Rigorous Bayesian Improvements to Ballistics Armour Testing Methods

M. McKibben[1], L. Stabile[1], A. Iwaskiw[1] and E. Cawi[1]

[1] *Johns Hopkins Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, 20723 MD., USA, Michael.Mckibben@jhuapl.edu*

Abstract. While testing methods such as Neyer and 3POD offer marked improvements over earlier test methods such as Bruceton and the step method in NIJ Standard 0101.07 [1], there is still some susceptibility to the initial user guess (μ_{Guess} and σ_{Guess}). Johnson et al [2] offered a rigorous comparison of all testing methods currently in the field, and noted that a mis specified σ_{Guess} can lead to large bias for all tested methods. Additionally, while in some cases prior data such as manufacturer testing and previous first articles tests may be available to inform μ_{Guess} and σ_{Guess}, there is not an existing framework for how this data should be included in a model. Note that updating σ_{Guess} is also consistent with the movement to incorporate unbiased error measurements in the broader design of experiments community initially spearheaded by Gilmour and Trinca [3]. In this paper, we offer both a rigorous Bayesian framework to include prior data, as well as additional considerations on when to include and how to weight that information. We also include a simulation study demonstrating the implications both of updating σ_{Guess}, as well as of a Bayesian design more generally. The results have implications for a variety of testing methodologies; for example, it is well known that Robbins Monro may not be able to achieve an unbiased estimate of V50 even asymptotically when initial parameters are incorrectly specified (i.e. when σ_{Guess} is too small).

1. INTRODUCTION

Personal protective equipment (PPE), which is designed to stop ballistic threats, is an important tool to protect military and law enforcement personnel. In order to ensure optimal performance of these products, PPE must be tested using live ammunition to evaluate their protective capabilities. This is primarily achieved by employing a scheme of testing where armour performance (scored as a partial or complete penetration) is assessed at various ballistic conditions (ambient versus elevated temperature, etc.) in a limited range of firing velocities. Various testing methodologies [4], [5], [6] describe velocity selection procedures, tools, measurements and techniques to test and assess the ballistic performance of PPE. These methods are used to determine various ballistic performance metrics, such as resistance to penetration, where 0 percent of events is expected to result in complete penetrations (RTP), and the ballistic striking velocity where 50 percent of events will be complete penetrations ($V50$). However, these tests are costly and therefore there is a strong desire to reduce the testing resources needed to calculate these ballistic measures to a specific confidence, as well as to expand the characterization of the PPE ballistic response curve. Using advanced statistical approaches may reduce the resources needed to accurately assess ballistic PPE performance [7]. This would also allow practitioners to calculate additional metrics, such as the ballistic striking velocity where 10 percent of events will be complete penetrations ($V10$).

Most statistical methods used in ballistics testing are Frequentist [7]. Bayesian methods allow the incorporation of information from previously collected data into the selection of tested velocities, through the use of priors. In complicated live-fire destructive testing firing schemes, Bayesian methods could better inform both initial test conditions and the subsequent testing sequence. These methods could also reduce the number of tests needed to achieve a certain level of confidence with respect to ballistic performance.

Studies of existing methods have generally not included the perturbation of target velocity to actual velocity in their simulations, i.e. when the striking velocity differs from the intended velocity. This additional source of variance is essential when considering the efficacy of existing methods in practice. This paper compares commonly used ballistic firing sequences (Neyer, 3POD, Bruceton) via a simulation study in which this variation is captured to assess the ability to reduce the number of tests needed through these approaches. Additionally, we introduce the idea of a Bayesian methodology to incorporate additional data that is currently unused in the test generation of initial conditions estimates (μ_{Guess} and σ_{Guess}). In certain scenarios, it is hypothesized that this will allow the practitioner to find the Zone of Mixed Results (ZMR) faster, and produce more accurate response curves than currently used methods.

 https://doi.org/10.52202/081621-0039

2. LITERATURE REVIEW

One of the earliest ballistics testing procedures to gain popularity was Bruceton [8], a binary search procedure published in 1948. While Bruceton is consistently able to find V50, it is generally inefficient unless an optimal starting step size is chosen. The Robbins-Monro and Delayed Robbins-Monro [9], [10] from 1961 and 1985, respectively, improved upon this method by dynamically altering the step size throughout the test. Robbins-Monro Joseph [11] incorporates the notion of V50 estimation following a normal distribution in shot selection, which was built upon by later methods. Langlie [12] combined several of these early ideas in 1962 into a sequential testing methodology. Ultimately, many of these methods are inefficient when compared to the advancements made in the statistical field of modern optimal design methodology [13] and may be overly dependent upon starting parameters such as μ_{Guess}, σ_{Guess}, and the initial shot velocity. For a more substantive review of historical ballistics testing methods, the reader is directed to Johnson et al. [2].

