Assessing Some Aspects of Factor Screening with Non-normal Responses

Muhammad Azam Chaudhry and John Tyssedal
Department of Mathematical Sciences,
Norwegian University of Science and Technology (NTNU)
7491 Trondheim Norway
Email: john.tyssedal@ntnu.no
azam@math.ntnu.no

Abstract

Non-normally distributed response values, such as count data for instance, create challenges for factor screening. One problem is that variances may vary from run to run. Another is the choice of screening design for such responses. In this paper, we assess some screening performances for three popular screening designs: a definite screening design, a minimum resolution IV design and a Plackett-Burman design. Four distributions, two binomial, one gamma and one Poisson, are chosen for the response values. For each distribution we test out if it is best to use the raw data, a variance stabilizing transformation of the data or perform a generalized linear modelling assuming three factors are active. From our investigations, two-level non-regular designs gave the highest success rate in identifying the subset of active factors and a variance stabilizing transformation turned out to perform equally good or better than generalized linear modelling in most cases.

Keywords: screening designs, variance-stabilizing transformations, generalized linear models.

1. Introduction

At the early stage of an experimental work, there may be many factors that potentially can be active. It is then common to perform a screening to find out which factors that really affect the response. Often the subspace of active factors is of low dimension, maybe 2, 3 or 4 (Box and Tyssedal[1,2]). Box and Meyer[3] suggested 0.25 to be a reasonable probability that a factor is active.

Two-level designs have been the preferred choices for factor screening. Responses are often assumed to be normally distributed, models assumed to be linear and the analysis, although not always trivial, has benefited from the widely developed theory of linear models.

However, there are many practical situations where the assumptions of linear models and normally distributed responses are not even close to being valid. Typical examples are count data, often modeled as binomial or Poisson distributed. If the data are counts of rare events, the normal distribution is clearly not a good approximation and, for the aforementioned two distributions, the variances and expectations are in all cases closely related. One option is then to use a variance stabilizing transformation of the response values whose distribution is often better approximated with a normal distribution. Another possibility is to use the framework of generalized linear models (GLM) at the expense of stepping out of the world of linear models.

The simplest approach for count data is to use the method of least squares on a transformation of the response values obtained from a standard design. The standard variance stabilizing transformation for a binomial distributed random variable, X, is given by

$$h(\mathrm{X}) = \arcsin\left(\sqrt{\frac{\mathrm{X}}{n}}\right), \tag{1}$$

where n is the number of trials, and for a Poisson distributed random variable, Z, it is $g(\mathrm{Z}) = \sqrt{\mathrm{Z}}$. Various improvements exist, see Yu[4] . For binomial distributed data, Mee[5] suggested to use the Freeman and Tukey[6] transformation given by

$$h(\mathrm{X}) = \arcsin\left(\sqrt{\frac{\mathrm{X}}{n+1}}\right) + \arcsin\left(\sqrt{\frac{\mathrm{X}+1}{n+1}}\right). \tag{2}$$

This can be shown to give an approximate constant variance over a wider range of the probability, p, than (1). The Freeman and Tukey[6] transformation for the Poisson case is

$$h(Z) = \sqrt{Z} + \sqrt{Z+1}. \tag{3}$$

For the number of expected counts greater or equal 5, this transformation has better variance stabilizing properties than the single square root but performs considerably worse when the expected count is less than 2.5 (Mee[5]).

For binomial distributed data, the probability range for which the transformation (1) and (2) give an approximate constant variance increases with n. A typical such range is the

interval $\left[\frac{1}{n}, 1-\frac{1}{n}\right]$. However, in performing an experiment certain level combinations may cause p to be close to its boundary in which case the response variances will be smaller even if they are transformed. Also, for Poisson distributed response values, small counts will cause a problem. It is therefore tempting to use generalized linear models for modelling such data.

Aguirre[7] developed a Bayesian approach for analyzing data from experiments where the relation between the response and its explanatory variables was given by a generalized linear model and demonstrated its performance on data from a binomial distribution, a Poisson distribution and a gamma distribution. All the data were obtained from a definitive screening design (DSD) (Jones and Nachtsheim[8]) with 13 runs. His procedure, particularly designed for the DSD, is a sequence of several steps and is, using his own words, not to be considered as a rigid procedure.

The motivation for this paper is three-fold. First, Aguirre[7] simulated data from a DSD for his analysis. How will other types of screening designs, like two-level Plackett-Burman (Plackett and Burman[9]) and the minimum run resolution IV (Webb[10]) designs perform for data obtained from the same type of distributions? Second, for normally distributed data there are several alternative analyzing methods (Tyssedal and Hussain[11]). Would any of these also work well for non-normal responses? Third, screening is about separating out the subset of active factors from the others. With data generated from one of the three types of designs given above, is it then preferable to use a variance stabilizing transformation on the responses, GLM modelling or simply the raw data?

