JOURNAL OF STATISTICAL THEORY AND PRACTICE
2017, VOL. 11, NO. 2, 339–360
http://dx.doi.org/10.1080/15598608.2017.1299055

Minimum aberration designs for discrete choice experiments

Jessica Jaynes[a], Hongquan Xu[b], and Weng Kee Wong[c]

[a]Department of Mathematics, California State University, Fullerton, Fullerton, California, USA; [b]Department of Statistics, University of California, Los Angeles, California, USA; [c]Department of Biostatistics, University of California, Los Angeles, California, USA

ABSTRACT

A discrete choice experiment (DCE) is a survey method that gives insight into individual preferences for particular attributes. Traditionally, methods for constructing DCEs focus on identifying the individual effect of each attribute (a main effect). However, an interaction effect between two attributes (a two-factor interaction) better represents real-life trade-offs, and provides us a better understanding of subjects' competing preferences. In practice it is often unknown which two-factor interactions are significant. To address the uncertainty, we propose the use of minimum aberration blocked designs to construct DCEs. Such designs maximize the number of models with estimable two-factor interactions in a DCE with two-level attributes. We further extend the minimum aberration criteria to DCEs with mixed-level attributes and develop some general theoretical results.

ARTICLE HISTORY
Received 2 February 2016
Accepted 21 February 2017

KEYWORDS
Aliasing; blocked design; estimation capacity; fractional factorial design; multinomial logit model; orthogonal array

AMS SUBJECT CLASSIFICATION
62K15; 62J15; 62K10

1. Introduction

A discrete choice experiment (DCE) is a survey method used to quantify subject preferences for various attributes, and to gain insight into how attributes influence subject choices. In a DCE, subjects are offered choice sets that contain questions. Each choice set is made up of several options and each option is made up of several attributes with two or more levels. Subjects are asked to select a single option in each choice set. Grossmann and Schwabe (2015) reviewed various designs for constructing DCEs and Lancsar and Louviere (2008) and Johnson et al. (2013) provided checklists for good research practice in conducting a DCE.

DCEs combine ideas from economic theory with experimental design. In each choice set, the option chosen by the subject is assumed to have the highest utility, where the utility is the benefit that the subject experiences by selecting a particular option. In this article, we assume the option chosen by the subject implies an implicit trade-off between attributes, and the responses from a DCE are modeled using the multinomial logit (MNL) model, which is the sum of two parts: (1) an explainable systematic component based on the observed attributes and (2) a nonexplainable random component that captures other attributes that may be relevant but not specified. The parameters in the MNL model provide information on the relative importance of each attribute (i.e., its main effect) or its interaction with other attributes (i.e., its two-factor interaction).

CONTACT Jessica Jaynes jjaynes@fullerton.edu Department of Mathematics, California State University, Fullerton, Fullerton, CA 92831, USA.

The design of the DCE plays a critical role because it determines how the attributes and their levels are combined to form choice sets. One common method is to use fractional factorial designs (FFDs) to construct DCEs; see Street and Burgess (2007) and Bush (2014), among others. These designs are based on a starting design that is either a full factorial or an FFD, for which the entries represent the first option in each choice set. Generators are then added component-wise to the starting design to form the remaining options in each choice set. These methods are flexible for constructing DCEs for estimating main effects only. Because both main effects and two-factor interactions can jointly determine whether DCEs are successfully used to accurately assess real-life decision-making processes, designs that can also accurately estimate two-factor interactions are more desirable. Our work in this article focuses on constructing more effective designs for estimating main effects and two-factor interactions simultaneously.

Jaynes et al. (2016) proposed using existing blocked fractional factorial designs (BFFDs) to construct DCEs for estimating main effects and select two-factor interactions. Such an approach assumes that it is known in advance which two-factor interactions are significant. This is problematic as significant interactions are often unknown in practice. Here we take an alternative approach and propose using minimum aberration (MA) criteria for selecting BFFDs to construct DCEs. MA designs are model robust and tend to have large capacity in estimating various models involving two-factor interactions (Mukerjee and Wu 2006; Cheng 2014). We review several MA criteria for comparing BFFDs and present examples to illustrate the benefits of MA designs. Our main innovations are to extend the MA criteria to DCEs with mixed-level attributes and to develop some general theoretical results.

Section 2 briefly reviews two-level FFDs and BFFDs. Section 3 describes how BFFDs can be used to construct DCEs for the MNL model and a simulation study to compare DCEs constructed from different BFFDs. In section 4, we introduce the MA criteria, present examples to illustrate the advantages of MA designs, and justify the MA criteria in the concept of estimation capacity. Section 5 discusses construction methods and section 6 shows how to construct DCEs with mixed-level attributes under generalized MA criteria. Section 7 offers a summary and a discussion on the use of MA BFFDs to construct DCEs.

2. Fractional and blocked fractional factorial designs

A FFD with k two-level attributes is said to be a $2^{-p}th$ fraction of the full 2^k design if it has 2^{k-p} runs. The fraction is determined by p defining words, where a word describes the relationship between columns in an FFD. The p defining words, and their products, form the treatment defining contrast subgroup (Wu and Hamada 2009). In the treatment defining contrast subgroup, there are $2^p - 1$ distinct words plus the identity, where each element within the treatment defining contrast subgroup is called a word, except the identity. The number of letters in a word is called its length. The length of the shortest word in the treatment defining contrast subgroup is the *resolution* of the design. The larger the resolution, the better is the design. The resolution of a design also determines which effects can be identified. Let $A_{i,0}$ be the number of words of length i ($i = 1, \ldots, k$) in the treatment defining contrast subgroup, such that $\sum_{i=1}^{k} A_{i,0} = 2^p - 1$. Then the vector $W_t = (A_{1,0}, A_{2,0}, \ldots, A_{k,0})$ is called the treatment wordlength pattern. In practice, we use designs with resolution III or higher with $A_{1,0} = A_{2,0} = 0$, so for simplicity, we write $W_t = (A_{3,0}, \ldots, A_{k,0})$.

Example 1. Suppose we have $k = 5$ two-level attributes and we wish to construct a half fraction from the full 2^5 factorial design. We set $p = 1$ and we determine the fraction by specifying a defining relation, say $E = ABCD$. Since there is only a single defining word for this design, this defining word forms the treatment defining contrast subgroup $I = ABCDE$. This results in $2^p - 1 = 2^1 - 1$ distinct words plus the identity I. Since the only word in the treatment defining construct subgroup is of length five, this design is said to have resolution V with treatment wordlength pattern $W_t = (A_{3,0}, A_{4,0}, A_{5,0}) = (0, 0, 1)$.

In experimental design, BFFDs are commonly used for reducing systematic variations and increasing precision of parameter estimation. To construct a BFFD we confound an interaction effect with a block effect. This means that the design is unable to estimate the two effects separately. To construct a two-level BFFD, we block a 2^{k-p} FFD in 2^q blocks defined by q block defining words, with blocks of size 2^{k-p-q}, which leads to two defining contrast subgroups: the treatment defining contrast subgroup and the block defining contrast subgroup. The q block defining words and their products form the block defining contrast subgroup, which consists of $2^q - 1$ distinct words.

Any effects, including any aliased effects, associated with these blocking variables are confounded with the blocks (Wu and Hamada 2009). This means that if an effect is confounded with a block effect, it cannot be estimated, and if an effect is aliased with another effect (not a block effect), it can be estimated only if all the aliased effects are negligible. A main effect or a two-factor interaction is *clear* in a BFFD if it is not aliased with any other main effects or two-factor interactions, or confounded with any block effects (Wu and Hamada 2009). A clear main effect or two-factor interaction can be estimated without having to assume negligibility of other two-factor interactions that may be of interest.

In a BFFD, each block effect is confounded with 2^p treatment words (or effects). Let $A_{i,1}$ be the number of treatment words of length i that are confounded with a block effect, such that $\sum_{i=1}^{k} A_{i,1} = 2^p(2^q - 1)$. Then the vector $W_b = (A_{1,1}, A_{2,1}, \ldots, A_{k,1})$ is called the block wordlength pattern. However, a blocking scheme is only feasible if none of the main effects are confounded with block effects, that is, $A_{1,1} = 0$, and we write $W_b = (A_{2,1}, \ldots, A_{k,1})$.