Neyer was the first to propose a systematic method for generating an optimal design in 1994 [14]. The method consists of two phases: (1) an initial series of shots to establish the appropriate stimulus range and break separation (i.e. finding the Zone of Mixed Results). This is necessary as unlike many traditional Design of Experiments (DoE) settings, the bounds are not known a priori. This is followed by stage (2) the use of a D-Optimal design for subsequent shots [14]. Sequential D-Optimal designs have seen substantial use in the broader DoE applications literature [13], however, Neyer was the first to apply it to destructive sensitivity testing. One of Neyer's contributions was the inclusion of the first phase of the experiment to establish the bounds of the experimental range. This differs from many traditional DoE settings where the bounds are known a priori through subject matter expertise.

3POD is a three-phase sequential optimization procedure that aims to find performance metrics of interest, such as V50, more efficiently than Neyer [15]. Similar to Neyer, the first phase establishes the appropriate experimental range. This is followed by finding the Zone of Mixed Results (breaking separation), and adding points to optimize the remaining tests. Phase 1 of 3POD chooses different points from Neyer to break separation more quickly, though the underlying approach is similar. Unlike Neyer, 3POD explicitly adds a third step with n_2 points to optimize for specific quantiles of interest, as specified by the user. This follows a stochastic search procedure proposed by Joseph [11] to estimate specific quantile of interest; while this method has desirable asymptotic properties, there is some question as the how it performs in the presence of both small n and when the target shot velocity is perturbed. Follow on work for 3POD 2.0 [16] offers some technical improvements, while GONOGO [17] offers an implementation of both 3POD, Neyer, and other methods.

This paper also drew inspiration from modern design techniques in other spaces; specifically, the Bayesian D-Optimal designs of DuMouchel and Jones [18], the Bayesian perspective of Dror and Steinberg [19], and work in the clinical trials space [20], [21], [22], [23]. Collectively, the promise of Bayes is two-fold. First, it allows for the logistic regression curve to be estimated before separation has been broken. This can be particularly useful when choosing future points, as current techniques rely on an average of velocities. Second, these methods have repeatedly demonstrated the power of relevant prior information when creating an optimal design and applying the results to a sequential optimization procedure such as 3POD.

3. METHODS

3.1 A Bayesian Approach

In the broader design community, there is a movement to use Bayesian statistics in both the creation of design matrices and in analysis. Note that priors are generally said to be strong when the prior has a large impact on the analysis or test points selected, and weak or vague when the impact is minimal. In some sense, all the methods discussed in the literature review use prior information. Specifically, the μ_{Guess} and σ_{Guess} parameters in 3POD and Neyer are often based on previous step-down experiments conducted before the test of interest. This is then incorporating summary statistics of prior information, while a Bayesian approach would use the actual test information. Other methods, such as Bruceton, require some estimates related to $V50$ range and variability. There is generally minimal issue with using additional information in choosing initial starting points, provided those starting points are not part of an analytical solution once the data has been collected.

The study presents two particular recommendations. The first is the use of a Bayesian framework to generate subsequent test points, i.e. target velocities. The priors allow for the easy and relevant incorporation of prior information into the design matrix. The second is for additional prior information to be included for the sole purpose of optimal points for shot selection. In many cases, manufacturers have conducted extensive internal testing prior to submission to the government for acceptance or rejection (First Article Testing, i.e. FAT). In some cases, this testing is contracted to the same facility that will perform the FAT. It is easy to demonstrate that incorporating data from those manufacturer tests into the determination to accept or reject a piece of equipment would create various moral hazards; for this reason, many regulations generally state that only information from the official tests may be used for analysis. It is also clear that, in some cases, the manufacturer information could be used to form an unbiased estimate of the response curve that would lead to better testing points. Specifically, the manufacturer data would be used to identify optimal target velocities for an acceptance, but would not be included in the acceptance test analysis. This would allow for a variety of benefits; separation would be broken faster, the response curve would fit more accurately by choosing more optimal test points, and higher confidence and lower variability would be gained in final estimates.

