Journal of Chemometrics

WILEY

Journal of CHEMOMETRICS

RESEARCH ARTICLE OPEN ACCESS

A Non-Linear Model for Multiple Alcohol Intakes and Optimal Designs Strategies

Irene Mariñas-Collado[1] | Juan M. Rodríguez-Díaz[2] | M. Teresa Santos-Martín[2]

[1]Department of Statistics and Operations Research and Mathematics Didactics, University of Oviedo, Oviedo, Spain | [2]Institute of Fundamental Physics and Mathematics, Department of Statistics, University of Salamanca, Salamanca, Spain

Correspondence: Irene Mariñas-Collado (marinasirene@uniovi.es)

Received: 6 June 2024 | **Revised:** 22 July 2024 | **Accepted:** 5 August 2024

Funding: This research was partially supported by the Spanish Ministry of Science, Innovation and Universities project PID2021-125211OB-I00 (all authors) and by the Junta de Castilla y León project 'SA217P23' (J.M.R.-D. and M.T.S.-M.).

Keywords: alcohol clearance | covariance structure | D-optimality | ethanol pharmacokinetics | non-linear models | optimal design of experiments | Widmark equation

ABSTRACT

This study addresses the complex dynamics of alcohol elimination in the human body, very important in forensic and healthcare areas. Existing models often oversimplify with the assumption of linear elimination kinetics, limiting practical application. This study presents a novel non-linear model for estimating blood alcohol concentration after multiple intakes. Initially developed for two different alcohol incorporations, it can be straightforwardly extended to the case of more intakes. Emphasising the significance of accurate parameter estimation, the research underscores the importance of precise experimental design, utilising optimal experimental design (OED) methodologies. Sensitivity analysis of model coefficients and the determination of D-optimal designs, considering correlation structures among observations, reveal a strong linear relationship between support points. This relationship can be used to obtain nearly optimal designs that are highly efficient and much easier to compute.

1 | Introduction and Background

Understanding the dynamics of alcohol elimination in the human body has great interest for various fields such as forensics, healthcare and substance abuse management. Although there is extensive evidence indicating that ethanol pharmacokinetics exhibit non-linear behaviour, many researchers still rely on linear, zero-order kinetics to determine ethanol elimination in the human body [1], using the classic Widmark equation, which was first developed in the 1930s [2]. Over time, improvements to the Widmark equation have been made, and new scientific findings on the pharmacology of alcohol have been published. Various models, ranging from one- to three-compartment models, have been proposed, such as the Norberg model, with two distribution compartments and elimination by Michaelis-Menten kinetics [3] or the Wilkinson model, with one distribution compartment and elimination by Michaelis-Menten kinetics [4], for example. However, in everyday forensic practice, these advancements are frequently disregarded due to their complexity, making them challenging for non-experts to utilise them effectively. This oversight poses significant issues, particularly considering that individuals driving under the influence of alcohol often face severe penalties, including substantial fines, revoked driving privileges, and in some cases, arrest and imprisonment, producing legal, economic, and social repercussions. Consequently, alternative models should undoubtedly find their place in forensic practice and expert witnessing [5].

In Mariñas-Collado et al. [6], a non-linear model based on the Gamma function that adequately captures the different

Abbreviations: BAC, blood alcohol concentration; BrAC, breath alcohol concentration; FIM, Fisher Information Matrix

stages of the ethanol pharmacokinetic process in the human body, such as absorption, distribution, metabolism and elimination, without any constraints on the density function, was proposed. The rationale behind this model stemmed from the necessity of a non-linear trend, particularly a hill-shaped function, to effectively illustrate both the absorption and clearance phases observed in alcohol consumption. The new model can be seen as a streamlined adaptation of the Gamma model used in previous research [7] or as a modified form of the Arrhenius model. A fundamental aspect of the proposed model lies in its simplicity, contrasting with the intricacies often associated with compartmental models. By prioritising simplicity, it offers a more accessible yet comprehensive depiction of ethanol kinetics. On the other hand, one limitation of this model is that it assumes that the alcohol is consumed in a single instance, which is seldom the case. In fact, alcoholic beverages are usually consumed at various time intervals, making it necessary to develop a model capable of capturing these increments.

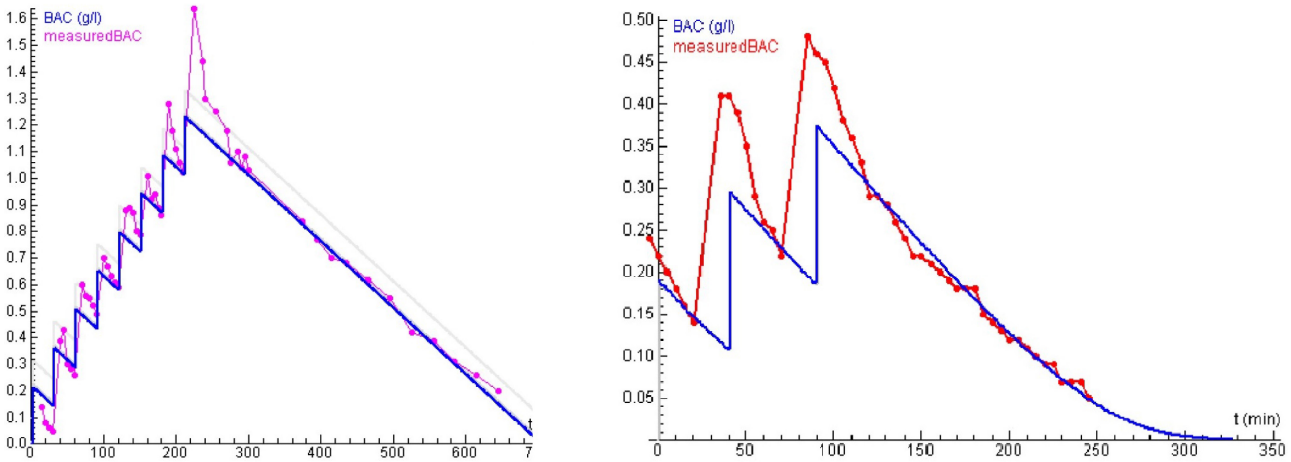
Although extensive literature exists regarding the various forms of modelling elimination processes and the different factors contributing to variability, such as gender, the effect of food, body size, genetic polymorphism and alcohol concentration on ethanol [8–10], most of these studies are centred solely on singular ethanol intake scenarios [11]. This prompted our investigation into multiple ethanol intake scenarios and the subsequent development of a corresponding model for blood alcohol concentration (BAC) within real-life settings.

When looking at literature on models considering multiple intakes, the usual assumption is instantaneous inputs (see, for instance, Heck et al. [12]), producing saw-shaped models (Figure 1). This results in the intake being modelled as a vertical straight line, which appears to diverge from physiological reality. In reality, the absorption of alcohol unfolds gradually. Controlled drinking experiments consistently demonstrate that the peak concentration usually emerges between 10 and 60 min following the conclusion of drinking [13], depicting a more intricate and less abrupt process that deviates from the assumed instantaneous incorporation.

In addition to the prevalent assumption of instantaneous inputs, another significant challenge arises from the methodology employed in experiments. Usually, experiments are conducted in order to find the model/parameters that best describe BAC. However, a common oversight lies in the lack of thorough consideration regarding how the selection of observational points might significantly influence the final model estimation. Consider, for illustrative purposes, the investigation conducted by Mumenthaler et al. [14], which highlights the ramifications of inadequate observation point selection in model construction.

The study focused on 27 healthy white women, from 20 to 40 years old, in good physical conditions, who had four drinks. Breath alcohol concentration (BrAC) was measured at different times, as shown in Figure 2. Their reasoning behind simplifying the model with instantaneous input was based on the requirement for subjects to consume specified quantities of alcohol within a 5-min timeframe. These quantities, represented as q, q, q, and 0.6 q ($q = 0.186$ g/L), were relatively small. However, the problem arose not from the assumption of instantaneous inputs, but from the observed time intervals. The mean BrAC-time (with ±SD) and the fitted curve are shown in Figure 3.

When performing the experiment, one would expect a estimated model similar to that of Figure 1b but having four peaks instead of three; however, Figure 3 appears far from



(a) Widmark model for regular consumption of 8 drinks.

(b) Wagner model for regular consumption of 3 standard units

FIGURE 1 | Screen shots from Heck et al. [12] depicting BAC for different number of intakes, representing both the modelled BAC (blue) and the measured BAC (pink/red).