Example 2. Suppose we wish to divide the 2^{5-1} FFD in Example 1 into $2^q = 2^2$ blocks, each of size $2^{k-p-q} = 2^{5-1-2}$. This design has treatment defining contrast subgroup $I = ABCDE$ and we can choose the block defining contrast subgroup $b_1 = AB, b_2 = AC,$ and $b_3 = b_1 b_2 = BC$ (which consists of $2^q - 1 = 2^2 - 1$ distinct words). With this design, additional effects are confounded with the three block effects. For instance, when we multiply the treatment defining contrast subgroup $I = ABCDE$ with the block defining contrast subgroup, we obtain $b_1 = AB = CDE, b_2 = AC = BDE,$ and $b_3 = b_1 b_2 = BC = ADE$. In this design, all five main effects are clear plus seven two-factor interactions. This design has block wordlength pattern: $W_b = (A_{2,1}, A_{3,1}, A_{4,1}, A_{5,1}) = (3, 3, 0, 0)$; that is, three two-factor interactions (AB, AC, BC) and three three-factor interactions (CDE, BDE, ADE) are confounded with block effects.

3. Blocked fractional factorial designs for discrete choice experiments

A main advantage of using BFFDs to construct a DCE is that the entire aliasing structure of a BFFD is known in advance and consequently we also know which effects are estimable in the DCE (Jaynes et al. 2016). The number of blocks in the BFFD represents the number of choice sets in the DCE, and the size of the block represents the number of options within each choice set. If we use a 2^pth fraction of a 2^k experiment in 2^q blocks, the number of choice sets in a DCE is 2^q and the number of options in each choice set is 2^{k-p-q}.

3.1. Multinomial logit model

The multinomial logit (MNL) model is a common model for modeling responses and analyzing data from a DCE. The parameters in the model measure the usefulness of the attributes and their interactions with other attributes. Specifically, suppose the DCE has S choice sets and J options in each choice set. We assume that the responses from the subjects are analyzed using random utility theory and define the utility for a subject that chooses option j in choice set s to be

$$U_{js} = \boldsymbol{x}'_{js}\beta + \epsilon_{js}. \tag{1}$$

Here $\mathbf{x}_{js}$ is a $k^* \times 1$ vector containing the model expansion of the attribute levels of option j in choice set s, k^* is the number of parameters to be estimated, β is the $k^* \times 1$ vector of model parameters representing the effect of the attribute levels on the utility and ε_{js} is an error term following an independent identically distributed extreme value type 1 distribution.

Under the MNL model, the probability that a subject selects option j in choice set s is

$$\mathrm{p}_{js} = \frac{e^{(\boldsymbol{x}'_{js}\beta)}}{\sum_{r=1}^{J} e^{(\mathbf{x}'_{rs}\beta)}}, \tag{2}$$

where β is estimated using maximum likelihood estimation. It is assumed in the MNL model that β is the same for every subject and that subjects' preferences for the attribute levels are homogeneous across the population (Kessels et al. 2011). We also assume all subjects are given the same choice sets and the choice of the option in each choice set is independent because the errors are assumed to be independent. The log-likelihood function for the MNL model is

$$l(\beta) = \sum_{s=1}^{S}\sum_{j=1}^{J} y_{js}\log(p_{js}), \tag{3}$$

where y_{js} is a choice indicator, which equals 1 if the subject chooses option j in choice set s, and zero otherwise (Gerard et al. 2008).

The optimal design $\mathbf{X} = [\mathbf{x}'_{js}]$ for estimating β in the MNL model depends on the Fisher information matrix (Kessels et al. 2011). This matrix is the covariance of the derivative of the log-likelihood function with respect to β (Sandor and Wedel 2001):

$$\mathbf{M}(\mathbf{X}, \beta) = \sum_{s=1}^{S} \mathbf{X}'_s(\mathbf{P}_s - \mathbf{p}_s\mathbf{p}'_s)\mathbf{X}_s, \tag{4}$$

where $\mathbf{X}_s = [\mathbf{x}_{1s}, \ldots, \mathbf{x}_{Js}]'$ is a submatrix of $\boldsymbol{X}$ that corresponds to choice set s, $\mathbf{p}_s = [p_{1s}, \ldots, p_{Js}]'$ and $\mathbf{P}_s = diag[p_{1s}, \ldots, p_{Js}]$. When all subjects are shown the same choice sets, $\mathbf{X}_s$ is the same for all subjects. If the information matrix is diagonal, estimates of the parameters are uncorrelated.

The design with the largest determinant of the information matrix $\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{\beta})$ is said to be the D-optimal design. Such an optimal design provides the most precise estimates for the model parameters (Atkinson and Donev 1992). However, the optimal design depends on the unknown parameter β, so design strategies cannot be implemented unless the parameters are known. One approach to overcome this problem is to construct a locally optimal design assuming nominal values for the parameters are available from pilot studies or experts' opinion.

We construct locally D-optimal designs and assume that each option has an equal probability of selection, that is, the nominal values are $\beta = \mathbf{0}_{k^*}$, where $\mathbf{0}_{k^*}$ is a $k^* \times 1$ vector of zeros. When $\beta = \mathbf{0}_{k^*}$, the information matrix for a locally optimal choice design under the MNL model is proportional to the information matrix for a BFFD with blocks of size J under the general linear model (Kessels et al. 2011). Consequently, "locally optimal DCEs obtained assuming $\beta = \mathbf{0}_{k^*}$ are exactly the same as optimal designs for blocked experiments when the model of interest is linear and the block effects are treated as fixed parameters" (Kessels et al. 2011, 176).

3.2. *Simulation study*

We now perform a simulation study using various DCEs constructed from different BFFDs to compare estimates for the parameters in the MNL model. We consider a DCE with five two-level attributes, and four choice sets each with four options. The three designs labeled S1, S2, and S3 in Table 1 are 2^{5-1} FFDs in 2^2 blocks taken from Table 4 in Sun et al. (1997). Each design has a different treatment defining construct subgroup and block defining words, which leads to different treatment, and block, wordlength patterns. Table 1 also displays the two-factor interactions confounded with block effects, the aliasing structure between main effects and two-factor interactions, and the aliasing structure between two-factor interactions. S1 is the same design used in Example 2. For the simulation study, we assume that the true model has five main effects plus three two-factor interactions given by

$$\mu = 0.5x_A - 0.5x_B + 0.5x_C - 0.5x_D + 0.5x_E + 0.25x_Ax_C - 0.25x_Ax_D + 0.25x_Bx_E, \tag{5}$$

where μ is the utility for the option (A, B, C, D, E). We first compute the MNL probability of selecting each option within each choice set using Eq. (2). We then use these probabilities to simulate a response according to the multinomial distribution for each of the three designs. Each DCE is replicated 500 times to represent 500 subjects.

To illustrate the consequences of confounding and aliasing, for each design, we fit two models: (i) a model with main effects only, and (ii) a model with all main effects and all clear two-factor interactions plus one two-factor interaction from each aliased set that is not confounded with a block effect. Tables 2 and 3 show the parameter estimates and standard errors, respectively, from each design. We observe from Table 2 that parameter

Table 1. Three BFFDs used in the simulation study.

Design	Treatment defining words	Block defining words	W_t	W_b	Confounded effects	Aliased effects
S1	$I = ABCDE$	$b_1 = AB, b_2 = AC, b_3 = BC$	$(0, 0, 1)$	(3,3,0,0)	$b_1 = AB, b_2 = AC, b_3 = BC$	—
S2	$I = ABCE$	$b_1 = ACD, b_2 = BCD, b_3 = AB$	$(0, 1, 0)$	(2,4,0,0)	$b_3 = AB = CE$	$AC = BE, AE = BC$
S3	$I = ABE$	$b_1 = AC, b_2 = ABCD, b_3 = BD$	$(1, 0, 0)$	(2,3,1,0)	$b_1 = AC, b_3 = BD$	$A = BE, B = AE, E = AB$

Note. $W_t = (A_{3,0}, A_{4,0}, A_{5,0})$ and $W_b = (A_{2,1}, A_{3,1}, A_{4,1}, A_{5,1})$. The confounded and aliased effects focus on main effects and two-factor interactions.

estimates from the model with only main effects are not consistent with the coefficients in the true model (5) because the model does not contain the significant two-factor interactions in the true model (5). Under the MNL model, these missing significant two-factor interactions bias the estimates of the main effects even though all main effects are clear for designs S1 and S2, and C and D are clear for design S3. This would not be possible in a linear model.

For design S1, three two-factor interactions (AB, AC, BC) are confounded with the three block effects and cannot be estimated. The remaining seven two-factor interactions are clear. Comparing the parameter estimates from design S1 in Table 3 with those from the true model (5), we observe that all five main effects are consistent with the coefficients in Eq. (5). Both AD and BE, which are included in Eq. (5), are also consistent because they are clear in the BFFD. However, AC, which is included in Eq. (5), cannot be estimated because it is confounded with the block effect b_2. Even though AC cannot be estimated it does not bias the estimates of the clear effects in the true model (5).