3.2 Method Description

The proposed procedure in this section is named Sequential Bayesian Optimal Procedure (SBOP) and is broadly inspired by 3POD [15]. This method is most suitable when prior information can be leveraged to inform estimates of the input parameters. These estimates may be obtained from a previous step-down test as is common for 3POD and Neyer, or from manufacturer data described in the previous section. The method has three phases that closely align with the phases of 3POD:

- Phase I, in which initial shots are added and separation is broken
- Phase II, in which points are added to optimally estimate the logistic response curve
- Phase III, in which points are added to better estimate quantiles of interest

3.2.1 Phase I

The goal of the first phase of the experiment is to establish valid test regions and break separation by varying the velocity of the projectile. This phase can be divided into the following steps: (I) Obtaining at least one partial and complete penetration (II) Breaking separation, i.e. finding the Zone of Mixed Results.

3.2.2 M.1 Obtain partial and complete penetrations

When the priors for location (μ_{Prior}) and scale (σ_{Prior}) are identical to μ_{Guess} and σ_{Guess}, the initial shot velocities for shots 1 and 2 of 3POD and SBOP are the same as shown in Equations 1 and 2:

$$\text{Shot 1} = \mu_{Prior} - 2\,\sigma_{Prior} = \frac{3}{4}(\mu_{Guess} - 4\,\sigma_{Guess}) + \frac{1}{4}(\mu_{Guess} + 4\,\sigma_{Guess}) \tag{1}$$

$$\text{Shot 2} = \mu_{Prior} + 2\,\sigma_{Prior} = \frac{1}{4}(\mu_{Guess} - 4\,\sigma_{Guess}) + \frac{3}{4}(\mu_{Guess} + 4\,\sigma_{Guess}) \tag{2}$$

Define x_{min} and x_{max} as the lowest and highest actual velocities, respectively. These values are updated after each shot. There are four possible scenarios described below for shots one and two, and each of these has implications for subsequent shot selections. Let y_1 be the outcome of shot one and y_2 be the outcome of shot 2, where 1 indicates a complete penetration (CP) and 0 indicates a partial penetration (PP).

A. [y_1=0, y_2 = 1] This represents the ideal (and also most common) scenario. The practitioner is testing in the appropriate range, and can move to step 1 in M.2

B. [y_1=0, y_2 = 0] Both shots are partial penetrations, and it is inferred that the actual range of interest is further occurs at a higher velocity. The next target velocity is

$$x_{next} = x_{max} + 1.5\,\sigma_{Prior}, \tag{3}$$

and tests are iteratively added until a complete penetration is observed.

C. [y_1=1, y_2 = 1] Both shots are complete penetrations, and it is inferred that the actual range of interest is occurs at a lower velocity. The next target velocity is

$$x_{next} = x_{min} - 1.5\,\sigma_{Prior}, \tag{4}$$

and tests are conducted iteratively until a partial penetration is observed.

D. [y_1=1, y_2 = 0] This scenario represents the case that the actual variance is substantially larger than the specified σ_{Guess}. Two additional shots are added:

$$\text{Shot 3} = x_{min} - 3\,\sigma_{Prior}, \tag{5}$$

$$\text{Shot 4} = x_{max} + 3\,\sigma_{Prior}, \tag{6}$$

 https://doi.org/10.52202/081621-0039

In the event that shot 3 is a PP and shot 4 is a CP, we revise the estimate of σ_{Prior} and move to step 1 in M.2.

For any other result, the authors recommend assessing the input parameters and priors for reasonableness, as this suggests the values have deviated very strongly from both subject matter expert and prior data expectations.

There may be cases where the target velocity is substantially different from the actual velocity; historically, these shots have been omitted from the test matrix and subsequent analysis. The authors recommend retaining these data points in the test matrix if the shot is otherwise within specification for other factors such as yaw and obliquity, even if it is necessary to remove them from subsequent analyses related to pass/fail. Existing methods generally do not comment on this issue.

3.2.3 M.2 Break Separation

Once the initial boundaries are established, the next step is to break separation. The practitioner adds shots until a zone of mixed results is established, i.e. the highest velocity of a partial penetration ($M0$) is greater than the lowest velocity of a complete penetration ($m1$). Similar to 3POD, shots are added based on the size of the separation interval, defined as $sep = M0 - m1$:

- If ($sep < 0$): move to the next stage
- If ($sep > 1.5\sigma_{Prior}$): $x_{next} = \widehat{V50}$
- If ($0 \leq sep < 1.5\sigma_{Prior}$):
    - If (Number of CPs > Number of PPs): $x_{next} = m1 + .3\sigma_{Prior}$ *(7)*
    - Else: $x_{next} = M0 - .3\sigma_{Prior}$

SBOP recommends iteratively adding a minimum of 4 shots near $\widehat{V50}$ until $m1 - M0 < \sigma_{Prior}$. Once the separation interval is sufficiently small, the next points added are at the 40th and 60th quantile estimates on the curve. This is broadly in line with the $mean(X) \pm .3\sigma_{Guess}$ recommended by 3POD. Note that the so-called "separation trap" [15] is not an issue in this methodology or 3POD as the chosen points are outside of the separation interval.