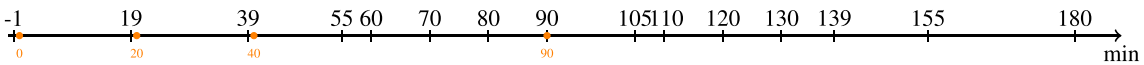


FIGURE 2 | Time points for sample collection (above the line) and alcohol intake events (below the line) in the study from Mumenthaler et al. [14].

this saw-shape. The reason of this difference relies on the design employed: Figure 4a shows the ideal mean BrAC model assuming four drinks and instantaneous inputs and displays as well the sampling point design and the intakes (described in Figure 2). Figure 4b shows that by measuring only at those selected time-points part of the information is left uncollected, losing the saw-shape. The interpolated BrAC (dashed red line) only shows two peaks, which aligns with the results obtained in Mumenthaler et al. [14]. This further highlights the pivotal role of optimal designs in obtaining accurate estimations. The inadequacy observed in capturing information at specific intervals emphasises the necessity for precise sampling strategies over an abundance of data points.

In this work, an extension of the simplified gamma model is presented, considering two intakes. Moreover, D-optimal designs are calculated. It is important to emphasise that the times shown in the optimal designs are those which, under experimental conditions, allow the best estimation of the model parameters. Once a reliable model is established through experimental (laboratory) testing, it can be employed in practice, where sampling times will often be dictated by specific circumstances, making it difficult to use optimal designs. The case of multiple (more than two) intakes will be addressed in the Conclusion section.

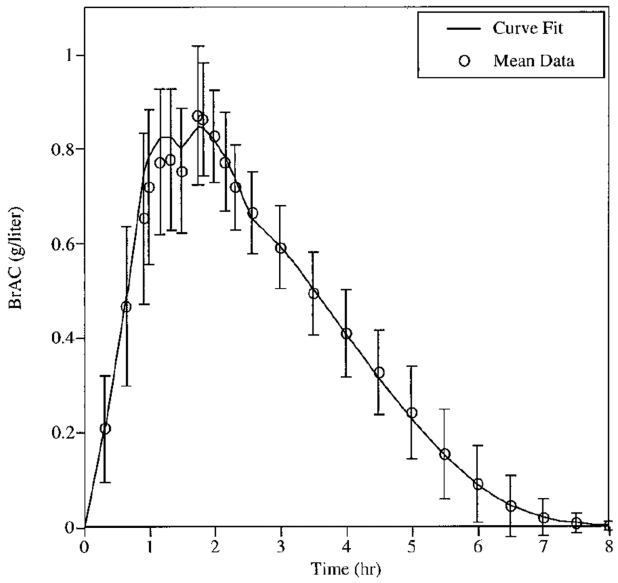


FIGURE 3 | Mean BrAC-time curve and curve fit to the mean BrAC data of 27 subjects [14].

2 | Initial Approaches and Model Definition

Starting from the simplified gamma model proposed to measure BAC concentracion at time t:

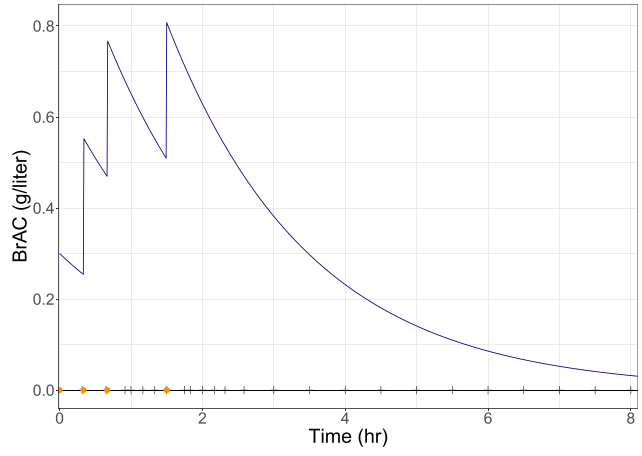
$$C_{SG}(t) = st^a e^{-bt}, \tag{1}$$

we extended our work to incorporate a second alcohol intake at time m (where time is measured in hours). The initial intuitive approach was to fit a piecewise-defined model:

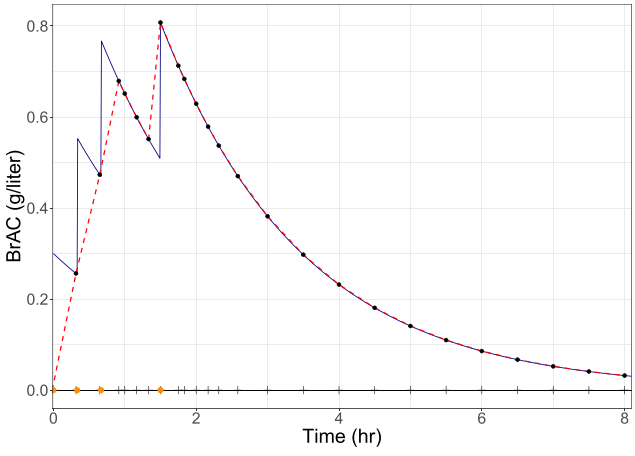
$$C_{SG_2*}(t) = \begin{cases} st^a e^{-bt}, & t \le m, \quad (2a) \\ s\left(t^a e^{-bt} + (t-m)^a e^{-b(t-m)}\right), & t > m. \quad (2b) \end{cases}$$

The challenge of dealing with this piecewise model, arising from multiple intakes whose effects accumulate from previous ones, lies in the lack of differentiability at the points where new incorporations occur. This lack of differentiability hampers the application of standard tools for computing and validating optimal designs. Therefore, it is convenient to approximate this piecewise model with an alternative that possesses similar properties but the advantage of being differentiable. Furthermore, this initial proposal still yields a relatively abrupt (instantaneous) rise, thus seeking ways to smooth the transition both before and after the secure incorporation aligns more closely with the objective of achieving a model that closely approximates the expected reality.

Several solutions were explored (see Figure 5). Initially, employing an Alpha function (ranging from 0 to 1) to blend the two functions around m, utilising a blending interval on each side of that point ($m \pm \mathcal{L}$), was considered. Although this approach improves the model, it falls short of achieving differentiability. Subsequently, including a logistic function to transition between the two function pieces was attempted. Here, adjusting the function to centre the transition near m involved the inclusion of a new parameter, the *lag*. However, a challenge arises as the logistic function does not reach



(a) Expected BrAC (blue solid line) with instantaneous inputs at minutes 0, 20, 40 and 90 (orange dots).



(b) Expected BrAC at observed times (black dots) and the interpolated BrAC (red dashed line).

FIGURE 4 | Expected and estimated BrAC from the design.

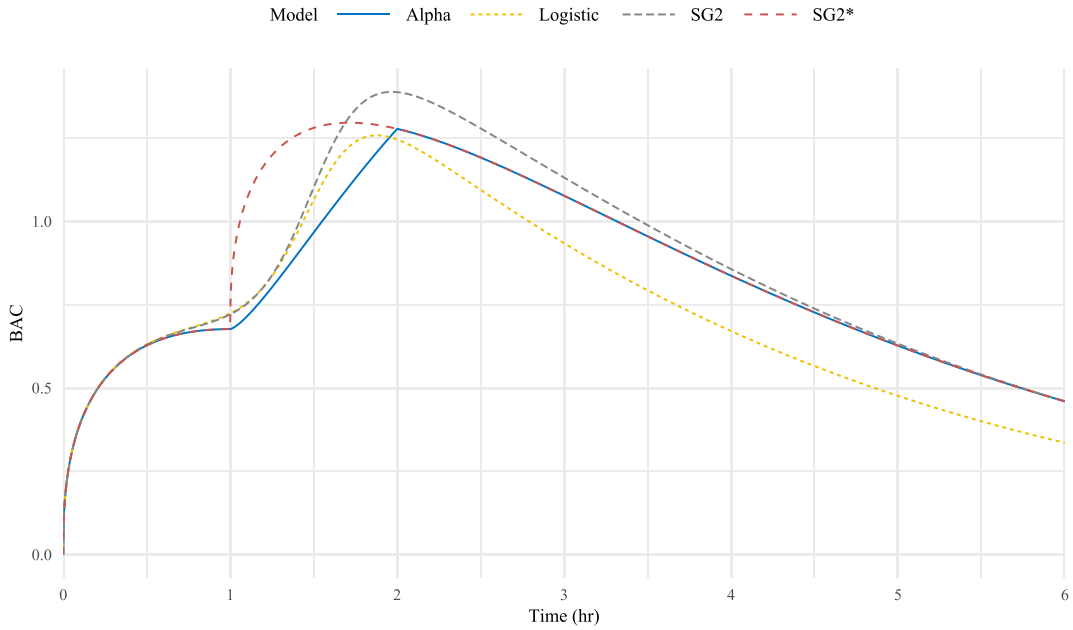


FIGURE 5 | Different approaches for the modification of C_{SG_2}.

complete zero for $t < m$. The final adjustment (*SG approx*) addresses this by accounting for the maximum time, ensuring the final BAC concentration remains consistent, regardless of the employed lag. In addition, this final approximation results in a more subtle and less abrupt incorporation, aligning more closely with the expected behaviour as previously discussed in the introduction.