For design S2, four two-factor interactions (AD, BD, CD, DE) are clear. The remaining six two-factor interactions form three alias sets. One of the alias set is confounded with b_3, and cannot be estimated. We include one two-factor interaction (say AC, AE) in the model from each of the other two alias sets. Comparing the parameter estimates for design S2 in Table 3 with the true model (5), we see that all five main effects are consistent with the coefficients in Eq. (5). The estimate for the two-factor interaction AD, which is included in Eq. (5), is consistent with the coefficient in Eq. (5) because AD is clear in the BFFD. Since AC and BE are aliased, and are both included in Eq. (5), in Table 3 the estimate for AC is the sum of the estimates for AC and BE in the the true model (5).

Table 2. Main effect estimates (and standard errors) from the simulation study.

Effect	Design S1	Design S2	Design S3
A	0.604 (0.032)	0.772 (0.033)	0.889 (0.041)
B	−0.455 (0.032)	−0.391 (0.033)	−0.567 (0.040)
C	0.509 (0.032)	0.609 (0.029)	0.471 (0.032)
D	−0.557 (0.027)	−0.502 (0.026)	−0.606 (0.028)
E	0.387 (0.026)	0.341 (0.029)	0.512 (0.037)

Note. True model: $\mu = 0.5x_A - 0.5x_B + 0.5x_C - 0.5x_D + 0.5x_E + 0.25x_Ax_C - 0.25x_Ax_D + 0.25x_Bx_E$.

Table 3. Main effects plus two-factor interactions (and standard errors) from the simulation study.

Effect	Design S1	Design S2	Design S3
A	0.506 (0.048)	0.555 (0.044)	0.792 (0.051)
B	−0.524 (0.048)	−0.503 (0.045)	−0.533 (0.051)
C	0.549 (0.048)	0.464 (0.044)	0.497 (0.051)
D	−0.484 (0.048)	−0.435 (0.046)	−0.456 (0.041)
E	0.454 (0.048)	0.482 (0.045)	0.502 (0.051)
AB	—	—	—
AC	—	0.467 (0.045)	—
AD	−0.246 (0.048)	−0.28 (0.038)	−0.262 (0.039)
AE	0.028 (0.048)	−0.025 (0.041)	—
BC	—	—	0.029 (0.050)
BD	0.038 (0.048)	0.039 (0.045)	—
BE	0.239 (0.048)	—	—
CD	−0.005 (0.048)	−0.015 (0.038)	0.011 (0.029)
CE	0.012 (0.048)	—	−0.053 (0.051)
DE	−0.017 (0.048)	−0.041 (0.045)	−0.057 (0.041)

Note. True model: $\mu = 0.5x_A - 0.5x_B + 0.5x_C - 0.5x_D + 0.5x_E + 0.25x_Ax_C - 0.25x_Ax_D + 0.25x_Bx_E$.

For design S3, two two-factor interactions (AC, BD) are confounded with two block effects, b_1 and b_3, and cannot be estimated. Three two-factor interactions (BE, AE, AB) are aliased with main effects and cannot be estimated either. The remaining five two-factor interactions are clear. Comparing the parameter estimates of the main effects from design S3 in Table 3 with those from the true model (5), we observe that only four of the main effects are consistent with the coefficients in Eq. (5). The main effect A, which is aliased with BE, is biased by BE because BE is included in the true model (5). Comparing the estimates of the two-factor interactions from design S3 in Table 3 with those from the true model (5), AD is consistent because AD is clear in the BFFD. Since AC is confounded with the block effect b_1, AC cannot be estimated.

From this simulation study, we have shown that a misspecified model can lead to biased and misleading estimates, even if the effects are clear. This illustrates the importance of including all significant effects in the model, particularly significant two-factor interactions. For example, if there are significant two-factor interactions (such as in our true model), and a main effects only model is fit, then the estimates of the main effects are biased by the significant two-factor interactions, even if the main effects are clear. By considering a BFFD, we present the following advantages: (1) Effects confounded with block effects are not estimable, but do not bias the estimate of other effects; (2) aliasing causes bias, but aliased effects are estimable if all the aliases are negligible; and (3) aliasing or missing a significant two-factor interaction can bias the estimation of main effects even if all main effects and two-factor interactions are clear. Hence, it is essential at the design stage to know the aliasing and confounding structure of the designs in order to construct an efficient DCE.

Viney et al. (2005), Bliemer and Rose (2011), and Burgess et al. (2011; 2015) reported some empirical comparisons of DCEs and concluded that the choice of designs is not as crucial when the sample size is reasonable. When the sample size becomes smaller, the choice of designs matters more. Our simulation shows that the three designs differ substantially when some two-factor interactions are included in the true model, even though they are equally good when the true model contains the main effects only. In the next section, we propose the MA criteria for choosing BFFDs to construct DCEs.

4. Minimum aberration criteria

The choice of the BFFD for constructing a DCE depends on the number of attributes k, the desired size of the choice set or the number of options, and which effects are to be identified as clear. Sun et al. (1997), Sitter et al. (1997), Chen and Cheng (1999), Cheng and Mukerjee (2001), Cheng and Wu (2002), Xu (2006), Xu and Lau (2006), and Xu and Mee (2010), among others, discussed optimal choice of blocking schemes for FFDs. Jaynes et al. (2016) focused on the choice of BFFDs to maximize the number of clear main effects and two-factor interactions. This method is beneficial if it is known in advance which two-factor interactions are significant; however, in practice, it is not known in advance which two-factor interactions are significant. One way to select a BFFD is to use the total number of clear effects to compare and rank order the different blocked 2^{k-p} designs. However, this is not always the best approach because it depends on the aliasing structure of the designs being compared.

In this article, we propose the use of the minimum aberration (MA) criteria to select BFFDs to construct efficient DCEs, which maximizes the number of models with estimable two-factor interactions by minimizing the confounding or aliasing of two-factor interactions. There are various approaches for applying MA criteria to select a BFFD because of the presence of the two defining contrast subgroups, one for the treatment effects and one for the block effects. One approach is to apply the MA criterion to the treatment and block wordlength patterns separately; however, an MA design with respect to one wordlength pattern may not have MA with respect to the other wordlength pattern. Another approach is to combine the treatment and block wordlength patterns into one sequence and apply the MA criterion to the combined wordlength pattern. With this approach, the MA criterion ranks BFFDs according to their combined treatment and block wordlength patterns. Several combined wordlength patterns have been proposed in the literature:

$$W_{scf} = (A_{3,0}, A_{2,1}, A_{4,0}, A_{3,1}, A_{5,0}, A_{4,1}, \ldots) \tag{6}$$

$$W_{cc} = (3A_{3,0} + A_{2,1}, A_{4,0}, 10A_{5,0} + A_{3,1}, A_{6,0} \ldots) \tag{7}$$

$$W_1 = (A_{3,0}, A_{4,0}, A_{2,1}, A_{5,0}, A_{6,0}, A_{3,1}, \ldots) \tag{8}$$

$$W_2 = (A_{3,0}, A_{2,1}, A_{4,0}, A_{5,0}, A_{3,1}, A_{6,0}, \ldots). \tag{9}$$

These sequences were proposed by Sitter et al. (1997), Chen and Cheng (1999), and Cheng and Wu (2002). Based on these combined wordlength patterns, several authors have provided collections and tables of MA BFFDs based on the W-criteria for both two and three-level attributes:

- Sitter et al. (1997): provide MA BFFDs based on the W_{scf} criterion for all 8 and 16 run designs; for 32 run designs up to 15 attributes, and for 64 and 128 run designs up to 9 attributes.
- Chen and Cheng (1999): provide MA BFFDs based on the W_{cc} criterion for 8, 16, and 32 runs up to 19 attributes.
- Cheng and Wu (2002): provide MA BFFDs based on the W_1 and W_2 criteria for all 27 run designs, and for 81 run designs up to 10 attributes.
- Xu and Lau (2006) and Xu (2006): provide MA BFFDs based on the W_{scf}, W_1, W_2, and W_{cc} criteria for all 32 run designs, for all 81 run designs, and for 64 runs up to 32 attributes.
- Xu and Mee (2010): provide MA BFFDs based on the W_1 criterion for 128 runs and up to 64 attributes.

Several authors have compared and commented on the advantages and disadvantages of the four sequences (6)–(9); see Chen and Cheng (1999), Zhang and Park (2000), Cheng and Wu (2002), and Xu and Mee (2010). Xu and Lau (2006) and Xu (2006) summarized the situations in which MA BFFDs differ under the different criteria (6)–(9). Cheng and Wu (2002) argued that both W_1 and W_2 are appropriate sequences because they allow for a large number of two-factor interactions to be estimated. The following example illustrates the benefits of using the W_1 and W_2 criteria for selecting BFFDs to construct DCEs.