3.2.4 Phase II:

Phase II adds points to optimize estimates for the logistic regression curve parameters. In contrast to methods that use a traditional generalized linear model (GLM) for this step, Bayesian Methods use a Markov Chain Monte Carlo (MCMC) to estimate the logistic response curve. In this implementation, SBOP uses MCMCpack [24]. This has the effect of centering the initial estimates at the prior, while allowing the data to play a larger role as the sample size increases. Optimizing for subsequent points has the form of optimizing the information matrix.

Note that n_1 points are assigned to Phase I and Phase II; the faster separation is broken, the more points will be available to optimize the estimate of the logistic regression curve. Future work will include dynamically altering the distribution of points between Phase I, Phase II, and Phase III, as it is likely that some points should always be assigned to Phase II prior to entering Phase III. To avoid cases of poor estimation, clipped estimates of the mean and scale parameter are used as shown in the equations below.

$$\widetilde{\mu_k} = max\left(\underline{x}, min(\widehat{\mu_k}, \bar{x})\right), \widetilde{\sigma_k} = min(\widehat{\sigma_k}, \bar{x} - \underline{x}), \text{ where } \underline{x} = min(x_1, \ldots, x_k) \text{ and } \bar{x} = max(x_1, \ldots, x_k) \quad (8)$$

3.2.5 Phase III:

Similar to 3POD, SBOP adds n_2 points that optimize for quantiles of interest. 3POD and SBOP use the modified stochastic optimization procedure suggested by Joseph [11]. Previous criticism of Robbins-Monro-Joseph is that it may converge slowly when the algorithm starts far away from the true value. This is less of a concern in the SBOP and 3POD implementations given n_1 runs have already occurred. However, as noted by Joseph, "The optimal choice of a_n in small samples has not been investigated, although in most experiments this is the most interesting case." An area for future work is deriving guarantees for small values of n_2. This is particularly relevant in the face of perturbations between the target velocity and actual velocity. Additionally, it is reasonable to compare the performance of the stochastic optimization procedure against a sequential c-optimal design for the quantities of interest for various n_1 and n_2.

4. SIMULATION STUDY

The following simulation study was designed to answer two essential questions:

- Does the Bayesian methodology have potential and warrant further study?
- What is the impact when the priors are correct, i.e. have perfect information, or incorrect, i.e. are substantially different from the truth?

Broadly, two classes of scenarios were considered: one in which the priors and guesses were centred to be exactly on the true response curve (i.e. a given method is provided with oracle guesses), and a second in which they were set to be substantially far apart from the true curve. The response curve equation is shown in Equation 9, where p is the probability of CP. Note that there is an equivalence between the prior β_0 and β_1 for a given logistic response curve in a single variable and the location-scale parameterization with scale $=\frac{1}{\beta_0}$ and location $= -1 * \frac{\beta_1}{\beta_0}$. Code was run in R version 4.4.2 [25], with the base glm function used to calculate the maximum likelihood estimates (MLE) once separation had been broken, and MCMCpack [24] used when a Bayesian logistic regression estimate was needed. The test matrix is shown in

Table *1*.

$$log(p/(1-p)) = \beta_0 + \beta_1 * Velocity \tag{9}$$

Table 1: Test matrix for simulation study

Scenario	μ_{Guess}	σ_{Guess}	β_0 *True*	β_1 *True*	Velocity $\mu_{Perturb}$	Velocity $\sigma_{Perturb}$	μ_{Prior}	σ_{Prior}	n_1	n_2
Perfect Information, N = 40	2000	10	-200	0.1	0	5	2000	10	20	20
Perfect Information, N = 10	2000	10	-200	0.1	0	5	2000	10	8	2
Imperfect Information, N = 10	1800	50	-200	0.1	0	5	1800	50	8	2

For each method in each scenario, the scenario was run over 1000 iterations using the same true ballistic response curve shown in Figure 1. For each iteration:

1. Each method was followed to select the desired velocity for the starting and subsequent shots until $n_1 + n_2 = n$ shots were recorded
2. The desired shot velocity was perturbed according to a $N(0,5^2)$ distribution and the actual velocity was output
3. The outcome of PP/CP was drawn according to the true ballistic response curve for the actual velocity
4. The outcome was recorded, as were the desired and actual velocities

4.2 Results

The ultimate goal of an analysis of ballistics armour is to understand the behaviour of the protective gear under a variety of threats and conditions. While this is often simplified to the V50 of a given threat or condition (e.g. temperature), understanding the full ballistic response curve allows for more informed decisions to be made regarding warfighter and personnel safety.