This paper is organized as follows. In Section 2 we give some motivational examples. Section 3 compares the performance of different screening designs on some given non-normal responses. In Section 4 a simulation study is conducted to test the overall performance of these designs. We devote Section 5 for reflections and some comparisons with earlier results obtained for normally distributed responses. Concluding remarks are given in Section 6.

2. Motivational examples

To investigate the performance of his proposed analysis procedure for non-normal responses, Aguirre[7] used the DSD to generate 13 response values for three models, one

binomial, one Poisson, and one gamma. The thirteen run DSD $\left(\mathrm{DSD}_{13}\right)$ with the response values is given below.

Table 1: The 13 run Definitive screening design for 6 factors, where the three responses are generated from the binomial, gamma, and Poisson models given in equation (4), (5) and (6).

Runs	A	B	C	D	E	F	$y_{binomial}$	y_{gamma}	$y_{Poisson}$
1	0	1	-1	-1	-1	-1	0	0.452	9
2	0	-1	1	1	1	1	10	14.368	156
3	1	0	-1	1	1	-1	5	0.043	16
4	-1	0	1	-1	-1	1	9	2.105	25
5	-1	-1	0	1	-1	-1	9	2.800	92
6	1	1	0	-1	1	1	1	2.159	39
7	-1	1	1	0	1	-1	7	3.961	26
8	1	-1	-1	0	-1	1	10	2.973	242
9	1	-1	1	-1	0	-1	10	14.455	254
10	-1	1	-1	1	0	1	0	0.138	4
11	1	1	1	1	-1	0	9	14.707	92
12	-1	-1	-1	-1	1	0	5	0.605	96
13	0	0	0	0	0	0	3	0.420	16

The DSDs are three level designs and as such allow the estimation of quadratic terms. The one in Table 1 consists of six mirror image pairs in scaled units -1 and 1 with two factor settings of zero added to each column and with the last run as a center run. With the center run added for all factors, the design given in Table 1 projects onto a full 3^2 design in every two dimensions. DSDs have several appealing properties. Main effect columns are not aliased with other main effect columns, with two-factor interaction columns or quadratic effect columns. The two-factor interaction columns, however, are aliased with each other as are the quadratic effect columns. Two-factor interaction columns and quadratic effect columns are also aliased.

The three models used by Aguirre[7] to obtain the three columns of response values were:

$$\mathrm{y_i} \sim \text{Binomial distributed } (10, p_i) \text{ with } p_i = \frac{1}{1+\mathrm{e}^{-\eta_i}} \text{ and } \eta_i = 2\,\mathrm{A} - 3\mathrm{B} + 3\mathrm{C} + 2\,\mathrm{BC}, \qquad (4)$$

$$\mathrm{y_i} \sim \text{Gamma distributed with } r = 2 \text{ and } \lambda_i = \frac{1}{2}\mathrm{e}^{\eta_i} \text{ where} \qquad (5)$$

$$\eta_i = 0.5\mathrm{A} - 0.5\mathrm{B} + \mathrm{C} + \mathrm{BB} + 0.5\mathrm{BC},$$

$$\mathrm{y_i} \sim \text{ Poisson distributed with } \mu_i = \mathrm{e}^{\eta_i} \text{ and } \eta_i = 3 + 0.5\mathrm{A} - \mathrm{B} + 0.5\mathrm{C} + \mathrm{BB} + 0.5\mathrm{BC}. \qquad (6)$$

For the gamma distribution, *r* is the shape parameter, λ the scale parameter, and its variance stabilizing transformation is the logarithm. We note that Aguirre (2016) used the canonical link for the binomial and Poisson distributed data and the variance stabilizing link for the gamma distributed data.

Screening procedures can mainly be classified as factor-based or effect-based. An explanation of the differences between these two procedures is given in Tyssedal and Niemi[12]. The goal of a factor-based procedure is to identify the subspace of active factors (Box and Tyssedal[2]). The functional relationship between the response and the factors may then be further explored afterwards. Standard factor-based procedures normally start with investigating how well all subsets of *m*-factors, $m = 1, 2, \ldots, k$, where k seldom exceeds 3 or 4, explain the variation in the data according to some measure(s) (Box and Meyer[3] , Tyssedal et al.[13]). An evaluation of the measure(s) is performed to decide how many factors and which subset of factors are likely active. The more factors that are active, the more difficult will it be to separate among subsets. Several methods of analysis are preferably used and expert knowledge may be taken into account. A list of candidate sets of active factors can be created (Tyssedal and Hussain[11]) and follow-up runs may be needed.

To investigate if such a procedure also could work for non-normally distributed data, we, for every subset of three factors, denoted $\{ x_1, x_2, x_3 \}$, defined and fitted a search model of the following type

$$Y = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i<j}^{3} \beta_{ij} x_i x_j + \sum_{i=1}^{3} \beta_{ii} x_i^2 + \varepsilon \qquad (7)$$

to the data, where β_0 is the intercept, β_i and β_{ij} represent half the main effects and half the two-factor interactions respectively, β_{ii} represent the sizes of the quadratic terms and ε is an error term with mean 0 and variance σ^2. Of course, all of the error terms are assumed to be independent. For the raw data and the transformed data the fit was evaluated using the mean squared error, MSE = $\dfrac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N-k-1}$, where *N* is the number of runs and k is the number of terms (intercept excluded), and $\hat{y}_i$ is the fitted response.