Then, the final *approximated* model proposed is

$$C_{SG_2}(t) = st^a e^{-bt} + \frac{s}{1 + e^{-6(t-m-lag)/\mathcal{L}}} \left(\frac{t(1-m)}{t_{max}} \right)^a e^{-b(t-m)}, \quad (3)$$

where $2\mathcal{L}$ represents the total time span over which the transition from (2a) to (2b) takes place; *lag* is the parameter incorporated to centre the logistic function's transition point near $m + lag$ and t_{max} is the maximum time considered.

3 | Optimal Designs of Experiments

Let t denote the time points for observations within the experiment. An exact design, ξ, consists of a collection of time points, $\mathbf{t} = \{t_1, t_2, \ldots, t_n\}$, $t \in \mathcal{T}$, where samples are to be taken. $\mathcal{T}$ is called the design space.

The observations vector $\mathbf{Y} = (y_1, \ldots, y_n)$ reflects a one-response linear model $y = f(t; \boldsymbol{\theta})$, with $\boldsymbol{\theta}$ as the p-parameter vector, and $\mathbf{X} = (\mathbf{f}(t_1), \ldots, \mathbf{f}(t_n))^T$ as the design matrix, where $\mathbf{f}(t) = (f_1(t), \ldots, f_p(t))^T$. The Fisher Information Matrix (FIM), which quantifies the information that the data provides about the unknown parameters, can be expressed as $\mathbf{M}(\xi, \boldsymbol{\theta}) = \mathbf{X}^T\mathbf{X}$. If accounting for the correlation between samples is desired, then it can be computed as $\mathbf{M}_{\boldsymbol{\Sigma}}(\xi, \boldsymbol{\theta}) = \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}$, with $\boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{Y})$. The inverse of the FIM is proportionally linked to the asymptotic covariance matrix for the maximum likelihood estimate of $\boldsymbol{\theta}$.

For non-linear models, the conventional approach involves linearising them by computing parameter-related derivatives. The information matrix depends on the unknown parameters that appear 'non-linearly' in the model [15], so the design must be obtained from an initial guess $\boldsymbol{\theta}^0$. In this case, the designs obtained will be locally optimal. When it is not possible to obtain an analytical expression of the model, the procedure developed in Rodríguez-Díaz and Sánchez-León [16] may be used. The usual procedure is to choose an optimality criterion, represented by ϕ and derived from the information matrix, that guides design selection. Notably, the D-criterion stands out as the favoured approach, aiming to minimise the volume of the confidence ellipsoid enclosing the model's parameters [17, 18]. For a given $\boldsymbol{\theta}^0$, D-optimality maximises the determinant of the FIM. If the number of support points of the D-optimal design is the number of parameters of the model, which happens quite often, the D-optimal design has equal weights. The general equivalence theorem [19] states that an approximate design, $\tilde{\xi}$, is D-optimal if and only if

$$\psi_a(\tilde{\xi}, \boldsymbol{\theta}, t) = \mathbf{f}(t, \boldsymbol{\theta})^T \tilde{\mathbf{M}}^{-1}(\tilde{\xi}) \mathbf{f}(t, \boldsymbol{\theta}) \leq p \quad \forall t \in \mathcal{T}, \quad (4)$$

where $\tilde{\mathbf{M}}$ denotes the information matrix of the approximate design $\tilde{\xi}$ that can denote any probability measure with finite support. An exact design is essentially an approximate design with equal weights across all points (each one equal to $1/n$). Then, for an exact design's FIM, $\mathbf{M}$, the corresponding approximate design's FIM, $\tilde{\mathbf{M}}$, equals $\mathbf{M}/n$. Thus, (4) can be expressed as

$$\psi(\tilde{\xi}, \boldsymbol{\theta}, t) = nf(t, \boldsymbol{\theta})^T \mathbf{M}^{-1}(\xi) f(t, \boldsymbol{\theta}) \leq p \quad \forall t \in \mathcal{T}, \quad (5)$$

with equality reached at the support points. The function $\psi(\xi, \boldsymbol{\theta}, t)$ is known as the *sensitivity function*.

Let $\boldsymbol{\theta}_T$ represent the true values of the parameters and ξ_D^* the locally D-optimal design obtained using these initial values.

Then the goodness of a design ξ can be evaluated by its efficiency when compared to ξ_D^*. If N experiments can be performed under design ξ, having a D-efficiency of w, this implies that the same accuracy (measured by the D criterion) in estimations could be achieved by conducting only wN experiments under the optimal design ξ_D^* [20]. The D-efficiency can be computed as

$$\frac{n_2}{n_1}\left[\frac{|\mathbf{M}(\xi,\theta_T)|}{|\mathbf{M}(\xi_D^*,\theta_T)|}\right]^{\frac{1}{p}}, \tag{6}$$

where n_1, n_2 denote the number of observations for ξ and ξ_D^*, respectively.

3.1 | Optimal Designs for the SG_2 Model

The simplified-Gamma model (SG) for two intakes introduced before, C_{SG_2}, is a triparametric model with parameter vector $\boldsymbol{\theta} = (s, a, b)^T$. The model can be linearised as

$$\mathbf{f}(t;\theta)^T = \frac{\partial C_S G_2(t;\theta)}{\partial \boldsymbol{\theta}} = \left(e^{-bt}t^a + \frac{e^{-b(t-m)}\left(\left(1-\frac{m}{t_{max}}\right)t\right)^a}{1+e^{-\frac{6(t-lag-m)}{\mathcal{L}}}},\right.$$
$$e^{-bt}st^a\log(t) + \frac{e^{-b(t-m)}s\left(\left(1-\frac{m}{t_{max}}\right)t\right)^a\log\left(\left(1-\frac{m}{6}\right)t\right)}{1+e^{-\frac{6(t-lag-m)}{\mathcal{L}}}},$$
$$\left. -e^{-bt}st^{1+a} + \frac{e^{-b(t-m)}s(mt)\left(\left(1-\frac{m}{t_{max}}\right)t\right)^a}{1+e^{-\frac{6(t-lag-m)}{\mathcal{L}}}} \right). \tag{7}$$

The design is considered to be minimally supported or saturated when the number of support points equals the number of parameters in the model. Because the criterion becomes a function of the p unknown support locations only, having a minimally supported design minimises the difficulty of computing optimal designs. Then, if the three-point design $\xi = \{t_1, t_2, t_3\}$, $t_i > 0$, is written as $\xi = \{t, t+d_1, t+d_1+d_2\}$ where $t_1 = t$, $t_2 = t+d_1$, $t_3 = t+d_1+d_2$ and $t, d_1, d_2 > 0$, then $X = [\mathbf{f}(t;\theta)^T, \mathbf{f}(t+d_1;\theta)^T, \mathbf{f}(t+d_1+d_2;\theta)^T]$, with $\mathbf{f}(t;\theta)^T$ from (7).

Due to the complexity of the information matrices, it is not possible to obtain analytical expressions for the optimal designs, which should be computed through numerical procedures. Using some initial values, optimal designs were calculated to test the possible effects of the choice of lag and $\mathcal{L}$, as well as the time of the second intake. Let us initially assume $\boldsymbol{\theta} = (s, a, b) = (1, 0.39, 0.39)$ as in Mariñas-Collado et al. [6]. Assuming that the second consumption was 30 min ($m = 0.5$ h) after the first one, $lag = 0.5$ and $\mathcal{L} = 1$, the D-optimal design is

$$\xi_D = \{0.0641, 1.4313, 4.8702\}, \tag{8}$$

that is, to measure BAC 3.85, 85.87 and 292.18 min after the first consumption. The D-optimality of the design is checked using (5), which is shown in Figure 6.