To this end, several metrics were studied as part of this analysis:

- Percent of runs where a Zone of Mixed Results is identified (i.e. separation is broken)
- Median shot number where overlap is achieved
- Properties of β_0 and β_1 estimates
- Properties of V50, and V10 estimates
- Width of V50, V10, β_0 and β_1 80% intervals

While $V50$ is an oft-reported metric, there is increasing interest in the performance of protective gear at other quantiles, such as $V10$, which may be important metrics under alternative safety and performance thresholds; there is related interest in better understanding the behaviour of the whole curve as defined by β_0 and β_1. These curves can be expanded to include additional variables such as yaw, which would allow practitioners to model armour performance

https://doi.org/10.52202/081621-0039

under more complex conditions. For experimental efficiency, it is preferable to have a smaller number of runs where the ZMR was not achieved (e.g. break separation faster), and to have a smaller median test number at which the ZMR is found. It is also preferable to have a smaller median absolute deviation (MAD) for estimates such as $V50$ and β_0. There were several cases of simulation runs with outliers for all methods; for this reason, MAD was chosen as a more robust metric than standard deviation. In practical applications, these tests would have been discontinued. Unless otherwise stated, all final results are calculated using GLM to calculate the logistic regression MLE coefficients; the stochastic approximation method of 3POD $V50$ estimation is considered separately. For all tables, the bolded values indicate the best result.

4.2.1 Perfect Information Results

All methods performed fairly well in the case where our starting parameters, σ_{Guess} and μ_{Guess}, match the truth and there is sufficient sample size (N=40). Table 2 shows the MADs for the metrics of interest for N =40 and N = 10. While the regression coefficients for SBOP were much closer to the ground truth of β_0=-200 and β_1=0.1 for N = 10, the MAD for $V10$ and $V50$ estimates was generally slightly better for Neyer and Bruceton. All methods performed similarly for N = 40.

Table 2: Median Absolute Deviations for perfect information scenarios

MAD	N = 40				N = 10			
Method	β_0	β_1	V50	V10	β_0	β_1	V50	V10
Bruceton	**54.2**	**0.0272**	**6.09**	**2.3**	157	0.0788	11.8	**4.37**
SBOP	55.2	0.0277	6.42	2.69	**116**	**0.0578**	11.8	5.43
Neyer	57	0.0286	**6.09**	2.48	153	0.0762	**11.2**	5.7
3POD	54.9	0.274	6.19	2.52	187	0.0938	11.8	5.79

With N = 40, all methods broke separation. For N = 10, Bruceton broke separation most often at 91.8%, followed by 3POD, Neyer and SBOP, as shown in **Error! Reference source not found.**. This is not surprising as Bruceton performs well when given an appropriate step size; velocity selection is highly influenced by step size. Additionally, this is expected in the presence of perfect information as the other methods use the first two shots (x_1 = 1960 and x_2 = 2040) to establish the appropriate range while Bruceton is already iterating through finding $V50$ (x_1=2000).

We calculate the difference between the 90th percentile and the 10th percentile for β_0, β_1, V50 and V10 as a measure of how closely these metrics are clustered for 80% of the simulation runs. The results are shown in

Table ***3***. SBOP has the smallest intervals around β_0 and β_1 for N =40, while Bruceton and Neyer have the smallest intervals around V50 and V10, respectively. For N = 10, Neyer has the smallest intervals around β_0 and β_1, followed closely by SBOP. 3POD performs relatively poorly in this metric for N = 10, but does have the smallest interval compared to the other methods around V10. When given a larger sample (i.e. N = 40), SBOP returns the tightest 80% intervals around β_0 and β_1 and performs comparably to the best method when given a smaller sample (i.e. N = 10).

