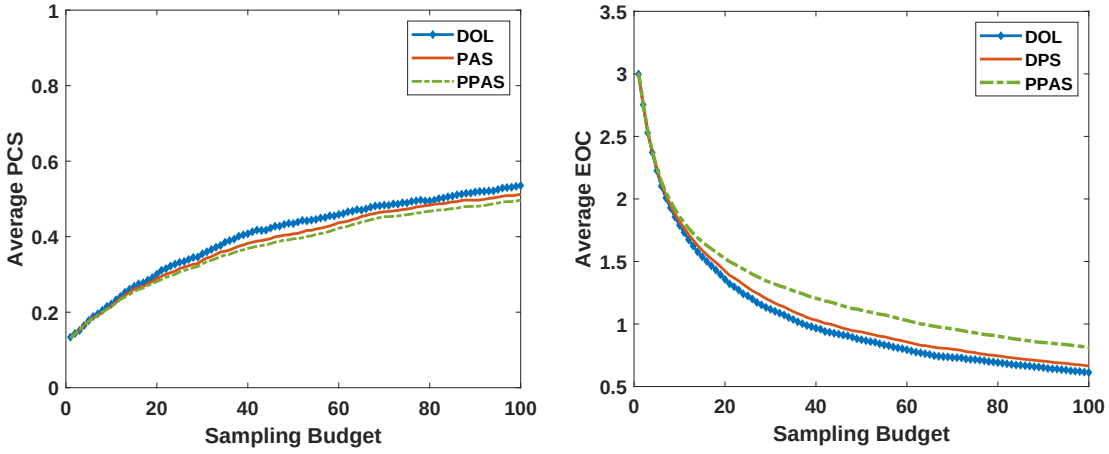


(d) Average number of allocations to each dose

Figure 11 Performance measures on average over both patient types for the experiment with equal sampling standard deviations of $\sigma_z = 5$ over all doses $z \in \mathcal{Z}$.

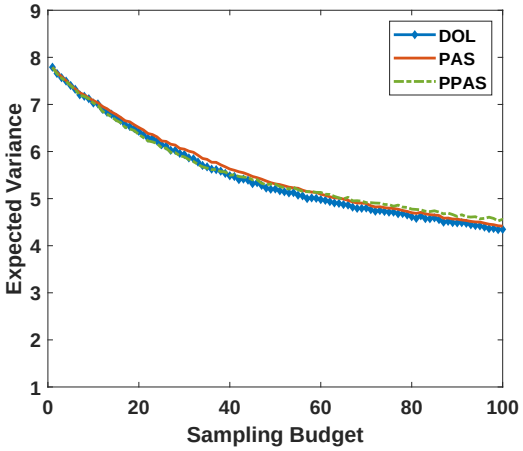
C.5. Sensitivity analysis with respect to correlation among doses

Here are the results of repeating the numerical experiment in Section 7 without considering the correlations among doses (Section 6) in the case where the environment is correlated (Section 4). To that end, we repeat the first experiment in this section for DOL, PAS, and PPAS by implementing a version of these policies that considers independent beliefs about the alternative doses and follows the updates mentioned in Section 6 and the corresponding algorithms in Section B.