The efficiency function (6) is used to study the dependency of the D-optimal designs on the values for lag and $\mathcal{L}$, which have to be chosen beforehand. Fixing $\boldsymbol{\theta} = (s, a, b)$, the efficiency of the D-optimal design (calculated with $lag = 0.5$ and $\mathcal{L} = 1$) when different values for lag and $\mathcal{L}$ are considered as *true* values are shown in Figure 7. It can be seen that the efficiency always surpasses 93%, which proves that the optimal designs are robust with respect to the choice of both lag and $\mathcal{L}$. Thus, from now on, $lag = 0.5$ and $\mathcal{L} = 1$ will be assumed.

On the other hand, it is assumed that the time of the second intake is also known, which might not be so clear sometimes. Hence, assessing the efficiency of the optimal design when considering potential deviations in the m parameter becomes imperative. In this case, it was set that the second intake occurred half an hour after the first one. In Figure 8, it can be seen that, within a reasonable error margin or consumption time range ($\pm$ 10 min), the efficiency remains almost at 100%. Even if the real m were 20 min earlier or later, the design efficiency still remains above 95%, which proves the robustness of the optimal design with respect to m.

3.1.1 | Nearly Optimal Designs

Apart from observing the efficiency of the design with a different m specification, it could also be beneficial to study how the D-optimal design depends on the second intake, m. In Figure 9,

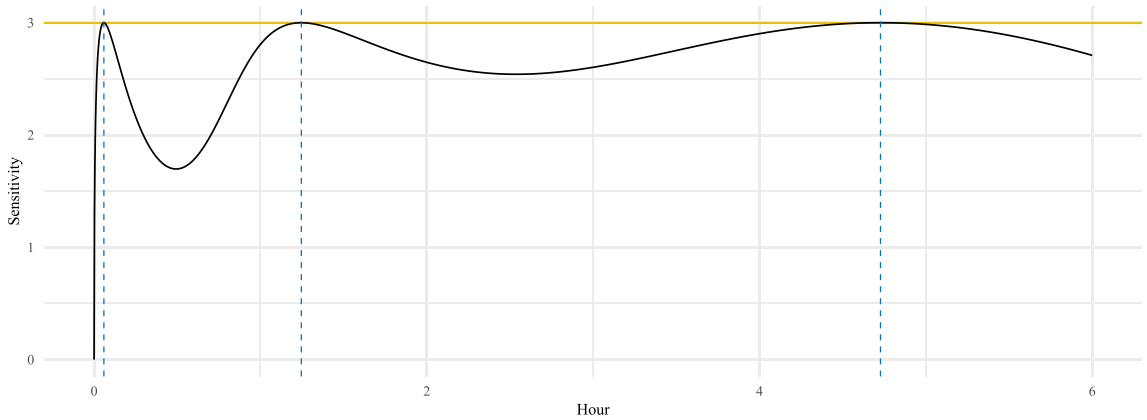


FIGURE 6 | Sensitivity function for the D-optimal design with $\boldsymbol{\theta} = (s, a, b) = (1, 0.39, 0.39)$, $lag = 0.5$ and $\mathcal{L} = 1$, over the time range (up to 6 h). Vertical lines represent the design support points.

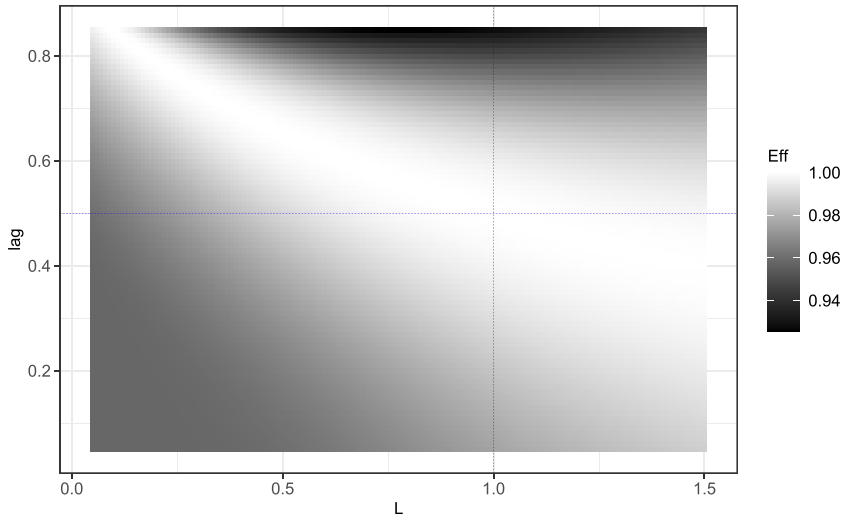


FIGURE 7 | Efficiency of the D-optimal design with respect to the choice of values for *lag* and $\mathcal{L}$.

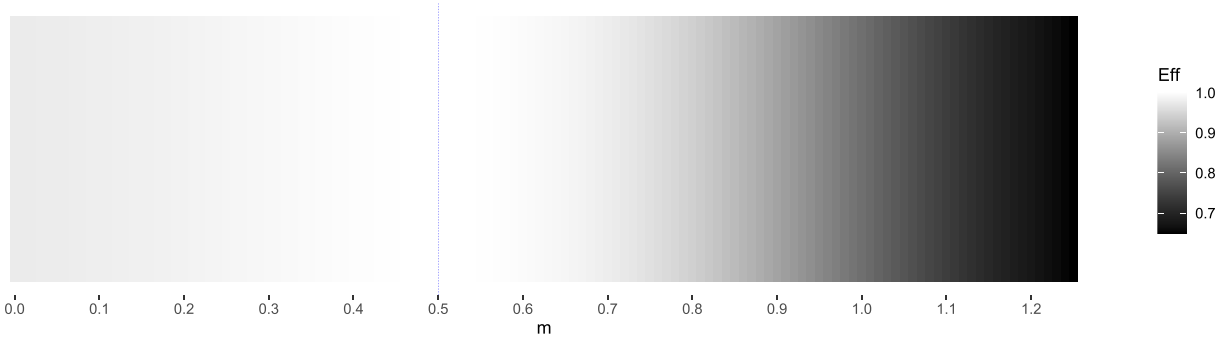


FIGURE 8 | Efficiency of the D-optimal design when the second intake is taken at a time m different from $m = 0.5$.

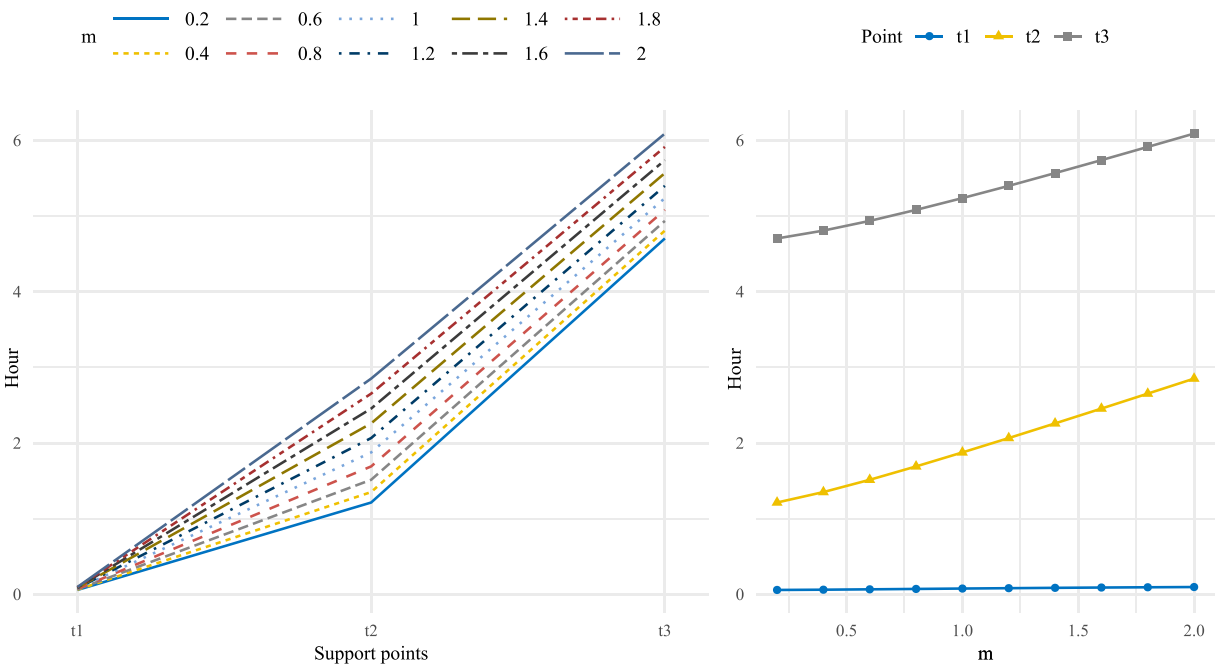


FIGURE 9 | Support points for D-optimal designs with different values of m.