Table 3: 80% intervals for key values, perfect information

80% Intervals	N = 40				N = 10			
Method	β_0	β_1	V50	V10	β_0	β_1	V50	V10
Bruceton	229.33	0.11	**9.18**	20.76	708.10	0.35	**17.80**	32.78
SBOP	**217.96**	**0.11**	10.09	21.35	672.68	0.34	21.35	40.06
Neyer	265.63	0.13	9.31	**19.35**	**652.71**	**0.33**	21.23	32.80
3POD	256.20	0.13	9.53	20.85	1,331.29	0.67	21.70	**32.34**

One area in which results differed from Johnson et al [2] is the performance of 3POD when the stochastic search approximation was used as recommended by 3POD, rather than the MLE estimates. The primary difference is that the simulation study in this paper included a perturbation of the true versus desired velocity, consistent with conversations with field technicians. This implication is that in practice, the prescribed small step sizes may instead be large or in the opposite direction.

This is a challenge as the stochastic approximation algorithm relies on achieving an increasingly smaller step sizes (trending towards zero) between the target velocity and actual velocity, with a guarantee of asymptotically achieving the desired quantile. However, in the presence of perturbations between target and actual velocity, most steps will have sizes substantially different from zero. One option is to use the actual velocity rather than the target velocity; while this is still centered at a $V50 = 2000$, there is substantially more variation than an MLE estimate, as pictured in Figure 1 and described in

Table 3 (Right): Separation Statistics for N = 10

Table *4*.

Table 3 (Right): Separation Statistics for N = 10

	N = 10	
Method	**Percent of Runs with Overlap**	**Median Overlap Test Number**
Bruceton	**91.8%**	**6**
SBOP	83%	7
Neyer	85.2%	7
3POD	85.3%	7

Table 4 (Below): V50 estimates for 3POD

Scenario	MLE	Stochastic Approximation
Perfect Info, N = 40	2.52	7.75
Perfect Info, N = 10	5.86	8.52
Imperfect Info, N = 10	7.79	10.2

Figure 1: Baseline true response curve (Left), distribution of V50 estimates for 3POD (Right). Velocity in ft/sec

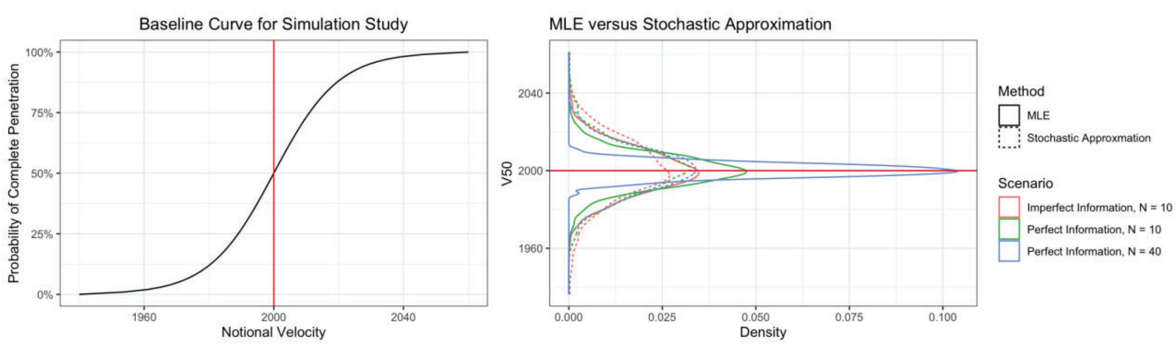

Table 6: Percent of simulation runs with overlap (Left)

Table 7: Median Absolute Deviation for imperfect information, N = 10 (Right)

MAD	N = 10			
Method	β_0	β_1	**V50**	**V10**
Bruceton	**50.1**	**0.248**	**9.86**	**7.37**
SBOP	79.8	0.04	17.6	8.97
Neyer	83.9	0.0418	19.9	12.4
3POD	74.6	0.0373	13.1	7.71

https://doi.org/10.52202/081621-0039

4.2.2 Distribution of V50 estimates for 3POD

The alternative option of using results from actual velocities and assigning them to the desired shot velocities is also problematic; for example, when at $V10$, perturbing the velocity up by 10 unit/second will not have an equal effect on the probability of a complete penetration as perturbing the velocity down by 10 units/second. Additional work is needed comparing the performance of a c-optimal design for various combinations of n_2 and variances.

4.2.3 Imperfect Information Results

In the case of imperfect information, SBOP does not perform well when both the μ_{Prior} or σ_{Prior} are incorrect and there is a small sample size (in our case, N = 10). When both the μ_{Prior} or σ_{Prior} are significantly incorrect, there is a cost in terms of how quickly overlap occurs, if it occurs at all. A prior standard deviation of σ^2=10^2 was used in MCMCpack function. The percent of runs where separation is broken and overlap is achieved is shown in **Error! Reference source not found.**.