The natural extension to the GLM is then to let

$$\eta(\boldsymbol{x}) = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i<j}^{3} \beta_{ij} x_i x_j + \sum_{i=1}^{3} \beta_{ii} x_i^2 \tag{8}$$

and then evaluate the fit using the deviance. The corresponding link function, $g(.)$ is then given by

$$g(\mu(\boldsymbol{x})) = \eta(\boldsymbol{x})$$

where $\mu(.)$ is the expectation function, and the deviances are given by:

$$D_{Binomial} = 2\sum_{i=1}^{N}\left[y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) + (n - y_i)\log\left(\frac{n - y_i}{n - \hat{\mu}_i}\right)\right], \tag{9}$$

$$D_{Gamma} = D_{Poisson} = 2\sum_{i=1}^{N} y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) \tag{10}$$

for models with a constant term.

The model given in (7) is motivated by the fact that the DSD_{13} is a three level design and with a center run added for all factors allows the estimation of all 10 terms with a reasonable D-efficiency when applied to normally distributed data (Tyssedal and Chaudhry[14]). Clearly, for four active factors such a model has more terms than can be estimated.

In fitting a model with 10 terms to 13 response values, several subsets of factors may have a low MSE/deviance due to the fact that factor effects columns in the search model and those in the correct model may be "correlated", and the correct subset of active factors may not necessarily be the one with the smallest MSE/deviance. In addition, the amount of noise matters. The correct subset of active factors will, however, often be among the ones with the smallest MSE (Wolters and Bingham[15]) or smallest deviance. Tables 2, 3 and 4 show the effect of this procedure on the data in Table 1 handling them as untransformed, transformed or with the same link functions for GLM as in Aguirre[7]. Used on the binomial distributed data, GLM picked a clear winner while both for the untransformed and transformed responses there are four subsets that need further investigation. For the gamma distributed response values, there are in each case two models that separate out from the others and only the untransformed case had the correct subset on the top. All three procedures gave a clear winner for the Poisson distributed response values.

Table 2: The five subsets of factors ranked according to the smallest MSE for untransformed and transformed data and according to smallest deviance using logistic regression (GLM) for the binomial data in Table 1 .

Untransformed		Transformed		GLM	
Subset	MSE	Subset	MSE	Subset	Deviance
B C E	1.80	B C D	0.021	A B C	0.029
B C D	2.14	A B C	0.023	B C D	6.78
A B C	2.92	B C E	0.025	B C E	8.57
B C F	3.40	B C F	0.035	B C F	11.61
B D E	13.02	A B D	0.185	A B E	15.20

Table 3. The five subsets of factors ranked according to the smallest MSE for untransformed and transformed data and according to smallest deviance using GLM with variance stabilizing link for the gamma distributed data in Table 1.

Untransformed		Transformed		GLM	
Subset	MSE	Subset	MSE	Subset	Deviance
A, B, C	3.08	B, C, E	0.37	B, C, E	1.775
A, C, D	3.87	A, B, C	0.40	A, B, C	1.896
A, C, E	6.81	B, C, F	0.54	B, C, F	2.575
A, C, F	6.88	B, C, D	0.58	B, C, D	2.637
B, C, D	12.30	A, C, E	0.62	A, C, E	2.947

Table 4: The five subsets of factors ranked according to the smallest MSE for untransformed and transformed and according to the smallest deviance using GLM with canonical link for the Poisson distributed data in Table 1.

Untransformed		Transformed		GLM	
Subset	MSE	Subset	MSE	Subset	Deviance
A, B, C	497.18	A, B, C	3.580	A, B, C	3.795
A, B, E	866.52	A, B, E	16.334	A, B, D	69.287
A, B, F	1093.88	A, B, F	17.344	A, B, D	70.556
A, B, D	1109.72	A, B, D	18.013	A, B, F	74.063
B, C, E	3374.60	B, C, E	32.643	B, C, F	88.453

To investigate if we were able to identify the correct terms and maybe resolve such ambiguities as happened for the binomial and gamma distributed data in the transformed and the untransformed case, subsets with approximately the same MSE were expanded to a model of the form given in (7). Backward elimination with alpha to remove = 0.15 was then performed to remove terms. Similarly, subsets with approximately the same deviance were investigated for the GLM modelling. However, this did not change the ranking among models. As Aguirre[7], we were successful for the Poisson distributed data. For the gamma distributed data Aguirre did not find factor A, while we at least have the correct subset of factors as one out of two plausible explanations for the variation in the data. In our case, the GLM was superior for the binomial distributed data, and backward elimination resulted in the model given in (4) and one additional

AC interaction. Aguirre's procedure ended up with the correct model. The success of the GLM modelling compared to the others in this case is not surprising considering the fact that the data have five values at the extreme and several others rather close, and our procedure also benefited from using the correct link function.