the support points of the D-optimal designs with different m values are shown. On the left side, each design is represented by a line connecting the three time points designated for sample collection. The right side illustrates the connection between support points across different designs. A noticeable trend emerges when linking t_1, t_2 and t_3 for varying m values. This trend demonstrates a consistent behavior in t_1, remaining relatively stable as m increases. In contrast, t_2 and t_3 show a tendency to shift towards later time points (larger values) as m increases. This suggests the possibility of computing t_1, t_2 and t_3 values based on the observed slope in these lines, potentially eliminating the need for explicit calculation of the D-optimal designs (but the ones for the extreme possible values of m).

Based on the preceding outcomes, the following procedure can provide nearly optimal designs:

- Compute the optimal support points for the extreme values of m such as, for instance, $m = 0.5$ and $m = 2$ for each pair (a, b).
- Generate connecting lines among the respective support points $\{t_1, t_2, t_3\}$.
- Approximate the optimal support points for any given value of m from these established lines.

D-optimal designs and their corresponding nearly optimal design are displayed in Figure 10, together with a bar representing the effectiveness of the nearly optimal designs with respect to the D-optimals (note the different scales employed to plot the three lines in order to show the differences, differences that are really small for the first point of the designs, at the bottom). Some instances are also shown in Table 1.

4 | Case Study

Li et al. [7] conducted an experiment using non-linear regression models with a non-zero right-skewed bell-shaped assumption. Using their results as initial values for the parameters,

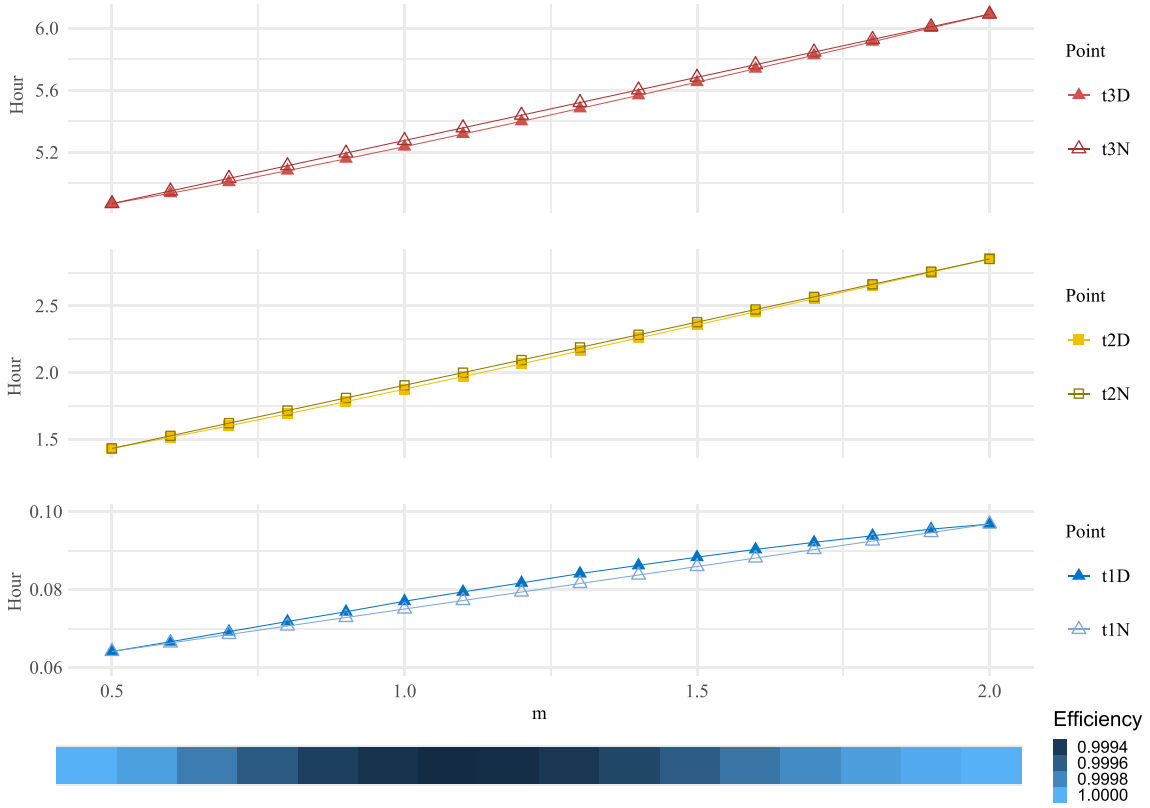


FIGURE 10 | Comparison of D-optimal designs (D, solid shapes) and nearly optimal designs (N, hollow shapes). Different scales have been used for the three points, in order to show the differences between the designs.

TABLE 1 | D-optimal designs, ξ_D, nearly optimal designs, ξ_N, and the efficiency of the nearly optimal designs with respect to the D-optimals.

m	0.7	0.9	1.1	1.5	1.7
ξ_D	{0.069,1.601,5.008}	{0.074,1.783,5.159}	{0.079,1.970,5.317}	{0.088,2.357,5.653}	{0.092,2.554,5.825}
ξ_N	{0.069,1.621,5.033}	{0.073,1.809,5.196}	{0.077,1.999,5.358}	{0.086,2.378,5.684}	{0.090,2.568,5.847}
Eff.	0.9998	0.9995	0.9993	0.9996	0.9998

Note: The designs are expressed as $\{t_1, t_2, t_3\}$.

TABLE 2 | D-Optimal designs (in minutes) for 38-year-old men: 45 kg, 40 g alcohol consumption per intake, $\theta = (48, 0.36, 0.42)$.

D-opt	$\lambda = 0.25$	$\lambda = 0.50$	$\lambda = 1$	$\lambda = 2, \lambda = \infty$
$m = 30$	{3, 76, 228}	{3, 78, 254}	{3, 81, 268}	{3, 82, 270}
$m = 45$	{3, 92, 240}	{3, 93, 267}	{3, 95, 280}	{3, 96, 282}
$m = 60$	{4, 107, 253}	{4, 108, 279}	{4, 110, 292}	{4, 110, 294}
$m = 75$	{4, 123, 266}	{4, 124, 292}	{4, 125, 305}	{4, 125, 307}
$m = 90$	{4, 138, 279}	{4, 139, 305}	{4, 140, 319}	{4, 141, 320}
$m = 105$	{3, 154, 292}	{4, 154, 318}	{4, 155, 332}	{4, 156, 333}
$m = 120$	{5, 169, 305}	{5, 170, 332}	{5, 170, 346}	{5, 170, 347}

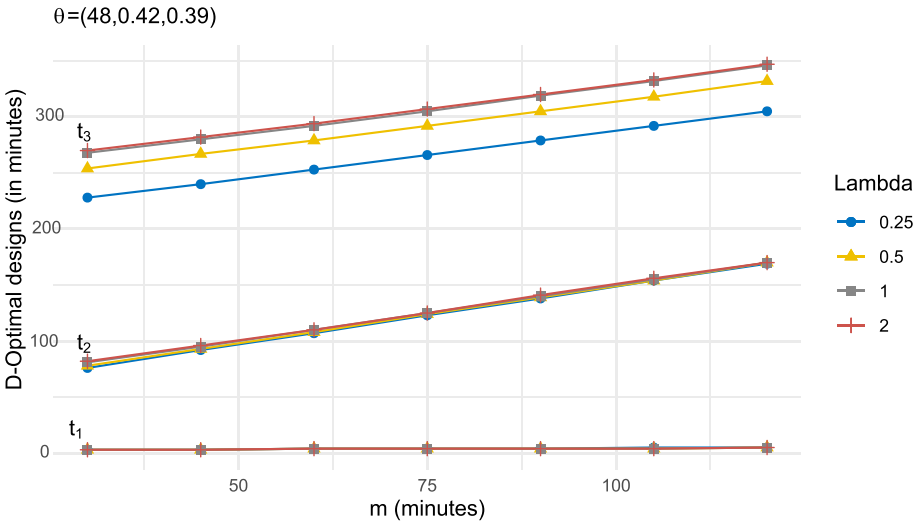


FIGURE 11 | Support points t_1, t_2 and t_3 for different values of m (in minutes) and different values of λ.