Example 3. Consider the three designs S1, S2, and S3 in our simulation study and two of the three designs used in the simulation study by Jaynes et al. (2016). For reference, we denote the other two designs as S4 and S5. S4 has treatment defining word $I = ABCE$ and block defining words $b_1 = AB = CE$, $b_2 = AC = BE$, and $b_3 = BC = AE$; S5 has treatment defining word $I = ADE$ and block defining words $b_1 = AB = BDE$, $b_2 = AC = CDE$, and $b_3 = BC = ABCDE$. A direct calculation shows that we have:

- Design S1: $W_1 = (0, 0, 3, 1, \ldots)$, $W_2 = (0, 3, 0, 1, 3, \ldots)$.
- Design S2: $W_1 = (0, 1, 2, 0, \ldots)$, $W_2 = (0, 2, 1, 0, 4, \ldots)$.
- Design S3: $W_1 = (1, 0, 2, 0, \ldots)$, $W_2 = (1, 2, 0, 0, 3, \ldots)$.
- Design S4: $W_1 = (0, 1, 6, 0, \ldots)$, $W_2 = (0, 6, 1, 0, 0, \ldots)$.
- Design S5: $W_1 = (1, 0, 3, 0, \ldots)$, $W_2 = (1, 3, 0, 0, 2, \ldots)$,

Both S3 and S5 have one word of length three ($A_{3,0} = 1$), which causes three two-factor interactions aliased with three main effects. They are worse than designs S1, S2, and S4 in terms of both W_1 and W_2. Both S2 and S4 have no words of length three ($A_{3,0} = 0$) and one word of length four ($A_{4,0} = 1$), but in design S2 two two-factor interactions are confounded with block effects ($A_{2,1} = 2$), while in design S4 six two-factor interactions are confounded with block effects ($A_{2,1} = 6$). Therefore, S2 is better than S4 in terms of both W_1 and W_2. Design S1 has MA with respect to W_1 because it has smaller $A_{4,0}$ than design S2 (0 vs. 1), which implies that in design S1 no two-factor interactions are aliased with other two-factor interactions, whereas in design S2 three sets of two-factor interactions are aliased with other two-factor interactions caused by one word of length four. Design S2 has MA with respect to W_2 because it has smaller $A_{2,1}$ than design S1 (2 vs. 3), which implies that two two-factor interactions are confounded with block effects in design S2 versus three two-factor interactions confounded with block effects in design S1. Both S1 and S2 are better than the other three designs in the capacity of estimating two-factor interactions.

Example 3 illustrated that by minimizing aliasing and confounding of two-factor interactions, we maximize the number of estimable two-factor interactions. The $A_{3,0}$ value captures the number of two-factor interactions aliased with main effects; the $A_{2,1}$ value captures the number of two-factor interactions confounded with block effects. By minimizing $A_{3,0}$ and $A_{2,1}$, we maximize the number of estimable two-factor interactions besides the estimation of main effects. This is further described by the concept of estimation capacity later.

The choice between the W_1 and W_2 criteria depends on whether aliased effects or confounded effects are viewed as less desirable. For resolution III and IV FFDs, the choice between W_1 and W_2 depends on whether $A_{4,0}$ or $A_{2,1}$ is less desirable, since both $A_{4,0}$ and $A_{2,1}$ pertain to either aliasing or confounding of two-factor interactions. Similarly for resolution V and VI FFDs, the choice between W_1 and W_2 depends on whether $A_{6,0}$ or $A_{3,1}$ is less desirable, since both $A_{6,0}$ and $A_{3,1}$ pertain to either aliasing or confounding of three-factor interactions.

The next example further illustrates the differences between choices of W_1 and W_2.

Example 4. Consider a DCE with eight two-level attributes and eight choice sets each with four options. We need a $2^{k-p} = 2^{8-3}$ FFD in $2^q = 2^3 = 8$ blocks each of size $2^{k-p-q} = 2^{8-3-3} = 2^2$, that is, eight choice sets each with four options. Table 5 in Xu and Lau (2006) lists two possible MA BFFDs, labeled as 8-3.1/B3(W_1) and 8-3.2/B3($W_2 W_{scf}$), which can be used to construct such a DCE. We call them D1 and D2, where D1 is optimal under the MA W_1 criterion and D2 is optimal under both the MA W_2 and W_{scf} criteria.

Design D1 has treatment defining contrast subgroup $I = ABCDF = ABEG = ACEH = CDEFG = BDEFH = BCGH = ADFGH$ and treatment wordlength pattern $W_t = (0, 3, 4, 0, 0, 0)$. This design has block defining words $b_1 = ABC, b_2 = AD, b_3 = AE$ and block wordlength pattern $W_b = (8, 16, 11, \ldots)$. For this design, all eight main effects and eight two-factor interactions ($BD, BF, CD, CF, DG, DH, FG, FH$) are clear.

Design D2 has treatment defining contrast subgroup $I = ABCDEF = ABCG = ABDH = DEFG = CEFH = CDGH = ABEFGH$ and treatment wordlength pattern $W_t = (0, 5, 0, 2, 0, 0)$. This design has block defining words $b_1 = AB, b_2 = ACD, b_3 = CE$ and block wordlength pattern $W_b = (7, 18, 10, \ldots)$. For this design, all eight main effects and four two-factor interactions (AE, AF, BE, BF) are clear.

Table 4 compares D1 and D2 and shows the main effects and two-factor interactions associated with the 31 columns in Yates order. Comparing the W_1 and W_2 combined wordlength patterns, D1 has $W_1 = (0, 3, 8, 4, \ldots)$ and $W_2 = (0, 8, 3, 4, \ldots)$, and D2 has $W_1 = (0, 5, 7, 0, \ldots)$ and $W_2 = (0, 7, 5, 0, \ldots)$. The MA W_1 criterion favors D1 because it has a smaller $A_{4,0}$ (3 vs. 5), while W_2 favors D2 because it only confounds seven two-factor interactions with blocks, $A_{2,1} = 7$ (vs. 8). Comparing the aliasing and confounding structure for each design: D1 has $(20 - 6 =)$ 14 degrees of freedom for two-factor interactions; that is, D1 has six sets of aliased two-factor interactions sacrificed for six block effects (for a total of eight two-factor interactions sacrificed for block effects) out of 20. However, D2 has only $(15 - 3 =)$ 12 degrees of freedom for two-factor interactions; that is, D2 has only three sets of aliased two-factor interactions sacrificed for three block effects (for a total of seven two-factor interactions sacrificed for block effects) out of 15. Furthermore, D1 has eight clear two-factor interactions (vs. D2 which has four clear two-factor interactions); therefore, D1 under the MA W_1 criterion may be preferred.

In Example 4, the W_1 optimal design, D1, has less aliasing and is less likely to require a follow-up experiment than the W_2 optimal design. If there is a follow-up experiment, the W_1 design will most likely be preferred because it has 14 degrees of freedom for two-factor interactions not confounded with blocks, whereas the W_2 design has 12. Xu and Mee (2010) argued that follow-up experiments are less likely for large experiments.

4.1. *Estimation capacity*

The MA criteria can be justified by the concept of estimation capacity. Cheng, Steinberg, and Sun (1999) showed that for unblocked FFDs, the MA criterion is a good surrogate for some model-robustness criteria. We now extend this justification for blocked FFDs. Assume that the main effects are of primary interest and their estimation is required. A model can be estimated by a design D if all the effects in the model are jointly estimable. For $i = 1, \ldots, \binom{k}{2}$, let $E_i(D)$ be

Table 4. Comparison of two 2^{8-3} designs in 2^3 blocks.

Column	D1	D2
1	A	A
2	B	B
3	$AB = EG$	$AB = CG = DH$ = BLOCK
4	C	C
5	$AC = EH$	$AC = BG$
6	$BC = GH$	$BC = AG$
7	DF = BLOCK	G
8	D	D
9	AD = BLOCK	$AD = BH$
10	BD	$BD = AH$
11	CF	H
12	CD	$CD = GH$
13	BF	= BLOCK
14	AF = BLOCK	= BLOCK
15	F	$EF = DG = CH$
16	E	E
17	$AE = BG = CH$ = BLOCK	AE
18	$BE = AG$	BE
19	G	
20	$CE = AH$	$CE = FH$ = BLOCK
21	H	
22	= BLOCK	
23	$CG = BH$	$DF = EG$ = BLOCK
24	DE = BLOCK	$DE = FG$
25		= BLOCK
26	FH	= BLOCK
27	DG	$CF = EH$
28	FG	
29	DH	BF
30		AF
31	EF = BLOCK	F

Table 5. A 2^{5-1} design in 2^2 blocks.