In practice, the prior variance is generally a predefined weight or treated as a tuning parameter; this was not explored in this analysis. When the prior information is relatively weak, the prior variance on the tuning parameters can be increased. More work on the Bayesian method is needed to appropriately identify when the prior parameters appear to be incorrect, as this is valuable information to the experimenter and can be identified relatively early in the experiment.

Bruceton outperforms all other methods on the MAD of the variables under consideration in this scenario as shown in **Error! Reference source not found.**. Neyer is broadly accepted as a superior and more efficient methodology to Bruceton; it is unclear if Bruceton's superior performance with regards to achieving overlap is an artifact of the specific parameters tested, or a robustness to the differences in target and actual velocity. Future research will investigate if Bruceton is in fact more robust to velocity perturbations with different pairs of input parameters. Note that SBOP and 3POD perform similarly for β_0, β_1, and $V50$ while SBOP performs similarly to Neyer for $V10$. SBOP's performance is in line with other standard methods for key metrics when given imperfect starting parameters, which demonstrates the viability of the proposed approach.

Again, we calculate the difference between the 90th percentile and the 10th percentile for β_0, β_1, $V50$ and $V10$ as a measure of how closely these metrics are clustered for 80% of the simulation runs. The results are shown in **Error! Reference source not found.**. With a small sample size and imperfect starting parameters, Neyer and SBOP return β_0 and β_1 values within a smaller range than the other methods. Neyer has the tightest interval followed closely by SBOP for β_0, and β_1. 3POD has the tightest interval around V50 and Neyer has the tightest interval around $V10$.

Method	Percent of Runs with Overlap
Bruceton	**34%**
SBOP	16.3%
Neyer	15.2%
3POD	26%

Table 8: 80% intervals for key values, imperfect information

80% Intervals Method	N = 10			
	β_0	β_1	V50	V10
Bruceton	516.57	0.26	33.18	47.92
SBOP	111.55	0.06	40.33	48.39
Neyer	**107.74**	**0.05**	34.77	**43.43**
3POD	503.03	0.25	**31.83**	45.88

5. CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

This paper demonstrates the potential to use Bayesian methodology in ballistics testing, in order to allow for the incorporation of prior information in the selection of new test points without impacting the final analysis. Additionally, this is one of the first paper to incorporate velocity perturbation into a simulation study [2]; Note that the Institute for

Defense Analyses enforced a maximum bound on perturbation size, while no limit was applied in this study. Additional work should analyse the sensitivity of methods to the size of the perturbation, as conversations with technicians indicated there can be substantial variability in the variance based on factors such as threat, season, and day to day variance. Under some circumstances, this methodology breaks separation faster than Frequentist methods, although further development is needed. Specifically, while the coefficients for both β_0 and β_1 were substantially closer to the true values, this did not translate to a lower MAD for $V50$ and $V10$. This is a promising result, and it is likely that future iterations of SBOP can correct this issue.

5.2 Future Work

There are multiple avenues for future work. For work relating specifically to SBOP, the following are of particular interest:

- In Phase I, initial shots are based on estimates of μ_{Guess} and σ_{Guess}. However, these shots are not defined as optimal by either a design or risk-adjusted method. Given priors for a response curve, it would be possible to estimate the probability that multiple shots are partial penetrations or complete penetrations, and to adjust accordingly.
- In Phase I and Phase II, the priors play a role in determining the selecting subsequent shots. There is robust literature on the impact of tuning various parameters, such as the precision. However, this has not been done specifically in the ballistics testing space. Future work includes implementing this method in live-fire testing to understand the sensitivity and impact of these parameters.
- In Phase III, SBOP follows the stochastic approximation method proposed by [11] and used in 3POD. While this has desirable asymptotic properties, guarantees do not exist for small n_1 and n_2, and it is unclear how a c-optimal design would compare. A separate study would robustly analyze if a method was dominant for particular combinations of n_1 and n_2, particularly when more than one quantile is of interest.

Additionally, more analysis is needed to better understand the robustness of existing methods against the impact of velocity perturbation between the target and actual velocity is needed. Another area stems from discussion with practitioners, who note that for some threat and velocity combinations, the variance in velocity is heterogeneous. Expanding the design to account for cases where it is known that the shot variance is not homogeneous would be quite useful in these cases.

Acknowledgements

The authors would like to thank the PEO Soldier Product Manager Soldier Protective Equipment for sponsoring this effort. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of PEO Soldier or Naval Sea Systems Command (NAVSEA).