3. Comparing screening designs for non-normally distributed responses

The use of generalized linear models for modelling raises several issues with respect to the experimental design used for collecting the data. One is that inference is based on asymptotic theory but screening experiments seldom have large samples. Inferential problems related to the use of GLM for analysing very small unreplicated experiments are well demonstrated in Aguirre-Torres and Vara[16] but considered less problematic for experiments with larger run sizes (Lewis et al.[17], Myers et al.[18]). Optimal design criteria are frequently used, at least to have a benchmark for the choice of design. Woods et al.[19] demonstrated the use of the Bayesian information capacity criterion for selecting designs where the response variable is approximated by a generalized linear model. Optimal design criteria are, however, more directed towards estimation and prediction than towards screening, and for nonlinear models the obtained designs will not only depend on knowing the model, but also the distribution, the size of the parameters and, for a GLM modelling, also the link function. We refer to Khuri et al.[20] and Atkinson and Woods[21] for reviews of optimal designs for generalized linear models.

It seems therefore worthwhile to investigate how effective standard screening designs, normally evaluated based on projection properties and/or desirable alias structure, are when responses are non-normal. Examples of two-level designs with good projection properties are the minimum run resolution IV (MinresIV) designs (Webb[10]) and the Plackett-Burman (PB) designs (Plackett and Burman[9]).

A MinresIV design for six factors in twelve runs (MinresIV_{12}) is given in Table 5. The MinresIV designs belong to the class of nonregular two-level designs and can accommodate k factors in $2k$ runs. Thereby they have very flexible run sizes, but at the expense of the number of experimental factors allowed. Their runs consist of k mirror image pairs, and as a result, main effects and two-factor interactions are not aliased. This is one of the reasons for their attractiveness. However, not all main effect columns are orthogonal and

main effects may thus be aliased with each other. The same applies to two-factor interaction columns.

Table 5: The MinResIV_{12} design for 6 factors in 12 runs. The design is taken from Design Expert.

Run	A	B	C	D	E	F
1	-1	1	-1	-1	-1	-1
2	1	-1	1	1	1	1
3	-1	-1	1	-1	-1	1
4	1	1	-1	1	1	-1
5	-1	-1	1	1	1	-1
6	1	1	-1	-1	-1	1
7	-1	-1	-1	-1	1	1
8	1	1	1	1	-1	-1
9	1	-1	-1	1	-1	-1
10	-1	1	1	-1	1	1
11	1	-1	1	-1	1	-1
12	-1	1	-1	1	-1	1

Another well-known class of screening designs is the orthogonal nonregular two-level designs. These apparently exist for all number of runs, $N = 4s$, $s \geq 3$, where s is an integer. Their projection properties are known to be good. Box and Tyssedal[1] defined projectivity of two-level designs as follows: *A $N \times k$ design with N runs and k factors each at two levels is said to be of projectivity P if the design contains a complete 2^p factorial in every possible subset of P out of the k factors, possibly with some points replicated.* Projection properties thus concern the properties of a design when restricted to a subset of P factors and thus fit well into the intention of screening. The alias pattern of the orthogonal nonregular two-level designs is often described as complex, but if only a few interactions are active, it is possible to take advantage of that effects are only partially aliased and hence can be estimated from the data.

An example of an orthogonal nonregular two-level design, the twelve run PB (PB_{12}) design, is given in Table 6.

Table 6: The PB_{12} design for 11 factor in 12 runs.

Run	A	B	C	D	E	F	G	H	I	J	K
1	1	1	-1	1	1	1	-1	-1	-1	1	-1
2	-1	1	1	-1	1	1	1	-1	-1	-1	1
3	1	-1	1	1	-1	1	1	1	-1	-1	-1
4	-1	1	-1	1	1	-1	1	1	1	-1	-1
5	-1	-1	1	-1	1	1	-1	1	1	1	-1
6	-1	-1	-1	1	-1	1	1	-1	1	1	1
7	1	-1	-1	-1	1	-1	1	1	-1	1	1
8	1	1	-1	-1	-1	1	-1	1	1	-1	1
9	1	1	1	-1	-1	-1	1	-1	1	1	-1
10	-1	1	1	1	-1	-1	-1	1	-1	1	1
11	1	-1	1	1	1	-1	-1	-1	1	-1	1
12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

The PB_{12} design has two non-isomorphic projections onto six factors (Lin and Draper[22]). One of the designs, design PB_{12} (6.1), has no mirror image run (columns A to F in Table 6**),** while the other design, PB_{12} (6.2), has two mirror image runs (for instance run 7 and 10 in columns A to E and G). According to Wang and Wu[23], design PB_{12} (6.1) has higher efficiency than design PB_{12} (6.2) and is therefore normally preferred when 6 columns are to be selected from a PB_{12} design.