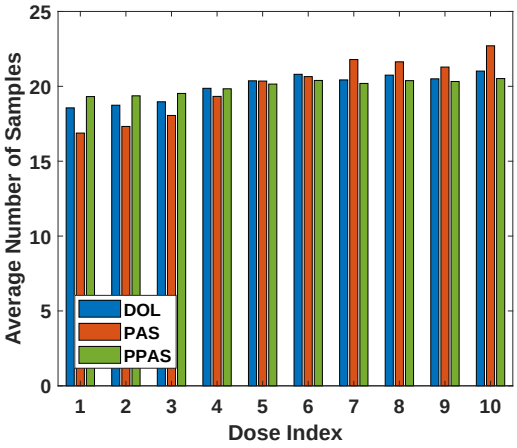


(a) Average PCS over both patient types

(b) Average EOC over both patient types



(c) EVar of the target dose over both patient types

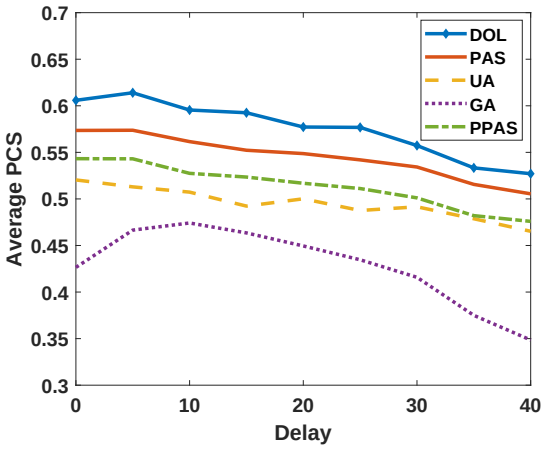


(d) Average number of allocations to each dose

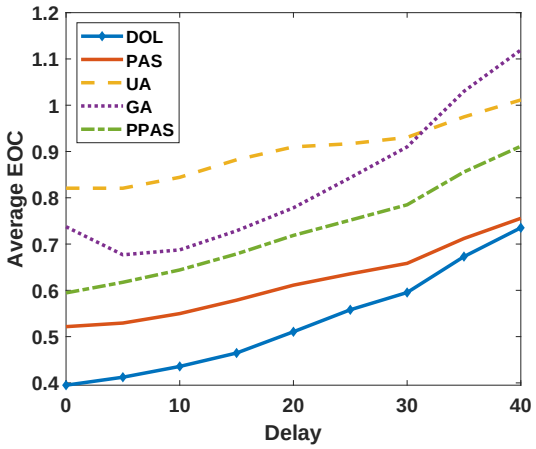
Figure 12 Performance of DOL, PAS, and PPAS policies, tailored not to consider correlation among doses.

C.6. Sensitivity analysis with respect to delay in observing the responses

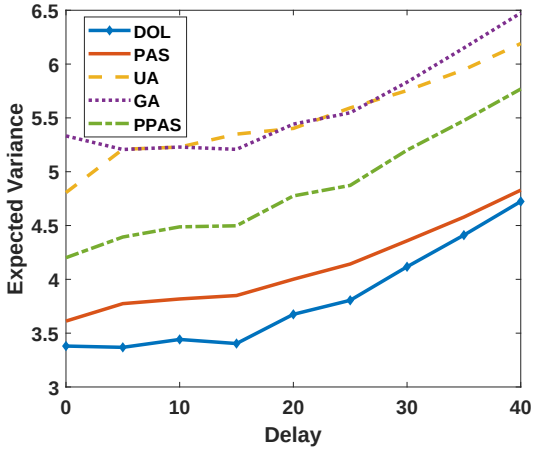
Here are the results of repeating the numerical experiment in Section 7 with a given delay in observing patients responses. To that end, we assume that the delay τ is a discrete number measured in unit of samples, i.e., we assume that the allocation decision at epoch $n \geq \tau$ should be made given all the information available at epoch $n-\tau$, i.e., the decision rule d^n is the function of $\mathcal{F}^{n-\tau}$. In this case, the sampling decisions for the first τ samples have to be made by just considering the initial prior belief $s^0 = (\mu^0, \Sigma^0)$ about the unknown parameter Θ. In order to see the effect of this delay, we repeat our numerical experiment of Section 7 for delays of $\tau = 0, 5, 10, \ldots, 40$ and plot the performance measures at the end of the trial ($N = 100$) given the delay τ. We can see that the drop in performance is not monotone for either of the policies, especially for GA policy for small delays. This is mainly because GA is inherently very exploitative (unlike UA which is the most explorative). Small delays have the most influence on the sampling allocations of exploitative policies.