this case study aims to explore various optimal designs based on the timing of the second alcohol intake. These initial data serve as a foundation for assessing different scenarios, allowing an examination of how optimal designs vary concerning the temporal structure of intake. Additionally, the consideration of correlation among observations is introduced, a pivotal aspect in accurately estimating the model parameters. The covariance between observations of the same subject taken at different points is assumed to depend on the distance between points, that is, $Cov[y(t_i), y(t_j)] = \rho(|t_i - t_j|)$, which is a reasonable and widely accepted assumption. For the stationary covariance kernel, ρ (assumed known), the widely used exponential kernel is chosen: $\rho(d) = e^{-\lambda d}$, where λ is characteristic of the subject. Then, the covariance matrix for the design $\{t, t + d_1, t + d_1 + d_2\}$ is:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & e^{-\lambda d_1} & e^{-\lambda(d_1+d_2)} \\ e^{-\lambda d_1} & 1 & e^{-\lambda d_2} \\ e^{-\lambda(d_1+d_2)} & e^{-\lambda d_2} & 1 \end{pmatrix}.$$

Table 2 shows optimal designs ($\{t_1, t_2, t_3\}$, in minutes) for various λ values when the second consumption is taken at different times, m, expressed in minutes as well. This designs are also shown in Figure 11, where it can be more clearly seen that the designs are consistent with the patterns observed before. It can be highlighted that in the optimal designs the first observation (t_1) should be made as soon as possible, while the final one (t_3) should be done between 228 and 347 min (4 and 6 h approximately) after intake, which is when alcohol is thought to have nearly completely left the body. The middle point (t_2) should be taken about 45–50 min after the second intake m. Note that as $\lambda \to \infty$, $\Sigma \to I$, and consequently the designs incorporating a correlation structure converge to those obtained when the correlation is not accounted for. In this case this can be observed from $\lambda = 2$. It can be noted as well that correlation has little influence in the designs, and almost exclusively in the third observation, while the first two only change slightly for different values of λ. Additional tables for

TABLE 3 | Parameter estimations from the 27-subject simulations assuming $\theta = (48, 0.36, 0.42)$, for the designs with $\lambda = \infty$ shown in Table 2.

	$m = 60$	$m = 90$
Max	(51.07,0.2322,0.2999)	(55.57,0.1966,0.3361)
Mean	(48.02,0.3887,0.4232)	(49.47,0.3608,0.4326)
Min	(42.15,0.7434,0.6056)	(47.76,0.6053,0.6328)

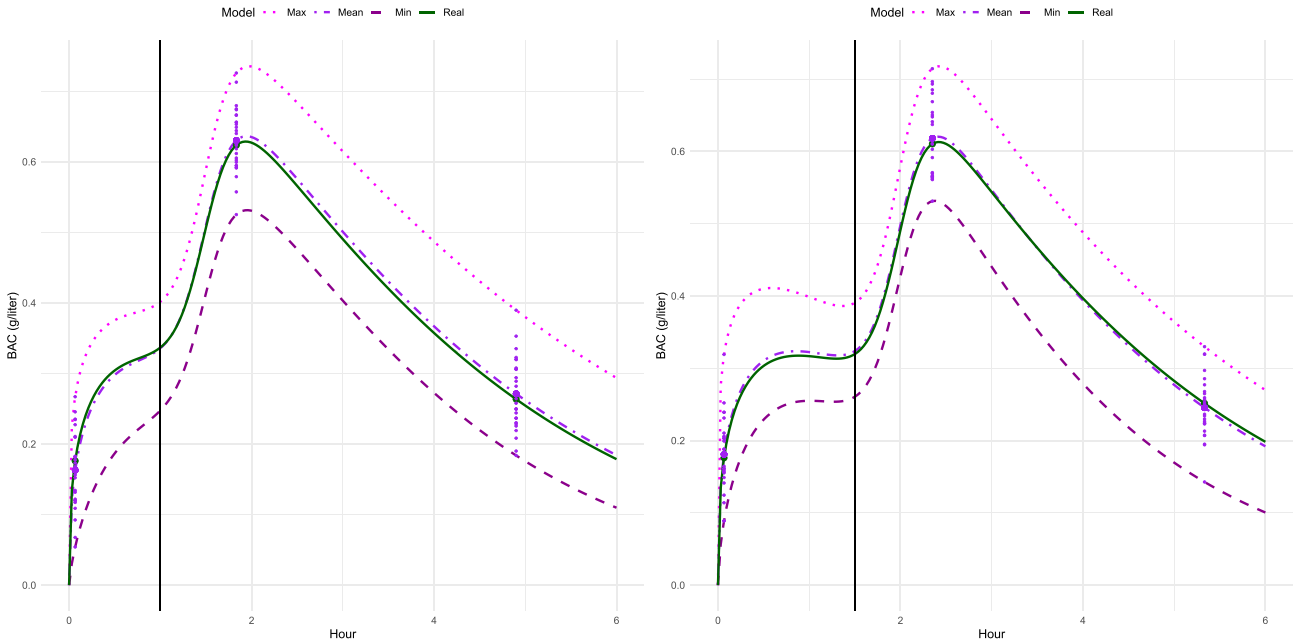


FIGURE 12 | The 27 simulated observations (dots) at the three design points, the assumed actual model (solid line) and the C_{SG_2}-type models based on parameter estimates from the minimum, maximum and mean values of the simulated observations (dashed lines). The vertical line marks the time of the second intake.

different cases (weight, alcohol consumption) and their corresponding figures are shown in Appendix A.

In order to demonstrate the effectiveness of the computed designs, simulations were performed assuming the C_{SG_2} model with actual parameter values set to $\boldsymbol{\theta} = (48,0.36,0.42)$, as used in Table 2. To emulate the approach in [14], samples of 27 people were simulated, with a normally distributed random error with 0 mean and a standard deviation of 0.05, added to the BAC values of the model at the design points. Then, the minimum, maximum and mean values of the 27 simulated observations in each design point were used to estimate the model parameters. The parameter estimations, using the designs obtained for two different values of m and assuming $\lambda = \infty$, are shown in Table 3, and the simulated observations and the C_{SG_2}-type models using these estimates are plotted in Figure 12. It can be seen that, when using the mean of the simulations, the parameter estimates are very close to the assumed true values, and thus, the model generated with those estimates fits the assumed actual model. Note that, differently than in Figure 3, only three samples have been taken for each subject.

5 | Conclusions

A novel model, C_{SG_2}, is proposed to characterise BAC following multiple alcohol intakes, a scenario of practical relevance often overlooked in forensic literature. The proposed model, initially designed for the case of two intakes but easily generalised, demonstrates simplicity, practicality, and affordability, making it accessible for the wider scientific community. Expanding on this, it is important to emphasise the model's flexibility, especially when dealing with many small alcohol intakes. In these situations, it can be seen (see Figure 1a) that the model can be approximated by a single intake curve. Addressing challenges identified in existing literature related to the inadequate treatment of multiple alcohol intakes, the study highlights the significance of proper experimental design for parameter estimation.

Out of the various optimality criteria aimed at achieving precise estimations, the D-criterion stands out as the most popular, emphasising the minimisation of the confidence ellipsoid volume surrounding the estimators of model parameters. Consequently, D-optimal designs have been computed for the proposed model, guiding the selection of observation times for BAC measurements in order to attain the most accurate estimations of the model parameters. These parameters, including the correlation parameter λ, rely on factors such as the quantity and type of alcohol consumed, sex, age, weight, height and so forth and must be established through clinical trials. In the future, after conducting the necessary experimental studies, lambda could be selected to characterise the "standard" individual, meaning a value that captures the characteristics of the average subject. Furthermore, upon obtaining the model and its parameters through a clinical trial (which is currently beyond the authors' scope), the proposed model could be compared with alternative models to validate its ability to adequately fit true ethanol kinetics. The simulations of 27 observations at the design points (for different values of m) resulted in parameter estimates that closely match the assumed actual model, demonstrating the effectiveness of the computed designs.