	A	B	C	D	E	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD	Block
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2	0	0	0	1	1	0	0	1	0	1	1	0	1	1	1	1	1
3	0	0	1	0	1	0	1	0	1	0	1	1	0	1	1	1	2
4	0	0	1	1	0	0	1	1	1	1	0	1	1	0	0	0	2
5	0	1	0	0	1	1	0	0	1	1	0	1	1	0	1	1	3
6	0	1	0	1	0	1	0	1	1	0	1	1	0	1	0	0	3
7	0	1	1	0	0	1	1	0	0	1	1	0	1	1	0	0	4
8	0	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	4
9	1	0	0	0	1	1	1	1	0	0	0	1	1	1	0	1	4
10	1	0	0	1	0	1	1	0	0	1	1	1	0	0	1	0	4
11	1	0	1	0	0	1	0	1	1	0	1	0	1	0	1	0	3
12	1	0	1	1	1	1	0	0	1	1	0	0	0	1	0	1	3
13	1	1	0	0	0	0	1	1	1	1	0	0	0	1	1	0	2
14	1	1	0	1	1	0	1	0	1	0	1	0	1	0	0	1	2
15	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	1	1
16	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	0	1

Note. $E = ABCD$, $b_1 = AB, b_2 = AC$, and $b_3 = b_1 b_2 = BC$.

the number of models containing all main effects and i two-factor interactions, which can be estimated by design D. It is desirable to have $E_i(D)$ as large as possible. A design D_1 is said to dominate D_2 if $E_i(D_1) \geq E_i(D_2)$ for all i, with strict inequality for at least one i. A design is said to have maximum estimation capacity (Chen and Cheng 1999; Cheng and Mukerjee 2001) if it maximizes $E_i(D)$ for all i. It is easy to see that $E_1(D) = \binom{k}{2} - 3A_{3,0}(D) - A_{2,1}(D)$ so that minimizing $3A_{3,0}(D) + A_{2,1}(D)$ would maximize $E_1(D)$. Cheng and Mukerjee (2001) argued that further minimizing $A_{4,0}(D)$ tends to make other $E_i(D)$ large. For resolution IV or higher designs, $A_{3,0}(D) = 0$; therefore, the MA criteria are good surrogates of the maximum estimation capacity criterion although they are not exactly equivalent. For a 2^{k-p} design in 2^q blocks, there are k main effects and $2^q - 1$ block effects. We can estimate at most $f = 2^{k-p} - k - 2^q$ two-factor interactions so that $E_i(D) = 0$ for $i > f$ and we only consider $(E_1, \ldots, E_f)$.

Example 5. Consider the five designs in Example 3. Their estimation capacities are:

Design	E_1	E_2	E_3	E_4	E_5	E_6	E_7
S1	7	21	35	35	21	7	1
S2	8	26	44	41	20	4	0
S3	5	10	10	5	1	0	0
S4	4	6	4	1	0	0	0
S5	4	6	4	1	0	0	0

S4 and S5 have the same estimation capacity although they are different. Both S1 and S2 dominate the other three designs in terms of estimation capacity and MA. S1 can estimate all main effects and up to seven two-factor interactions (as $E_7 = 1$), whereas S2 can estimate all main effects and at most six two-factor interactions (as $E_7 = 0$). S1 can estimate more models than S2 if more than four two-factor interactions are important. On the other hand, S2 can estimate more models containing all main effects and up to four two-factor interactions than S1.

Example 6. Consider the two designs in Example 4 and a third design, called D3, which has the same treatment defining contrast subgroup as D1 but different block defining words. The independent block defining words for D3 are $b_1 = AB, b_2 = AC, b_3 = AE$ and the block wordlength pattern is $W_b = (15, 6, 12, \ldots)$. D3 has five more (13 vs. 8) clear two-factor interactions than D1 but it has larger $A_{2,1}$ (15 vs. 8) than D1. The estimation capacities are:

Design	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}	E_{13}	E_{14}	E_{15}	E_{16}
D1	20	184	1032	3942	10848	22180	34232	40081	35436	23292	11040	3568	704	64	0	0
D2	21	200	1142	4353	11665	22526	31572	31864	22576	10656	3008	384	0	0	0	0
D3	13	78	286	715	1287	1716	1716	1287	715	286	78	13	1	0	0	0

Both D1 and D2 can estimate more models than D3 even though D3 has more clear two-factor interactions. D3 is dominated by D1 in terms of both MA and estimation capacity. D1 can estimate more models than D2 and so is preferred if seven or more two-factor interactions are important. We note that there are many other designs that are dominated by either D1 or D2 in terms of MA and estimation capacity.

As Examples 5 and 6 show, the W_1 criterion would be a better choice if the number of possible two-factor interactions is large, while the W_2 criterion would be a better choice if that number is thought to be smaller.

5. Construction methods

We present two methods to construct MA BFFDs. A regular 2^{k-p} FFD can be viewed as k columns of an $N \times (N-1)$ matrix, which consists of $k-p$ independent columns and all possible interactions among them, where $N = 2^{k-p}$. To arrange a regular 2^{k-p} FFD in 2^q blocks, one must choose q columns from the remaining $N-1-k$ columns as possible generators. The method presented by Xu and Lau (2006) uses coding theory to screen out infeasible block schemes when searching over all possible $\binom{N-1-k}{q}$ combinations of q block generators, which is fast when q is small.

The method proposed by Xu and Mee (2010) directly partitions a regular 2^{k-p} FFD into 2^q blocks of size 2^m (with $m = k-p-q$). For two-level and multilevel blocked designs, Theorem 1 from Xu and Mee (2010) presents a method to partition a regular FFD directly into blocks. For convenience, we restate their theorem.

Theorem 5.1 (Xu and Mee, 2010). *A regular s^{k-p} design D can be properly partitioned into s^q blocks of size s^m (with $m = k-p-q$) if and only if there exists an $m \times k$ submatrix V of D such that V has full row rank and every column of V is not a null vector.*

Given an unblocked $N = 2^{k-p}$ design, we choose m rows from the $N-1$ nonzero rows to form a matrix V and check whether both conditions in Theorem 5.1 are satisfied. Theorem 5.1 is most useful when m is small.

Example 7. Consider the following example to block the MA 2^{5-1} design defined by $E = ABCD$, given in Example 2 and Design S1 from the simulation study. This design is given as the first five columns in Table 5. Theorem 5.1 states that there exists a matrix V that is a subset of the design matrix given in Table 5. To partition this design into 2^2 blocks, rows 15 and 16 satisfy both conditions in Theorem 5.1. Therefore, this 2^{5-1} MA FFD can be directly partitioned in 2^2 blocks of size 2^{5-1-2}. To determine the block generators we examine the 11 remaining columns. Looking at rows 15 and 16, there are three columns (AB, AC and BC) where both elements are zero at rows 15 and 16. These columns correspond to the block columns $b_1 = AB$, $b_2 = AC$, and $b_3 = b_1b_2 = BC$. The principal block consists of rows {1, 2, 15, 16}, rows {3, 4, 13, 14} form block 2, rows {5, 6, 11, 12} form block 3, and rows {7, 8, 9, 10} form block 4.

If $m = 1$, then V is a row vector, and a regular 2^{k-p} design can be partitioned into 2^{k-p-1} blocks if and only if the design consists of a row of k ones. In this case, the unblocked FFD is a foldover design, and each row and its foldover form a block (Xu 2006). A regular foldover design is known as an *even design*, where all of the treatment words are of even length. Xu (2006) presented Corollary 3 for this special case, which states: "A regular 2^{k-p} design containing the null treatment can be partitioned into maximal 2^{k-p-1} blocks as a regular main effect (RME) design if and only if it is an even design." Here, an RME design is a design such that no main effects are aliased with other main effects, and no main effects are confounded with block effects. The treatment word ($ABCDE$) for the design in Example 7 is not of even length, and hence, there is not a row of ones in this design. Consequently, the design is not an even design and cannot be partitioned into maximal blocks. Let us consider another example.

Table 6. A 2^{5-1} design in 2^3 blocks.

	A	B	C	D	E	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD	Block
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2	0	0	0	1	0	0	0	1	0	1	1	0	1	1	1	1	2
3	0	0	1	0	1	0	1	0	1	0	1	1	0	1	1	1	3
4	0	0	1	1	1	0	1	1	1	1	0	1	1	0	0	0	4
5	0	1	0	0	1	1	0	0	1	1	0	1	1	0	1	1	5
6	0	1	0	1	1	1	0	1	1	0	1	1	0	1	0	0	6
7	0	1	1	0	0	1	1	0	0	1	1	0	1	1	0	0	7
8	0	1	1	1	0	1	1	1	0	0	0	0	0	0	1	1	8
9	1	0	0	0	1	1	1	1	0	0	0	1	1	1	0	1	8
10	1	0	0	1	1	1	1	0	0	1	1	1	0	0	1	0	7
11	1	0	1	0	0	1	0	1	1	0	1	0	1	0	1	0	6
12	1	0	1	1	0	1	0	0	1	1	0	0	0	1	0	1	5
13	1	1	0	0	0	0	1	1	1	1	0	0	0	1	1	0	4
14	1	1	0	1	0	0	1	0	1	0	1	0	1	0	0	1	3
15	1	1	1	0	1	0	0	1	0	1	1	1	0	0	0	1	2
16	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	0	1

Note. $E = ABC$, $b_1 = AB, b_2 = AC, b_3 = AD, b_4 = b_1b_2 = BC, b_5 = b_1b_3 = BD, b_6 = b_2b_3 = CD$, and $b_7 = b_1b_2b_3 = ABCD$.