(a) Average PCS over both patient types



(b) Average EOC over both patient types

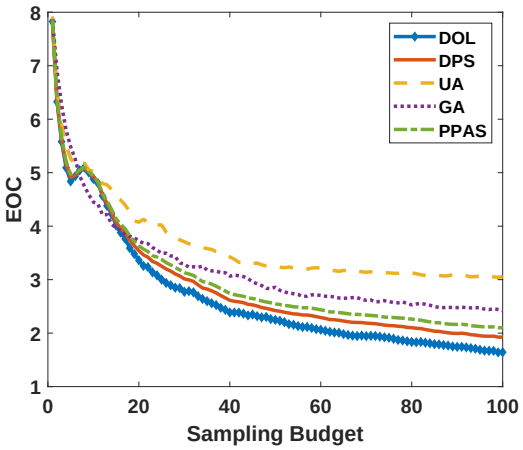


(c) EVar of the target dose over both patient types

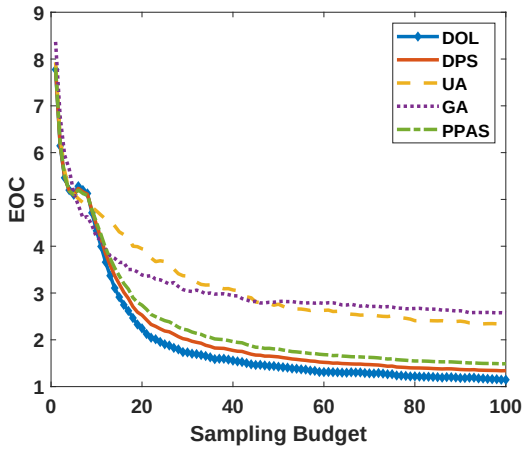
Figure 13 **Performance measures on average over both patient types at the end of the trial ($N = 100$) as a function of delay (in terms of number of samples).**

C.7. Additional plots for the real case study of Warfarin in Section 8

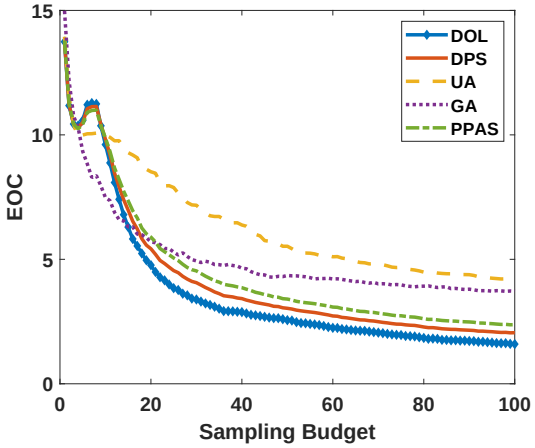
Figure 14 represents the EOC plots while Figure 15 shows the PCS and EVar plots across different patient types with different alleles for the real case study of Warfarin in Section 8. As can be seen, the EOC plots in Figure 14 concur with PCS and EVar plots in Figure 15 as DOL outperforms other benchmarks in each patient type as well as on average. Another interesting observation is how the performance measures for patients with AA allele is not as good as other two patient types. The main reason for this behavior is the fact that the true dose-response curve for patients with AA allele is much more flat (with less variation over the dose space as presented in Figure 4-e) compared to the other two patient types, which makes it harder under any allocation policy to correctly learn its target dose.