We used the same models as Aguirre[7], given in (4), (5) and (6), and simulated new response values, 13 with the DSD_{13} and 12 for the $MinresIV_{12}$ and PB_{12} (6.1) designs for each choice of distribution. The obtained response values are given in the Tables 7, 9 and 11. For every subset of three factors, we fitted a search model of the same form as in (7) with the natural extension (8) for the GLM modelling to the response values from the DSD_{13}, and a search model of the form

$$Y = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i<j}^{3} \beta_{ij} x_i x_j + \beta_{123} x_1 x_2 x_3 + \varepsilon \qquad (11)$$

to the response values from the two-level designs with the same natural extension to the GLM. Tables 8, 10 and 12 show the five subsets with the smallest MSE (Proj1 ,…, Proj5) for the untransformed and transformed case and the five with the smallest deviance when the GLM was used.

Table 7: Binomial distributed responses obtained from the DSD_{13}, $MinResIV_{12}$ and PB_{12} (6.1) designs.

Y_{DSD}	0 10 2 8 8 4 3 10 10 0 10 8 5
$Y_{Minres\ IV}$	0 10 10 0 9 0 3 10 10 3 10 0
Y_{PB}	0 4 10 0 9 6 10 0 10 3 10 8

Table 8. The five subset of factors with the smallest MSE/deviance (M/D) for all three designs with binomial distributed response values using untransformed (U), transformed (T) and GLM modelling (G) of the data.

	Subset	Proj1	M/D	Proj2	M/D	Proj3	M/D	Proj4	M/D	Proj5	M/D
U	DSD	B C E	3.24	B C F	3.24	A B C	3.24	B C D	4.86	B E F	10.44
	MinResIV	B C F	0.08	A B C	0.08	B C D	0.08	B C E	0.11	A B E	11.33
	PB	A B C	0.41	A B F	1.50	B C D	3.33	B D F	3.75	B C F	4.50
T	DSD	A B C	0.04	B C F	0.04	B C E	0.05	B C D	0.06	A B E	0.15
	MinResIV	B C D	0.001	B C F	0.001	A B C	0.001	B C E	0.002	A B E	0.16
	PB	A B C	0.003	A B F	0.03	B C D	0.04	B C E	0.05	B C F	0.05
G	DSD	A B C	4.8e-10	B C F	11.36	B C E	17.67	B C D	19.26	A B F	23.32
	MinResIV	A B C	1.43	B C F	1.43	B C D	1.43	B C E	2.26	A B E	35.92
	PB	A B C	1.18	A B F	7.22	B C D	13.96	B C E	16.08	B C F	18.11

Table 9: Gamma distributed response values obtained from the DSD_{13}, $MinResIV_{12}$ and PB_{12} (6.1) designs.

Y_{DSD}	0.39 3.12 0.80 1.19 0.20 4.05 7.24 2.67 4.58 0.09 11.72 0.74 0.51
$Y_{Minres\ IV}$	0.23 5.15 5.94 0.44 0.34 0.90 2.66 7.28 1.68 1.82 11.72 0.10
Y_{PB}	0.64 1.89 16.16 0.16 0.34 2.45 7.24 0.36 4.58 1.82 11.72 0.74

Table 10: The five subset of factors with the smallest MSE/deviance (M/D) for all three designs with gamma distributed response values using untransformed (U), transformed (T) and GLM modelling (G) of the data.

	Subset	Proj1	M/D	Proj2	M/D	Proj3	M/D	Proj4	M/D	Proj5	M/D
U	DSD	A B C	0.58	A C D	3.04	A C F	3.45	A C E	3.53	B C D	4.21
	MinResIV	A C D	2.41	A C F	3.73	A C E	3.83	B C D	4.75	D E F	4.78
	PB	A B C	1.89	A B F	2.28	A B D	3.37	A B E	3.65	A C F	4.47
T	DSD	A C E	0.13	A C D	0.18	A B C	0.29	A C F	0.29	C D E	1.37
	MinResIV	A C E	0.38	D E F	0.49	A C D	0.76	A D F	0.77	A B E	0.82
	PB	A B C	0.15	B E F	0.48	A C F	0.53	B C D	0.66	C D E	0.84
G	DSD	A C E	0.66	A C D	0.89	A B C	1.39	A C F	1.42	C D E	6.19
	MinResIV	A C E	2.13	D E F	2.64	A C D	3.96	A B C	4.08	A D E	4.23
	PB	A BC	0.89	B D F	2.64	A C F	2.96	B C D	3.48	A B F	4.37

Table 11: Poisson distributed response values obtained from the DSD_{13}, $MinResIV_{12}$ and PB_{12} (6.1) designs.

Y_{DSD}	9 140 19 13 79 33 41 243 259 4 93 83 17
$Y_{Minres\ IV}$	5 234 89 6 79 12 103 89 259 31 237 6
Y_{PB}	13 29 257 4 88 74 260 16 87 29 234 93

Table 12: The five subset of factors with the smallest MSE/deviance obtained for all three designs with Poisson distributed response values using untransformed (U), transformed (T) and GLM modelling (G) of the data.