Example 8. Suppose we wish to block the 2^{5-1} design defined by $E = ABC$ into maximal blocks. This design is given as the first five columns in Table 6. We note that row 16 does not contain any zeros in the first five columns. Based on Theorem 5.1, this design can be arranged into maximal blocks, that is, eight blocks (2^3) of size 2 ($2^m = 2^{5-1-3}$). The treatment word ($ABCE$) for this design is of even length, and thus this design is an even design, and each row and its foldover form a block. For example, row 16 is the foldover of row 1. Looking at row 16, there are seven columns (AB, AC, AD, BC, BD, CD, and $ABCD$) where the elements are zero. These seven columns form the block columns: $b_1 = AB, b_2 = AC, b_3 = AD, b_4 = b_1b_2 = BC, b_5 = b_1b_3 = BD, b_6 = b_2b_3 = CD$, and $b_7 = b_1b_2b_3 = ABCD$. Rows 1 and 16 form the principal block, rows 2 and 15 form block 2, rows 3 and 14 form block 3, and so on. With this design all five main effects are clear, and all two-factor interactions are confounded with block effects.

6. Extensions to mixed-level and nonregular designs

We now extend our approach to construct DCEs with mixed-level k attributes. Let $D = (T, B)$ be an $N \times (k+1)$ matrix, where T is an $(N \times k)$ matrix for the k attributes and B is an $(N \times 1)$ vector for b blocks or b choice sets. The matrix D is a mixed-level orthogonal array (OA) of strength 2 or higher, which can be used to construct a DCE with k attributes each at $s_1, \ldots, s_k$ levels, and b choice sets each with N/b options. We denote such an OA by $OA(N, s_1 \times \ldots \times s_k \times b)$. Table 5 shows the 2^{5-1} design in 4 blocks from Example 2 representable by $D = (T, B)$, where the first five columns form the matrix T and the last column forms the vector B. In this case, D is an $OA(16, 2^5 4^1)$. As another example, consider Table 6, where columns 1–5 and 13–15 define the 16×8 matrix T and the last column defines the vector B. Then $D = (T, B)$ is $OA(16, 2^8 8^1)$, which defines a DCE with eight two-level attributes in eight choice sets and each choice set consists of a foldover pair.

When we use one of the columns in an OA to define a blocking scheme, all main effects are orthogonal to block effects and the resulting block design is universally optimal for the main effects model (Dey and Mukerjee 1999, Theorem 7.4.1). The connection between the

MNL model and the linear model for blocked designs in section 3.1 implies that such a blocked design used as a DCE is a locally optimal design for the main effects MNL model assuming $\beta = \mathbf{0}$.

Theorem 6.1. *When $D = (T, B)$ is a mixed-level $OA(N, s_1 \times \ldots \times s_k \times b)$, all main effects are orthogonal to block effects. This is a locally optimal design for the main effects MNL model assuming $\beta = \mathbf{0}$.*

Although all mixed-level OAs are locally optimal designs for the main effects model, they have different properties when some two-factor interactions are significant. To further distinguish them, we extend the MA criteria to the mixed-level case. We first review the generalized minimum aberration (GMA) criterion due to Xu and Wu (2001).

Following Xu and Wu (2001), for design T with N runs and k attributes, the full analysis of variance (ANOVA) model is

$$\mathbf{y} = X_0\theta_0 + X_1\boldsymbol{\theta}_1 + X_2\boldsymbol{\theta}_2 + \ldots + X_k\boldsymbol{\theta}_k + \varepsilon \tag{10}$$

where $\mathbf{y}$ is the vector of N observations, θ_0 is the general mean, $\boldsymbol{\theta}_j$ is the vector of jth-order factorial effects, X_0 is the vector of 1's, X_j the matrix of orthonormal contrast coefficients for $\boldsymbol{\theta}_j$, and ε the vector of independent random errors. Note that jth-order factorial effects represent main effects when $j = 1$ and interactions when $j \geq 2$. Note that the contrast matrix X_j is different from $\mathbf{X}_s$ defined in section 3.1.

For $j = 1, \ldots, k$, Xu and Wu (2001) defined A_j, a function of X_j, to measure the overall aliasing between all jth-order factorial effects and the general mean. Specifically, let $X_j = [x_{il}^{(j)}]$ and define

$$A_j(T) = N^{-2}1'X_jX_j'1 = N^{-2}\sum_{l=1}^{n_j}\left(\sum_{i=1}^{N} x_{il}^{(j)}\right)^2 \tag{11}$$

where 1 is the $N \times 1$ vector of ones and n_j is the number of all jth-order factorial effects. The value of A_j is independent of the choice of the orthonormal contrasts used. The vector $(A_1, \ldots, A_k)$ is called the generalized wordlength pattern (GWLP), because for a two-level regular design, A_j is the number of words of length j. The GMA criterion (Xu and Wu 2001) is to sequentially minimize $A_1, A_2, A_3, \ldots$. A design that does this is said to have GMA.

To use a mixed-level design $D = (T, B)$ for a DCE, we define the treatment and block wordlength patterns similarly to the two-level FFDs and BFFDs presented in section 2. For a blocked design $D = (T, B)$, we define $A_i(D)$ as Eq. (11) by treating D as an unblocked (mixed-level) design, and then define two types of wordlength patterns:

$$A_{i,0}(D) = A_i(T) \tag{12}$$

and

$$A_{i-1,1}(D) = A_i(D) - A_i(T). \tag{13}$$

When $D = (T, B)$ is a mixed-level OA (of strength 2), $A_1(T) = A_2(T) = 0$ and $A_1(D) = A_2(D) = 0$ so that $A_{1,0}(D) = A_{2,0}(D) = 0$ and $A_{1,1}(D) = A_2(D) - A_2(T) = 0$. Then we can apply the GMA criterion to the sequences (6)–(9) for mixed-level designs as in the two-level designs. We use an example to show that for two-level designs the definitions of treatment and block wordlength patterns in this new formulation are consistent with the original ones.

Example 9. Consider the 2^{5-1} FFD in 2^2 blocks in Example 2 defined by $I = ABCDE$ and $b_1 = AB, b_2 = AC$. The treatment and block wordlength patterns are $W_t = (A_{3,0}, A_{4,0}, A_{5,0}) = (0, 0, 1)$ and $W_b = (A_{2,1}, A_{3,1}, A_{4,1}, A_{5,1}) = (3, 3, 0, 0)$. Table 5 displays the blocked design $D = (T, B)$, where the first five columns form the treatment matrix T and the last column is the block column B. It is obvious that $A_i(T) = A_{i,0}(D)$. To show the new formulation is consistent with the original one, we explain how to compute $A_i(D)$ as a mixed-level $OA(16, 2^5 4^1)$ according to the definition (11). The block column B has four levels, so it has three degrees of freedom, which can be represented by the contrasts $b_1 = AB, b_2 = AC, b_3 = BC$. From Eq. (10), X_1 has $5 + 3 = 8$ columns (i.e., five main effects plus three block effects), X_2 has $(5 \times 4)/2 + 5 \times 3 = 25$ columns (i.e., 10 two-factor interactions plus each block times each main effect), X_3 has $(5 \times 4 \times 3)/6 + 3 \times (5 \times 4)/2 = 40$ columns (i.e., 10 three-factor interactions plus each block times each two-factor interaction), and so on. From this, we can see that $A_i(D)$ is connected with the treatment and block wordlength patterns. For example, $A_3(D) = A_3(T) + A_{2,1}(D) = 0 + 3 = 3$ because three two-factor interactions are confounded with block effects. Similarly, we have $A_4(D) = A_4(T) + A_{3,1}(D) = 0 + 3 = 3$, $A_5(D) = A_5(T) + A_{4,1}(D) = 1 + 0 = 1$. In general we have $A_i(D) = A_i(T) + A_{i-1,1}(D)$ so Eq. (13) holds.