(a) EOC for AA allele

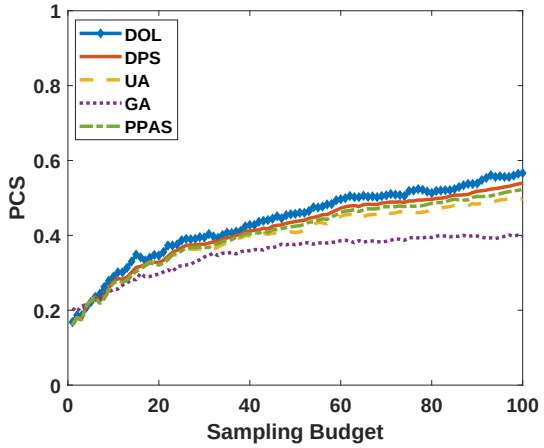


(b) EOC for AG allele

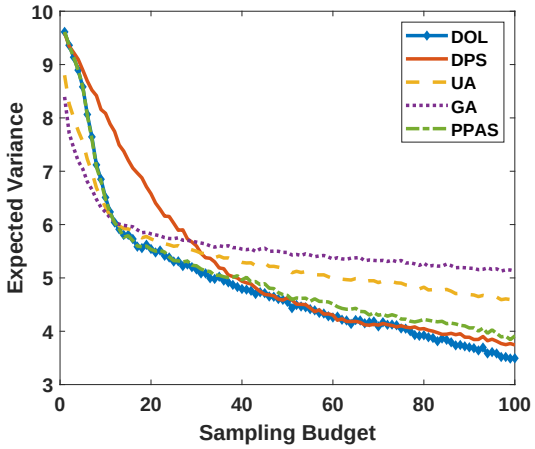


(c) EOC for GG allele

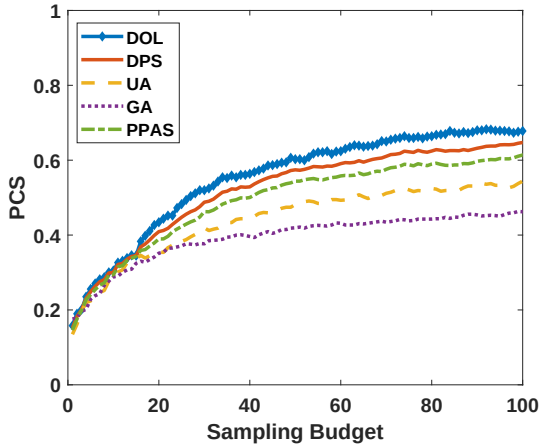
Figure 14 Expected opportunity cost across patients with different VKORC1 alleles for a real case study of Warfarin



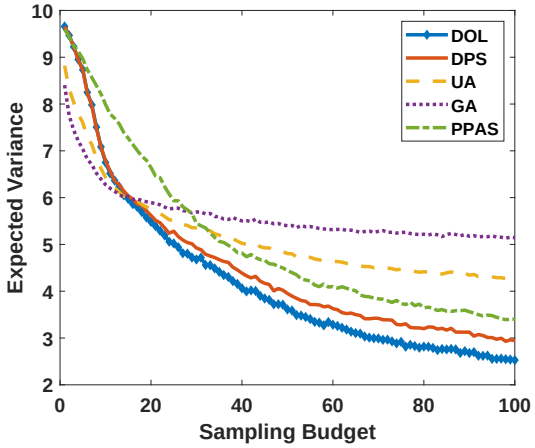
(a) PCS for AA allele



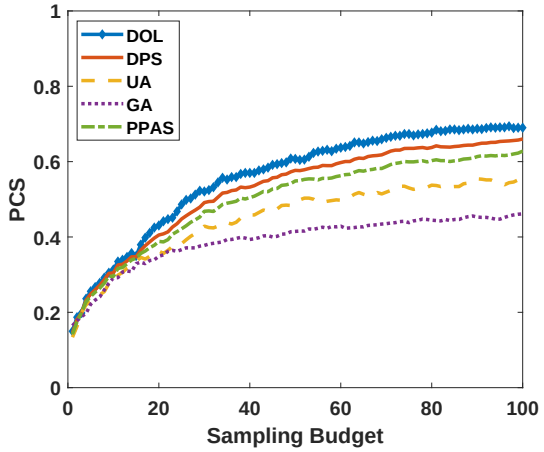
(b) EVar of the target dose for AA allele