	Subset	Proj1	M/D	Proj2	M/D	Proj3	M/D	Proj4	M/D	Proj5	M/D
U	DSD	A B C	383.22	A B E	797.2	A B D	1013.2	A B F	1076.5	B C E	3736.9
	MinResIV	A B C	12.16	A B E	542.9	A B F	682.83	A B D	696.83	B C E	2726.8
	PB	A B C	74.91	A B F	125.5	A B D	518.3	A B E	558.5	B C E	2093.8
T	DSD	A B C	0.92	A B E	14.39	A B D	15.22	A B F	15.27	B C E	36.34
	MinResIV	A B C	0.44	A B E	12.24	B C E	19.01	A B F	19.39	A B D	19.87
	PB	A B C	0.58	A B F	4.13	A B D	13.19	A B E	13.58	B C E	17.92
G	DSD	A B C	1.88	A B D	66.40	A B F	69.51	A B E	76.33	B C E	94.18
	MinResIV	A B C	2.74	A B E	69.66	B C E	106.42	A B F	107.43	A B D	110.12
	PB	A B C	3.55	A B F	24.26	A B D	76.41	A B E	78.73	B C E	78.73

For the Poisson data the correct subset of active factors was found with no ambiguity for all three designs, independent of if raw data, transformed data or GLM modelling was used. Only the PB_{12} (6.1) succeeded in all cases for the gamma distributed responses. For the binomial distributed responses, GLM modelling was successful for all three designs but the $MinresIV_{12}$ design had three subsets with equal values for the smallest deviance.

4. A simulation study to test the overall performance of the designs

The results obtained in Section 3 did not leave any clear answer about whether to transform the data or not, or whether a GLM modelling should be preferred. In fact, the choice of design seems to matter more. To get a better impression of the overall performance for each of the three ways of handling the data, we used four distributions: binomial with n=10 and n=100, gamma and Poisson. The η-functions were allowed to be of two forms:

$$\eta(\boldsymbol{x}) = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i<j}^{3} \beta_{ij} x_i x_j + \sum_{i=1}^{3} \beta_{ii} x_i^2 \qquad (12)$$

and

$$\eta(\mathbf{x}) = \beta_0 + \sum_{i=1}^{3} \beta_i\, x_i + \sum_{i<j}^{3} \beta_{ij}\, x_i\, x_j + \beta_{123}\, x_1\, x_2\, x_3 . \qquad (13)$$

As in Tyssedal and Chaudhry[14], we for each simulation, randomly draw each model coefficient uniformly from an interval. The constant term was held fixed at 2. The interval [-1, 1] was chosen to represent small values and the interval [-3, 3] to represent large values.

For the η-function given in (12) we investigated the following four cases:

1. All coefficients are drawn uniformly from the interval [-1, 1].
2. All terms in the set $[\beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{13}, \beta_{23}]$ are drawn uniformly from [-1, 1] and all terms in $[\beta_{11}, \beta_{22}, \beta_{33}]$ from the interval [-3, 3].
3. All terms in the set $[\beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{13}, \beta_{23}]$ are drawn uniformly from [-3, 3] and all terms in $[\beta_{11}, \beta_{22}, \beta_{33}]$ from the interval [-1, 1].
4. All coefficients are drawn uniformly from the interval [-3, 3].

Since the coefficients are chosen in a low/high manner, one way to summarize the results is to define two factors F_1 and F_2 where F_1 represents the interval from which each of $[\beta_{11}, \beta_{22}, \beta_{33}]$ are randomly chosen and F_2 similarly represents $[\beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{13}, \beta_{23}]$. For each of the four distributions, our simulation study will then constitute a $2\times2\times3\times3$ design with two two-level factors F_1 and F_2, both coded to -1 and 1 for low and high levels respectively, and two three level factors: "way of handling data" and "type of design". For the last two factors we introduced two contrasts for each as follows:

Way of handling data	T_1	T_2	Type of design	D_1	D_2
Untransformed	-1	-1	DSD_{13}	-1	-1
Transformed	1	0	$MinresIV_{12}$	1	0
GLM	0	1	$PB_{12}(6.1)$	0	1

Hence, T_1 and T_2 compare the effect of transforming the data and the effect of using GLM to having them untransformed respectively, and similarly D_1 and D_2 compare the $MinResIV_{12}$ and the $PB_{12}(6.1)$ to the DSD_{13}. For each simulation and for every subset of three factors, we for the untransformed and transformed data fitted a search model of the form given in (7) to the response values obtained from the DSD_{13} and a search model of the form given in (11) to the data obtained from the $MinresIV_{12}$ and $PB_{12}(6.1)$ designs with the natural extensions to the GLM. For each coefficient scenario, way of handling data and design, 1000 simulations were carried out. The number of times the model containing the correct subset of active

factors had the smallest MSE for the untransformed and transformed data was recorded and similarly for the GLM using the deviance. These numbers, divided by 1000, will be denoted success proportions.