In general, for $D = (T, B)$, if B has b blocks, it has $b - 1$ contrasts. A generalized word of length i in the mixed-level design D falls into one of two types: (i) It involves i factors from T only, which defines a treatment relation, and (ii) it involves $i - 1$ factors from T and one contrast from B, which defines a block relation. The numbers of words of these two types are $A_i(T)$ and $A_i(D) - A_i(T)$, respectively. This justifies the definition of the treatment and block wordlength patterns in Eqs. (12) and (13).

Example 9 shows that it is cumbersome to compute the GWLP according to the definition (11). Xu and Wu (2001) developed a fast computation method based on coding theory. The GWLP function in the R package "DoE.base" (Groemping, Amarov and Xu 2015) implements this method and can compute the GWLP for mixed-level designs efficiently.

Example 10. Table 7 gives an $OA(20, 2^8 5^1)$, which has 20 runs, eight two-level factors, and one five-level factor. Suppose we want to study five two-level attributes with five choice sets and four options in each choice set. We can choose any five two-level columns as the treatment design T and the last column as the block column B, which defines five blocks. There are in total $\binom{8}{5} = 56$ choices to form an $OA(20, 2^5 5^1)$. Consider three designs. The first design uses columns: 2, 3, 5, 6, 8, 9. The block and treatment wordlength patterns for this blocked design are $W_t = (A_{3,0}, A_{4,0}, A_{5,0}) = (0.4, 0.2, 0)$ and $W_b = (A_{2,1}, A_{3,1}, A_{4,1}, A_{5,1}) = (2.4, 2.8, 1.2, 0)$, respectively. The second design uses columns: 1, 2, 3, 4, 5, 9. The two wordlength patterns for this blocked design are $W_t = (0.72, 0.2, 0)$ and $W_b = (2.40, 2.48, 1.2, 0)$. The third design uses columns: 1, 3, 6, 7, 8, 9. The two wordlength patterns are $W_t = (0.72, 0.52, 0)$ and $W_b = (3.20, 1.68, 0.88, 0)$. Among these three designs, the first design is the best and the third design is the worst with respect to all four sequences (6)–(9). Indeed, it can be verified that the first design has GMA with respect to all four sequences (6)–(9) among all possible $OA(20, 2^5 5^1)$ derived from the $OA(20, 2^8 5^1)$ given in Table 7.

We now develop some general theoretical results. A two-level design T is called an even design if all its words have even length, that is, $A_i(T) = 0$ for odd i. It is known that all even regular designs are foldover designs and can be used as a paired comparison design with each foldover pair as a choice set; see section 5. This can be generalized to nonregular designs so that the number of blocks is not necessarily a power of two. It is known that a two-level design, regular or nonregular, is an even design if and only if it is a foldover design (Cheng, Mee, and Yee 2008). Together with Theorem 6.1, we have the following result.

Theorem 6.2. *A two-level OA, regular or nonregular, can be used to define a locally optimal paired comparison design for the main effects model if and only if it is an even (or foldover) design and each foldover pair forms a choice set.*

When a two-level regular even design is used to define a paired comparison design, all two-factor interactions are confounded with block effects; see Example 8 and Table 6. We cannot estimate any effects that are confounded with block effects, but they do not bias the estimation of main effects. This is true for paired comparison designs in general. When each choice set consists of a foldover pair, the probability p_{js} in Eq. (2) does not change whether some two-factor interactions are included in the MNL model (1) or not.

Theorem 6.3. *When a two-level foldover design, regular or nonregular, is used as a paired comparison design, the estimates of all main effects are not biased even if some two-factor interactions are significant.*

Table 7. $OA(20, 2^8 5^1)$.

	A	B	C	D	E	F	G	H	Block
1	1	1	1	1	1	1	1	1	1
2	1	−1	−1	−1	1	1	−1	−1	1
3	−1	1	−1	1	−1	−1	−1	1	1
4	−1	−1	1	−1	−1	−1	1	−1	1
5	1	1	1	1	−1	1	−1	−1	2
6	1	−1	−1	−1	−1	−1	−1	1	2
7	−1	1	−1	−1	1	1	1	1	2
8	−1	−1	1	1	1	−1	1	−1	2
9	1	1	1	−1	−1	1	1	1	3
10	1	−1	−1	1	1	−1	−1	1	3
11	−1	1	−1	1	−1	−1	1	−1	3
12	−1	−1	1	−1	1	1	−1	−1	3
13	1	1	1	−1	−1	−1	−1	−1	4
14	1	−1	1	1	1	−1	1	1	4
15	−1	1	−1	1	1	1	−1	−1	4
16	−1	−1	−1	−1	−1	1	1	1	4
17	1	1	−1	−1	1	−1	1	−1	5
18	1	−1	−1	1	−1	1	1	−1	5
19	−1	1	1	−1	1	−1	−1	1	5
20	−1	−1	1	1	−1	1	−1	1	5

Theorem 6.4. *If a two-level OA, regular or nonregular, has GMA among all even (or foldover) designs, then it can be used to define a paired comparison design that has GMA with respect to all four criteria.*

Bunch et al. (1996) and others used foldover pairs to construct DCEs. Theorem 6.3 provides a good theoretical justification for the popularity of two-level paired comparison designs for estimating main effects in practice. Theorem 6.4 further shows that such designs have GMA properties over all possible designs. The next result gives a sufficient condition for a blocked design to have GMA properties with respect to all four criteria. The corresponding result for regular designs was obtained by Xu (2006).

Theorem 6.5. *If T has GMA among all designs and $D = (T, B)$ as an unblocked design has GMA among all designs, then $D = (T, B)$ as a blocked design has GMA with respect to all four criteria.*

Example 11. Consider a paired comparison design with 2^{k-2} choice sets for k two-level attributes. The MA 2^{k-1} design has resolution k and GWLP: $A_k = 1$ and other $A_i = 0$. For even k, it is a foldover design and defines a GMA paired comparison design where each foldover pair forms a choice set. For odd k, the MA design is not a foldover design as $A_k = 1$. The regular 2^{k-1} design with resolution $k-1$ is a foldover design and has GMA among all possible 2^{k-1} even designs. By Theorem 6.4, this resolution $k-1$ design can be used to define a GMA paired comparison design; see Example 8.

Example 12. Suppose we have k three-level attributes for $2 \le k \le 6$ and we wish to construct a DCE with six choice sets and three options each. We start with any $OA(18, 3^6 6^1)$, and choose the six-level column as the vector B and any other k three-level columns as the matrix T. Xu (2003) showed that T has GMA among all possible designs with 18 runs and k three-level factors. We can further show that $D = (T, B)$ has GMA among all possible $OA(18, 3^k 6^1)$. Therefore, such a DCE has GMA with respect to all four criteria.

Butler (2004) showed that some two-level foldover designs have GMA among all possible designs for $N = 24$, 32, 48, 64 runs. These GMA foldover designs can be used to define GMA paired comparison designs. There are many other results on the construction of GMA designs; see Xu, Phoa, and Wong (2009) and Xu (2015) for recent developments of nonregular designs.

Cheng, Li, and Ye (2004) studied blocked nonregular two-level designs and proposed four versions of GMA criteria. It can be shown that two of their criteria are special cases of the GMA criteria defined here with respect to W_1 and W_2. Their other two criteria can be extended by considering projections and the concept of generalized resolution proposed by Groemping and Xu (2014) for mixed-level OAs. We do not pursue this here. We note that our formulation of blocked designs for nonregular designs is more natural and more general than the approach by Cheng et al. (2004), even for two-level designs. In their approach, blocks are defined by independent generators as in regular designs so that the numbers of blocks are limited to a power of two. In our approach, blocks are defined by an individual column of a mixed-level OA so that the number of blocks are not limited to a power or a multiple of two; see Examples 10 and 12. Our approach relies on the existence of mixed-level OAs. There are various studies of the existence and construction of mixed-level OAs; see, for example, Dey and Mukerjee (1999) and Hedayat, Sloane, and Stufken (1999). Many mixed-level OAs are available in the R package "DoE.base," and Kuhfeld and Tobias (2005) provide an SAS macro to generate thousands of mixed-level OAs.

7. Summary and discussion

In this article, we have illustrated the use of MA BFFDs for constructing DCEs by building on the research performed by Jaynes et al. (2016). By considering the use of MA BFFDs we can maximize the number of models with estimable two-factor interactions by minimizing the confounding or aliasing of two-factor interactions. We presented and compared various MA criteria for selecting BFFDs. The choice of which MA criteria to use to construct a DCE depends on the goals of the study. We focused on the choice between W_1 and W_2, depending on whether aliased effects or confounded effects are viewed as less desirable. With the simulation study and various examples thereafter, we illustrated the following: (1) Effects confounded with block effects are not estimable, but do not bias the estimate of other effects; (2) aliasing causes bias, but aliased effects are estimable if all the aliases are negligible; and (3) aliasing or missing a significant two-factor interaction can bias the estimation of main effects even if all main effects and two-factor interactions are clear. The MA criteria deal with the intrinsic aliasing and confounding of a design per se and so work for linear models as well as generalized linear models. In this article we proposed the use of MA criteria for selecting BFFDs for constructing DCEs assuming the MNL model. There is potential for future work considering various models other than the MNL model and their properties.