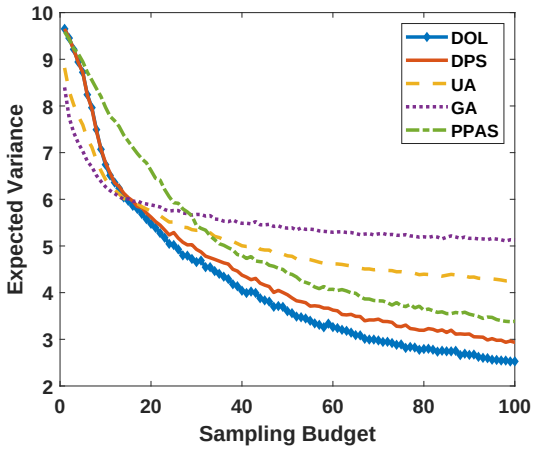
(c) PCS for AG allele



(d) EVar of the target dose for AG allele



(e) PCS for GG allele

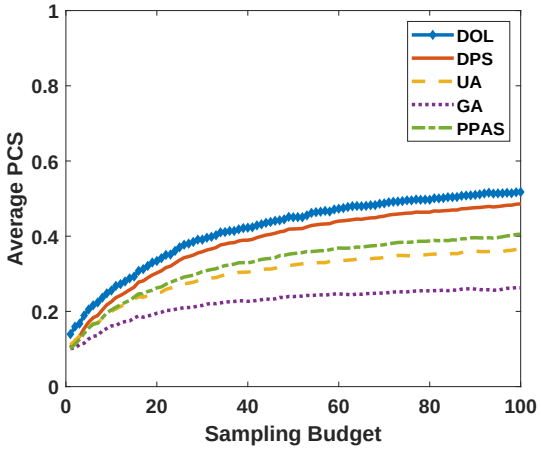


(f) EVar of the target dose for GG allele

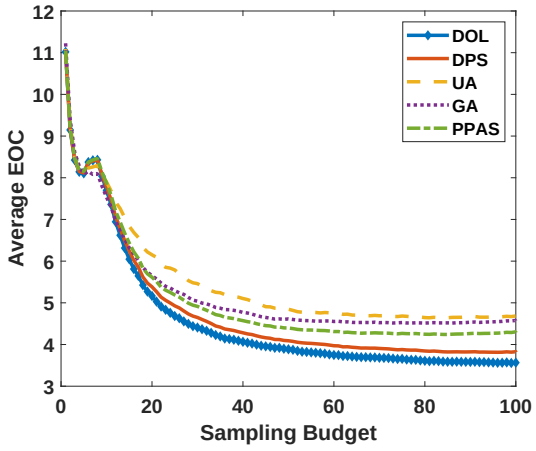
Figure 15 Probability of correct selection and variance of the target dose across patients with different VKORC1 alleles for a real case study of Warfarin

C.8. Value of Personalization in Dose-finding Trials

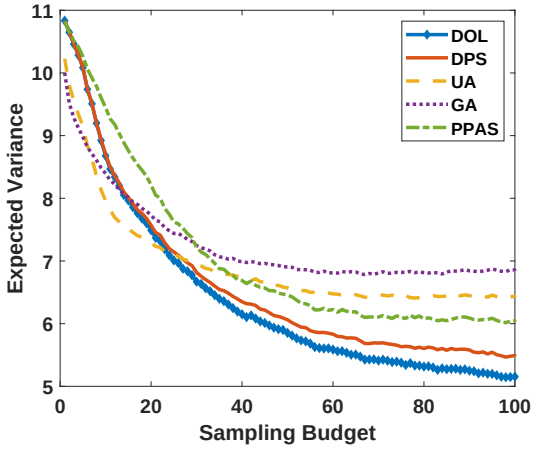
Here, we present the results of an experiment for the real case study of Warfarin where we assume that the patients are homogeneous, and hence, we consider an impersonalized model to find a single $\text{ED}_{0.95}$ for all the patients. Note that we are using the true dose-response curves presented in the figure of Section 8 to control the simulation, and the patients randomly enter the trial similar to the previous experiment. The only difference is that we use the responses to update a singular belief over all doses, i.e., we consider the unknown parameter Θ of the model to be a $Z \times 1$ vector and simulate the results assuming flat prior mean vector of all zeros similar to the experiment in Section 8.



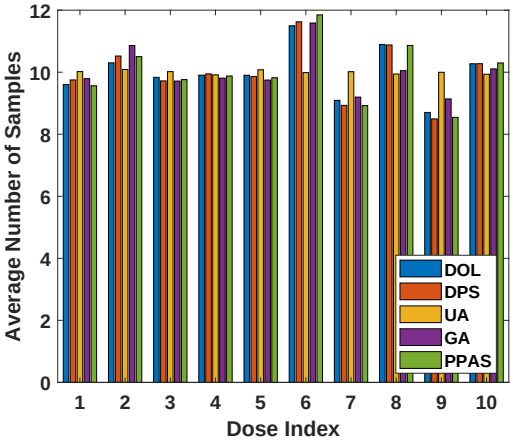
(a) Average PCS over three patient types