Our overall performance test does not follow all the steps of a factor based screening procedure. There are several reasons for our simplifications. First, the two level designs are both P=3 designs, and the DSD_{13} is also able to estimate all the terms in the search models given in (7) and (11) with natural extensions to the GLM modelling. Hence, to be able to identify three active factors, given that three really are, is something we hope and expect from these designs. If two or less factors are active, all three designs have convincing projection properties and the search models with extensions have rather few terms. We do not expect much difference between the designs in this case. For more than three active factors, our procedure runs into problems having too many terms. For some ways of handling this with the PB_{12} design, we refer to Tyssedal and Hussain[11] and Hussain[24]. However, the current state of the analysis methods for more than three active factors in 12/13 runs is such that in general one should not trust the outcome unless several methods can confirm it, see also Jones and Nachtsheim[25] when it comes to the DSD_{13}. We therefore concentrate on the case with three active factors, and think literature supports well that most often one ends up with not identifying more than the three most important factors in 12/13 run when data are real. Second, estimating success proportions only on how often the correct subset of active factors separates from the others in terms of lowest MSE/deviance is rather strict. In a single experiment, one or several of the additional steps given in Section 2 are natural to carry out, but they are not that easily automated in a simulation study. Our simplification likely makes the estimates of the screening success proportions conservative, but it is hard to imagine how this should affect the comparison between designs and between ways of handling data.

Table 13 shows the $2\times2\times3\times3$ design with four responses given as success proportions obtained from the respective distributions. With 13 two-factor interactions added (the interactions T_1T_2 and D_1D_2 were left out), the design was analysed using GLM with a logistic link. Table 14 shows the estimated coefficients in the logit for the respective terms after a backward elimination with α set to 0.05 has been performed. In order to interpret the results one should be aware that a 0.2 change in the link function can at most affect the probability of success with 0.05.

Table 13. Success proportions from the simulation study obtained with binomial distributions with n=10 and n=100, a gamma distribution and a Poisson distribution. The results are set out as a $2\times2\times3\times3$ design. Three factors are assumed active and

$$\eta(\boldsymbol{x}) = \beta_0 + \sum_{i=1}^{3}\beta_i x_i + \sum_{i<j}^{3}\beta_{ij}x_i x_j + \sum_{i=1}^{3}\beta_{ii}x_i^2 .$$

Run	F1	F2	T1	T2	D1	D2	$y \sim$ Bin n=10	$y \sim$ Bin n=100	$y \sim$ Gamma	$y \sim$ Poisson
1	-1	-1	-1	-1	-1	-1	0.255	0.486	0.262	0.356
2	1	-1	-1	-1	-1	-1	0.165	0.327	0.140	0.137
3	-1	1	-1	-1	-1	-1	0.395	0.465	0.356	0.414
4	1	1	-1	-1	-1	-1	0.346	0.423	0.312	0.291
5	-1	-1	-1	-1	1	0	0.379	0.813	0.286	0.728
6	1	-1	-1	-1	1	0	0.328	0.671	0.247	0.663
7	-1	1	-1	-1	1	0	0.690	0.922	0.437	0.884
8	1	1	-1	-1	1	0	0.645	0.884	0.435	0.837
9	-1	-1	-1	-1	0	1	0.342	0.841	0.212	0.744
10	1	-1	-1	-1	0	1	0.279	0.660	0.194	0.680
11	-1	1	-1	-1	0	1	0.700	0.955	0.259	0.895
12	1	1	-1	-1	0	1	0.606	0.897	0.245	0.798
13	-1	-1	1	0	-1	-1	0.241	0.439	0.333	0.478
14	1	-1	1	0	-1	-1	0.245	0.126	0.353	0.285
15	-1	1	1	0	-1	-1	0.463	0.746	0.695	0.440
16	1	1	1	0	-1	-1	0.354	0.408	0.714	0.297
17	-1	-1	1	0	1	0	0.371	0.821	0.475	0.832
18	1	-1	1	0	1	0	0.324	0.692	0.465	0.682
19	-1	1	1	0	1	0	0.743	0.946	0.903	0.978
20	1	1	1	0	1	0	0.684	0.878	0.895	0.883
21	-1	-1	1	0	0	1	0.331	0.871	0.474	0.783
22	1	-1	1	0	0	1	0.389	0.691	0.465	0.678
23	-1	1	1	0	0	1	0.753	0.985	0.962	0.943
24	1	1	1	0	0	1	0.677	0.914	0.955	0.891
25	-1	-1	0	1	-1	-1	0.139	0.578	0.326	0.488
26	1	-1	0	1	-1	-1	0.146	0.552	0.366	0.467
27	-1	1	0	1	-1	-1	0.144	0.574	0.707	0.430