Finally, we extended our approach to construct DCEs with mixed-level attributes through the use of mixed-level OAs, as a combination of an unblocked FFD and a column for blocks, that is, choice sets. This approach for constructing DCEs with mixed-level attributes relies on the existence of mixed-level OAs and is flexible for constructing DCEs because the blocks are defined by an individual column of a mixed-level OA and the number of blocks is not limited to a power or a multiple of two. We further extended the MA criteria to the mixed-level case and obtained some general theoretical results.

We demonstrated that MA designs tend to have large estimation capacity; that is, they tend to maximize the number of estimable models involving two-factor interactions. This is a desirable model-robustness property. To address uncertainty of potential important two-factor interactions, Li et al. (2013) proposed model-robust DCEs by considering models with all main effects and few two-factor interactions. They used a Bayesian approach to evaluate design performance in terms of an average information criterion. Their approach requires intensive computation and would not work well when the number of total runs or factors is large. On the other hand, the MA criteria are fast to compute and many MA designs have been tabulated for practical use.

Acknowledgments

The authors thank the editorial team for their helpful comments, which led to an improvement of the article.

Funding

This work was partially supported by the National Science Foundation grant DMS-1407560 for Xu and by the National Institute of General Medical Sciences of the National Institutes of Health under grant award number R01GM107639 for Wong.

References

Atkinson, A., and A. Donev. 1992. *Optimum experimental designs*. Oxford, UK: Oxford University Press.

Bliemer, M., and J. Rose 2011. Experimental design influences on stated choice output: An empirical study in air travel choice. *Transportation Research Part A: Policy and Practice* 45 (1):63–79.

Bunch, D. S., J. J. Louviere, and D. Anderson. 1996. A comparison of experimental design strategies for choice-based conjoint analysis with generic-attribute multinomial logit models. Working paper, Graduate School of Management, University of California, Davis, California.

Burgess, L., D. J. Street, and N. Wasi. 2011. Comparing designs for choice experiments: a case study. Journal of Statistical Theory and Practice 5 (1):25–46.

Burgess, L., S. A. Knox, D. J. Street, and R. Norman. 2015. Comparing designs constructed with and without priors for choice experiments: A case study. *Journal of Statistical Theory and Practice* 9:330–60.

Bush, S. 2014. Optimal designs for stated choice experiments generated from fractional factorial designs. *Journal of Statistical Theory and Practice* 8 (2):367–81.

Butler, N. A. 2004. Minimum G_2-aberration properties of two-level foldover designs. *Statistics Probability Letters* 67:121–32.

Chen, H., and C. S. Cheng. 1999. Theory of optimal blocking of 2^{n-m} designs. *Annals of Statistics* 27:1948–73.

Cheng, C. S. 2014. *Theory of factorial design: Single- and multi-stratum experiments*. Boca Raton, FL: Chapman and Hall/CRC.

Cheng, C. S., R. W. Mee, and O. Yee. 2008. Second order saturated orthogonal arrays of strength three. *Statistica Sinica* 18 (1):105–19.

Cheng, C. S., and R. Mukerjee. 2001. Blocked regular fractional factorial designs with maximum estimation capacity. *Annals of Statistics* 29:530–48.

Cheng, C. S., D. M. Steinberg, and D. X. Sun. 1999. Minimum aberration and model robustness for two-level factorial designs. *Journal of the Royal Statistical Society: Series B* 61:85–93.

Cheng, S. W., W. Li, and K. Q. Ye. 2004. Blocked nonregular two-level factorial designs. *Technometrics* 46 (3):269–79.

Cheng, S. W., and C. F. J. Wu. 2002. Choice of optimal blocking schemes in two-level and three-level designs. *Technometrics* 44 (3):269–77.

Dey, A., and R. Mukerjee. 1999. *Fractional factorial plans*. New York, NY: Wiley.

Fries, A., and W. G. Hunter. 1980. Minimum aberration 2^{k-p} designs. *Technometics* 22:601–8.

Gerard, K., M. Shanahan, and J. Louviere. 2008. Using discrete choice modeling to investigate breast screening participation. In *Using discrete choice experiments to value health and health care*, ed. M. Ryan, K. Gerard, and M. Amaya-Amaya, Dordrecht, The Netherlands: 117–137.

Groemping, U., B. Amarov, and H. Xu. 2015. DoE.base: R package version 0.27-1. http://CRAN.R-project.org/package=DoE.base. August 21.

Groemping, U., and H. Xu. 2014. Generalized resolution for orthogonal arrays. *Annals of Statistics* 42:918–39.

Grossmann, H., and R. Schwabe. 2015. Design for discrete choice experiments. In *Handbook of design and analysis of experiments*, ed. A. Dean, M. Morris, J. Stufken, and D. Bingham, 787–826 Boca Raton, FL: Chapman and Hall/CRC.

Hedayat, A. S., N. J. A. Sloane, and J. Stufken. 1999. *Orthogonal arrays: Theory and applications*. New York, NY: Springer-Verlag.

Jaynes, J., W. K. Wong, and H. Xu. 2016. Using blocked fractional factorial designs to construct discrete choice experiments for healthcare studies. *Statistics in Medicine* 35:2543–60. doi:10.1002/sim.6882.

Johnson, F. R., E. Lancsar, D. Marshall, V. Kilambi, A. Muhlbacher, D. A. Regier, B. W. Bresnahan, B. Kanninen, and J. F. P. Bridges. 2013. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value in Health* 16:3–13.

Kessels, R., B. Jones, P. Goos, and M. Vandebroek. 2011. The usefulness of Bayesian optimal designs for discrete choice experiments. *Applied Stochastic Models in Business and Industry* 27:173–88.

Kuhfeld, W., and R. Tobias. 2005. Large factorial designs for product engineering and market research applications. *Technometrics* 47:122–32.

Lancsar, E., and J. Louviere. 2008. Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics* 26 (8):661–77.

Li, W., C. J. Nachtsheim, K. Wang, R. Reul, and M. Albrecht. 2013. Conjoint analysis and discrete choice experiments for quality improvement. *Journal of Quality Technology* 45:74–99.

Mukerjee, R., and C. F. J. Wu. 2006. *A modern theory of factorial designs*. New York, NY: Springer.

Sandor, Z., and M. Wedel. 2001. Conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research* 38:430–44.

Sitter, R. R., J. Chen, and M. Feder. 1997. Fractional resolution and minimum aberration in blocked 2^{n-k} designs. *Technometrics* 39 (4):382–90.

Street, D., and L. Burgess. 2007. *The construction of optimal stated choice experiments: Theory and methods*. New York, NY: Wiley.

Sun, D. X., C. F. J. Wu, and Y. Chen, 1997. Optimal blocking schemes for 2^{n} and 2^{n-p} designs. *Technometrics* 39 (3):298–307.

Viney, R., E. Savage, and J. Louviere 2005. Empirical investigation of experimental design properties of discrete choice experiments in health care. *Health Economics* 14 (4):349–362.

Wu, C. F. J., and M. Hamada. 2009. *Experiments: Planning, analysis and optimization*, 2nd ed. New York, NY: Wiley.

Xu, H. 2003. Minimum moment aberration for nonregular designs and supersaturated designs. *Statistica Sinica* 13:691–708.

Xu, H. 2006. Blocked regular fractional factorial designs with minimum aberration. *Annals of Statistics* 34 (5):2534–53.

Xu, H. 2015. Nonregular factorial and supersaturated designs. In *Handbook of design and analysis of experiments*, ed. A. Dean, M. Morris, J. Stufken, and D. Bingham,339–70. Boca Raton, FL: Chapman and Hall/CRC.

Xu, H., and S. Lau. 2006. Minimum aberration blocking schemes for two- and three-level FFDs. *Journal of Statistical Planning and Inference* 136:4088–118.

Xu, H., and R. W. Mee. 2010. Minimum aberration blocking schemes for 128-run designs. *Journal of Statistical Planning and Inference* 140:3213–29.

Xu, H., and C. F. J. Wu. 2001. Generalized minimum aberration for asymmetrical fractional factorial designs. *The Annals of Statistics* 29 (4):1066–77.

Xu, H., F. K. H. Phoa, and W. K. Wong. 2009. Recent developments in nonregular fractional factorial designs. *Statistics Surveys* 3:18–46.

Zhang, R., and D. Park. 2000. Optimal blocking of two-level fractional factorial designs. *Journal of Statistical Planning and Inference* 91:107–21.