(b) Average EOC over three patient types



(c) EVar of the target dose over three patient types

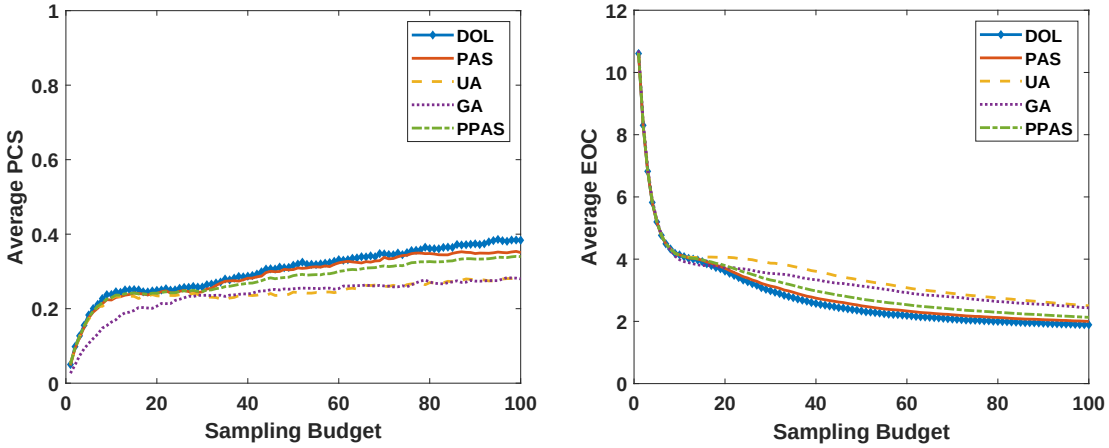


(d) Average number of allocations to each dose

Figure 16 Performance measures on average over all patient types for the real case study of Warfarin assuming homogeneous patients.

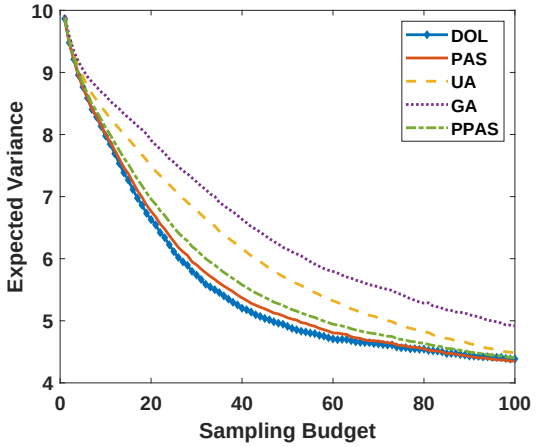
C.9. Value of Considering Correlations among Covariates

Here, we present the results of the last experiment for the real case study of Warfarin where we assume that the patients have different personalized target doses; however, we consider three separate impersonalized learning models to find three different target doses, where the information collected for one patient type is not shared with others. Here, while the patients are assumed to have different target doses, we only use the responses of each group of patients to individually update the belief about their dose-response relation.

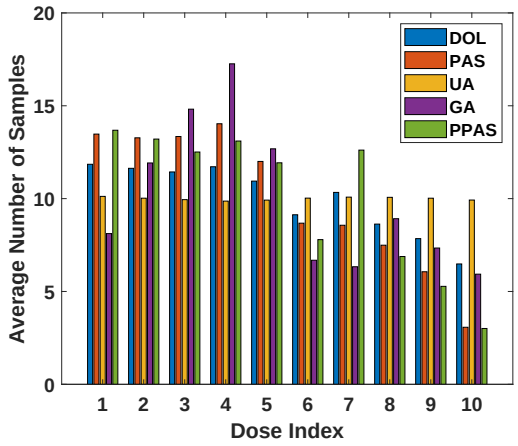


(a) Average PCS over three patient types

(b) Average EOC over three patient types



(c) EVar of the target dose over three patient types



(d) Average number of allocations to each dose

Figure 17 Performance measures on average over all patient types for the real case study of Warfarin assuming non-related patient types (independent covariates).

References

Frazier PI, Powell WB, Dayanik S (2008) A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization* 47(5):2410–2439.