After computing optimal designs for different values of the second intake m and the choice of lag and $\mathcal{L}$, a sensitivity analysis has been performed, proving that the designs computed are quite robust respect to the choice of the initial values of these parameters.

In examining three-point optimal designs, the analysis focused on the varying behavior of each specified point (initial, medium, final) for data collection as m changed. Different slopes were

identified, indicating distinct linear trends for each point based on m. The precision of the fitted lines for each point was noteworthy, allowing the proposal of 'nearly optimal designs'. These designs involve computing optimal configurations solely for the extreme values of m. It has been checked that the proposed nearly optimal designs are almost as efficient as the D-optimal ones, thus this procedure can be employed for each case study as a very easy way to obtain tables of optimal designs for different values of m.

In some cases, it may be interesting to treat the second intake, m, as an unknown parameter that needs to be estimated as well. This would result in a more complex treatment and computations, constituting an objective for future research.

Another consideration is the practical feasibility of optimal designs beyond laboratory settings, such as when conducting routine alcohol control tests for drivers on the road. In such cases, time constraints often conflict with the proposed designs, particularly the final observation. Exploring 'restricted optimal designs' could be a promising avenue for future research. When the model parameters are known, the interest may rest in the estimation of the initial intake, and thus c-optimal designs could be computed [21]. In addition, when different types of measures are to be taken (exhaled air, blood samples), multiresponse models should be consider, [22–24]. A user-friendly web application is being developed to assist non-specialised users in obtaining optimal (or near-optimal) designs for specific experiments. This tool will be applicable to both the single intake model proposed in [6] and the multiple intake model presented in this study, which may provide significant benefits.

Acknowledgments

This research was partially supported by the Spanish Ministry of Science, Innovation and Universities project PID2021-125211OB-I00 (all authors) and by the Junta de Castilla y León project 'SA217P23' (J.M.R.-D. and M.T.S.-M.).

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

1. A. W. Jones, "Evidence-Based Survey of the Elimination Rates of Ethanol From Blood With Applications in Forensic Casework," *Forensic Science International* 200, no. 1-3 (2010): 1–20.

2. P. E. Watson, I. D. Watson, and R. D. Batt, "Prediction of Blood Alcohol Concentrations in Human Subjects. Updating the Widmark Equation," *Journal of Studies on Alcohol* 42, no. 7 (1981): 547–556.

3. A. A. Norberg, B. Sandhagen, L.-E. Bratteby, et al., "Do Ethanol and Deuterium Oxide Distribute Into the Same Water Space in Healthy Volunteers?,"*Alcoholism: Clinical and Experimental Research* 25, no. 10 (2001): 1423–1430.

4. L. G. Sultatos, G. M. Pastino, C. A. Rosenfeld, and E. J. Flynn, "Incorporation of the Genetic Control of Alcohol Dehydrogenase Into a Physiologically Based Pharmacokinetic Model for Ethanol in Humans," *Toxicological Sciences* 78, no. 1 (2004): 20–31.

5. P. Zekan, N. Ljubičić, V. Blagaić, et al., "Pharmacokinetic Analysis of Ethanol in a Human Study: New Modification of Mathematic Model," *Toxics* 11, no. 9 (2023): 793. https://www.mdpi.com/2305-6304/11/9/793.

6. I. Mariñas-Collado, J. M. Rodríguez-Díaz, and M. T. Santos-Martín, "Optimal Designs for a Non-Linear Model for the Pharmacokinetics of Ethanol Elimination in the Human Body," *Chemometrics and Intelligent Laboratory Systems* 238 (2023): 104839.

7. Y. C. Li, N. N. Sze, S. C. Wong, K. L. Tsui, and F. L. So, "Experimental Study of the Temporal Profile of Breath Alcohol Concentration in a Chinese Population After a Light Meal," *PLoS One* 14, no. 9 (2019): e0221237.

8. E. Baraona, C. S. Abittan, K. Dohmen, et al., "Gender Differences in Pharmacokinetics of Alcohol," *Alcoholism: Clinical and Experimental Research* 25, no. 4 (2001): 502–507.

9. H. Kalant, "Effects of Food and Body Composition on Blood Alcohol Curves," *Alcoholism: Clinical and Experimental Research* 24, no. 4 (2000): 413–414.

10. M. Wedel, J. E. Pieters, N. A. Pikaar, and T. Ockhuizen, "Application of a Three-Compartment Model to a Study of the Effects of Sex, Alcohol Dose and Concentration, Exercise and Food Consumption on the Pharmacokinetics of Ethanol in Healthy Volunteers," *Alcohol and Alcoholism* 26, no. 3 (1991): 329–336.

11. B. Lee, H. Yoon, I. Baek, and K. Kwon, "Population Pharmacokinetics of Multiple Alcohol Intake in Humans," *Alcohol* 47, no. 2 (2013): 159–165.

12. A. Heck, "Modelling Intake and Clearance of Alcohol in Humans," *Electronic Journal of Mathematics and Technology* 1, no. 3 (2007): 232–244.

13. A. W. Jones, "Alcohol, Its Absorption, Distribution, Metabolism, and Excretion in the Body and Pharmacokinetic Calculations," *Wiley Interdisciplinary Reviews: Forensic Science* 1, no. 5 (2019): e1340.

14. M. S. Mumenthaler, J. L. Taylor, and J. A. Yesavage, "Ethanol Pharmacokinetics in White Women: Nonlinear Model Fitting Versus Zero-Order Elimination Analyses," *Alcoholism: Clinical and Experimental Research* 24, no. 9 (2000): 1353–1362.

15. P. D. H. Hill, "D-Optimal Designs for Partially Nonlinear Regression Models," *Technometrics* 22, no. 2 (1980): 275–276.

16. J. M. Rodríguez-Díaz and G. Sánchez-León, "Design Optimality for Models Defined by a System of Ordinary Differential Equations," *Biometrical Journal* 56, no. 5 (2014): 886–900.

17. A. Atkinson, A. Donev, and R. Tobias, *Optimum Experimental Designs, with SAS*, Vol. 34, (OUP Oxford, 2007).

18. V. V. Fedorov, *Theory of Optimal Experiments* (Elsevier, 2013).

19. J. Kiefer, "General Equivalence Theory for Optimum Designs (Approximate Theory)," *The Annals of Statistics* 2, no. 5 (1974): 849–879.

20. I. Mariñas-Collado, M. J. Rivas-López, J. M. Rodríguez-Díaz, and M. T. Santos-Martín, "A New Compromise Design Plan for Accelerated Failure Time Models With Temperature as an Acceleration Factor," *Mathematics* 9, no. 8 (2021): 836.

21. J. M. Rodríguez-Díaz, "Computation of c-Optimal Designs for Models With Correlated Observations," *Computational Statistics & Data Analysis* 113 (2017): 287–296.

22. J. M. Rodríguez-Díaz and G. Sánchez-León, "Efficient Parameter Estimation in Mulltiresponse Models Measuring Radioactivity Retention," *Radiation and Environmental Biophysics* 58, no. 2 (2019a): 167–182.

23. J. M. Rodríguez-Díaz and G. Sánchez-León, "Optimal Designs for Multiresponse Models With Double Covariance Structure," *Chemometrics and Intelligent Laboratory Systems* 189 (2019b): 1–7.

24. J. M. Rodríguez-Díaz, "Optimal Sample Plans for Multiresponse and Multisubject Experiments," *Chemometrics and Intelligent Laboratory Systems* 231 (2022): 104699.

Appendix A

D-Optimal Design in Minutes for Different Weights in 38-Year-Old Men

The following tables (Tables A1, A2, A3 and A4) show D-optimal designs in minutes, using initial values from Li et al. [7], for different weights in 38-year-old Chinese men consuming different quantities of alcohol. These designs are also shown in Figure A1. An intermediary value, such as $\lambda = 1$, could be used in the absence of prior knowledge about the value of λ.

TABLE A1 | 38-year-old men: 55 kg, 40 g alcohol consumption per intake, $\theta = (39.75, 0.33, 0.36)$.