References

[1] *Ballistic Resistance of Body Armour, NIJ Standard 0101.07*, National Institute of Justice, 2023.
[2] T. H. Johnson, L. Freeman, J. Hester, and J. L. Bell, "A Comparison of Ballistic Resistance Testing Techniques in the Department of Defense," *IEEE Access*, vol. 2, 2014, pp. 1442–1455.
[3] S. G. Gilmour and L. A. Trinca, "Optimum Design of Experiments for Statistical Inference," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 61, no. 3, May 2012, pp. 345–401.
[4] *V50 Ballistic Test for Armour*. 1997.
[5] *Ballistic Test Method for Personal Armour Materials and Combat Clothing*. NATO Standardization Agency, 2003.
[6] M. E. et al. Greene, *The Next Revision of the NIJ Performance Standard for Ballistic Resistance of Body Armour, NIJ Standard 0101.07: Changes to Test Methods and Test Threats.* 2018.
[7] S. Magnan, G. Pageau, and A. Bouamoul, "Beyond V50: A More Comprehensive and Efficient Methodology for Assessing Armour Performance," Personal Armour Systems Symposium, Dresden, Germany, Sep. 2023.
[8] W. J. Dixon and A. M. Mood, "A Method for Obtaining and Analyzing Sensitivity Data," *Journal of the American Statistical Association*, vol. 43, no. 241, 1948, pp. 109–126.
[9] H. Robbins and S. Monro, "A Stochastic Approximation Method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, Sep. 1951, pp. 400–407.
[10] V. Dupac and U. Herkenrath, "Stochastic Approximation With Delayed Observations."
[11] V. R. Joseph, "Efficient Robbins-Monro procedure for binary data," *Biometrika*, vol. 91, no. 2, Jun. 2004, pp. 461–470.

https://doi.org/10.52202/081621-0039

[12] H. J. Langlie, "Proceedings of the Eighth Conference on the Design of Experiments in Army Research Development and Testing," in *Technical Report U-1792, Third edition*, 1962, pp. 146–165.
[13] C. F. J. Wu and M. Hamada, *Experiments: Planning, Analysis, and Optimization*, 1st ed. John Wiley & Sons, Inc., 2021.
[14] B. T. Neyer, "A D-Optimality-Based Sensitivity Test," *Technometrics*, vol. 36, no. 1, 1994, pp. 61–70.
[15] C. F. J. Wu and Y. Tian, "Three-phase optimal design of sensitivity experiments," *Journal of Statistical Planning and Inference*, vol. 149, Jun. 2014, pp. 1–15.
[16] D. Wang, Y. Tian, and C. F. J. Wu, "Comprehensive Comparisons Of Major Sequential Design Procedures For Sensitivity Testing," *Journal of Technical Quality*, vol. 52, no. 2, Apr. 2020, pp. 155–167.
[17] P. Roediger, "GONOGO: An R Implementation of Test Methods to Perform, Analyze and Simulate Sensitivity Experiments," U.S. Army Armament Research, Development and Engineering Center, Picatinny Arsenal, New Jersey, Technical Report AREIS-CR-18007, 2018.
[18] W. Dumouchel and B. Jones, "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model," vol. 36, no. 1, 1994.
[19] H. A. Dror and D. M. Steinberg, "Sequential Experimental Designs for Generalized Linear Models," *Journal of the American Statistical Association*, vol. 103, no. 481, Mar. 2008, pp. 288–298.
[20] A. Sarma and M. Kay, "Prior Setting in Practice: Strategies and Rationales Used in Choosing Prior Distributions for Bayesian Analysis," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–12.
[21] M. Zondervan-Zwijnenburg, M. Peeters, S. Depaoli, and R. Van De Schoot, "Where Do Priors Come From? Applying Guidelines to Construct Informative Priors in Small Sample Research," *Research in Human Development*, vol. 14, no. 4, Oct. 2017, pp. 305–320.
[22] S. W. Hyun and W. K. Wong, "Multiple-Objective Optimal Designs for Studying the Dose Response Function and Interesting Dose Levels," *The International Journal of Biostatistics*, vol. 11, no. 2, Jan. 2015.
[23] T. Zhou and Y. Ji, "On Bayesian Sequential Clinical Trial Designs," *The New England Journal of Statistics in Data Science*, 2024, pp. 136–151.
[24] A. D. Martin, K. M. Quinn, and J. H. Park, "MCMCpack: Markov Chain Monte Carlo (MCMC) Package." Comprehensive R Archive Network, Feb. 21, 2003, pp. 1.7–1.
[25] R. C. Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2021.