	$\lambda = 0.25$	$\lambda = 0.50$	$\lambda = 1$	$\lambda = 2, \lambda = \infty$
$m = 30$	{2, 78, 257}	{3, 79, 287}	{2, 83, 298}	{2, 85, 300}
$m = 45$	{3, 93, 269}	{3, 95, 299}	{3, 96, 310}	{3, 98, 312}
$m = 60$	{3, 108, 282}	{3, 110, 311}	{3, 112, 323}	{3, 112, 324}
$m = 75$	{3, 124, 295}	{3, 125, 325}	{3, 126, 336}	{3, 127, 337}
$m = 90$	{4, 139, 308}	{4, 140, 338}	{3, 141, 349}	{3, 141, 350}
$m = 105$	{4, 155, 321}	{4, 156, 351}	{4, 156, 362}	{4, 156, 363}
$m = 120$	{4, 170, 334}	{4, 171, 364}	{4, 171, 375}	{4, 171, 376}

TABLE A2 | 38-year-old men: 65 kg, 40 g alcohol consumption per intake, $\theta = (34.29, 0.29, 0.31)$.

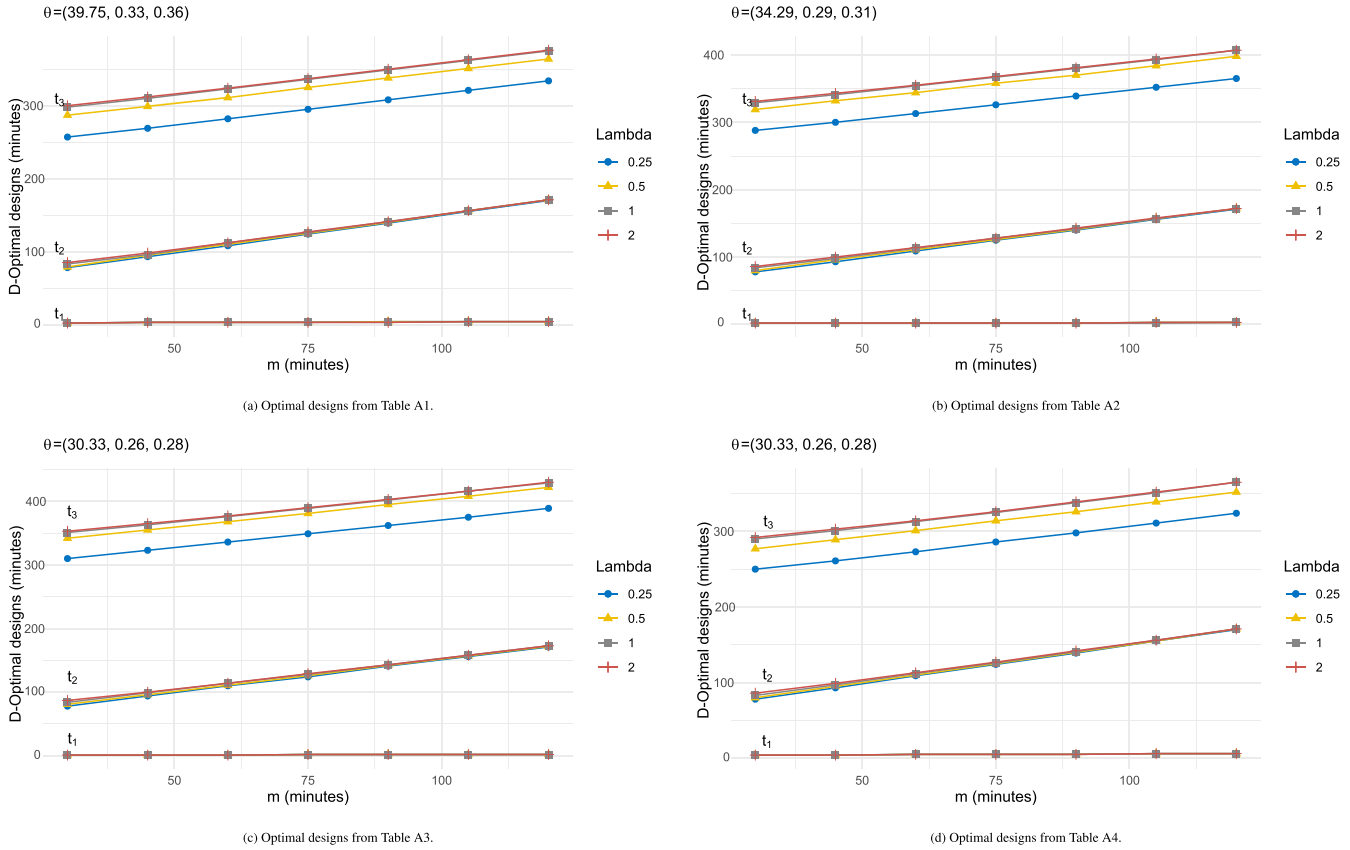
	$\lambda = 0.25$	$\lambda = 0.50$	$\lambda = 1$	$\lambda = 2, \lambda = \infty$
$m = 30$	{2, 78, 288}	{2, 80, 319}	{2, 84, 329}	{2, 86, 331}
$m = 45$	{2, 93, 300}	{2, 96, 332}	{2, 98, 341}	{2, 100, 343}
$m = 1$	{2, 109, 313}	{2, 111, 344}	{2, 113, 354}	{2, 114, 355}
$m = 75$	{2, 125, 326}	{2, 126, 358}	{2, 128, 367}	{2, 128, 368}
$m = 90$	{2, 140, 339}	{2, 141, 370}	{2, 142, 380}	{2, 143, 381}
$m = 105$	{3, 156, 352}	{3, 157, 384}	{3, 157, 393}	{3, 158, 394}
$m = 2$	{3, 171, 365}	{3, 172, 398}	{3, 172, 407}	{3, 172, 407}

TABLE A3 | 38-year-old men: 75 kg, 40 g alcohol consumption per intake, $\theta = (30.33, 0.26, 0.28)$.

	$\lambda = 0.25$	$\lambda = 0.50$	$\lambda = 1$	$\lambda = 2, \lambda = \infty$
$m = 30$	{1, 78, 310}	{1, 81, 342}	{1, 84, 351}	{1, 87, 353}
$m = 45$	{1, 94, 323}	{1, 96, 355}	{1, 99, 363}	{1, 100, 365}
$m = 60$	{1, 110, 336}	{1, 111, 368}	{1, 114, 376}	{1, 114, 377}
$m = 75$	{2, 124, 349}	{2, 126, 381}	{2, 128, 389}	{2, 129, 390}
$m = 90$	{2, 141, 362}	{2, 142, 395}	{2, 143, 402}	{2, 143, 403}
$m = 105$	{2, 156, 375}	{2, 157, 408}	{2, 158, 416}	{2, 158, 416}
$m = 2$	{2, 171, 389}	{2, 172, 422}	{2, 173, 429}	{2, 173, 430}

TABLE A4 | 38-year-old men: 66.7 kg, 60 g alcohol consumption, $\theta = (62.22, 0.39, 0.39)$.

	$\lambda = 0.25$	$\lambda = 0.50$	$\lambda = 1$	$\lambda = 2, \lambda = \infty$
$m = 30$	{4, 78, 250}	{4, 80, 277}	{4, 83, 290}	{4, 86, 292}
$m = 45$	{4, 93, 261}	{4, 95, 289}	{4, 97, 301}	{4, 99, 303}
$m = 1$	{5, 109, 273}	{5, 110, 301}	{5, 112, 313}	{5, 113, 314}
$m = 75$	{5, 124, 286}	{5, 125, 314}	{5, 126, 325}	{5, 127, 326}
$m = 90$	{5, 139, 298}	{5, 140, 326}	{5, 141, 338}	{5, 142, 339}
$m = 105$	{6, 155, 311}	{6, 155, 339}	{6, 156, 351}	{6, 156, 352}
$m = 2$	{6, 170, 324}	{6, 171, 352}	{6, 171, 365}	{6, 171, 365}



(a) Optimal designs from Table A1.

(b) Optimal designs from Table A2

(c) Optimal designs from Table A3.

(d) Optimal designs from Table A4.

FIGURE A1 | Support points t_1, t_2, t_3 for different values of m (in minutes) and different values of λ for different weights in a 38-year-old Chinese men consuming different quantities of alcohol.