28	1	1	0	1	-1	-1	0.142	0.580	0.715	0.390
29	-1	-1	0	1	1	0	0.364	0.822	0.464	0.787
30	1	-1	0	1	1	0	0.389	0.695	0.463	0.670
31	-1	1	0	1	1	0	0.702	0.944	0.945	0.955
32	1	1	0	1	1	0	0.662	0.882	0.718	0.893
33	-1	-1	0	1	0	1	0.340	0.847	0.475	0.820
34	1	-1	0	1	0	1	0.401	0.710	0.447	0.661
35	-1	1	0	1	0	1	0.761	0.975	0.973	0.978
36	1	1	0	1	0	1	0.680	0.903	0.728	0.913

Since the size of the coefficients are chosen at random, while type of design and way of handling data is something we can control, it is possible to think of our simulation experiment as a robust parameter design with type of design and way of handling data as control factors and F_1 and F_2 as noise factors. With that perspective, we, in this case, find it natural to divide the logit in a controlled part, $\eta_C\left(T_1,T_2,D_1,D_2\right)$, and an uncontrolled part, $\eta_U\left(F_1,F_2,T_1,T_2,D_1,D_2\right)$. For the binomial distribution with n=10 the controlled and uncontrolled parts become:

$$\hat{\eta}_C\left(T_1,T_2,D_1,D_2\right)=-0.32+0.17T_1-0.18T_2+0.42D_1+0.42D_2-0.14T_1D_1-0.10T_1D_2+0.20T_2D_1+0.28T_2D_2,$$

and

$$\hat{\eta}_U\left(F_1,F_2,T_1,T_2,D_1,D_2\right)=-0.08F_1+0.57F_2-0.06F_1F_2+0.07F_1T_2+0.06F_2T_1-0.08F_2T_2+0.12F_2D_1+0.17F_2D_2.$$

The advantage of using two-level designs is evident and also that the success proportions increase with large main effects and interactions. Large quadratic effects have a small negative effect. The effects of transforming or using a GLM analysis are not that easily observable from the equations. Considering the three ways of handling the data separately, we get:

Untransformed responses, $T_1=T_2=-1$,

$\hat{\eta}_C\left(T_1,T_2,D_1,D_2\right)=-0.31+0.36D_1+0.24D_2$, and

$\hat{\eta}_U\left(F_1,F_2,T_1,T_2,D_1,D_2\right)=-0.15F_1+0.43F_2-0.06F_1F_2+\left(0.12D_1+0.17D_2\right)F_2.$

Transformed responses, $T_1=1$, $T_2=0$,

$\hat{\eta}_C\left(T_1,T_2,D_1,D_2\right)=-0.15+0.28D_1+0.32D_2$, and

$$\hat{\eta}_U\left(F_1,F_2,T_1,T_2,D_1,D_2\right)=-0.08F_1+0.63F_2-0.06F_1F_2+\left(0.12D_1+0.17D_2\right)F_2.$$

Generalized linear modelling, $T_2=1$, $T_1=0$,

$\hat{\eta}_C\left(T_1,T_2,D_1,D_2\right)=-0.31+0.63D_1+0.70D_2$, and

$$\hat{\eta}_U\left(F_1,F_2,T_1,T_2,D_1,D_2\right)=-0.01F_1+0.49F_2-0.06F_1F_2+\left(0.12D_1+0.17D_2\right)F_2.$$

In a robust parameter design setting there will be focus on choosing designs and ways of handling data such that the effect of variation in the noise factors is as small as possible. In performing a screening we want to have high probability of identifying factors with large effects. Therefore, an increase in the logit by increasing F_1 and F_2 is desirable and leads to smaller variation in the probability of identifying active factors as long as this probability is greater than 0.5.

The $\mathrm{PB}_{12}(6.1)$ design together with a generalized linear modelling gives the highest estimated success proportion. We also notice that if a $\mathrm{DSD}_{13}\left(D_1=D_2=-1\right)$ is used, generalized linear modelling performs worse than analysing the raw data while a variance stabilizing transformation, in this case, gives a slight improvement. For the binomial with n=100, the gamma and the Poisson distributions, transforming or using GLM we get equally good results or improvements for all designs compared to using the raw data. In particular, it seems to be important to avoid using the raw data from the gamma and the Poisson distributions.

Tabel 14. Estimated coefficients for contrasts and interactions obtained by a GLM analysis with logistic link on the success proportions in Table 13.

Distr/Effects	Binomial, n=10	Binomial, n=100	Gamma	Poisson
Const	-0.32	1.25	0.13	0.92
F_1	-0.08	-0.41	-0.11	-0.30
F_2	0.57	0.56	0.76	0.43
T_1	0.17	0.10	0.67	0.10
T_2	-0.18	0.08	0.44	0.18
D_1	0.42	0.53	0.26	0.74
D_2	0.42	0.83	0.13	0.73
F_1F_2	-0.06		-0.07	
F_1T_1		-0.24	0.08	-0.08
F_1T_2	0.07	0.17	-0.10	0.07

F_1D_1			-0.05	
F_1D_2		-0.11	-0.07	
F_2T_1	0.06	0.25	0.34	
F_2T_2	-0.08	-0.14	0.12	
F_2D_1	0.12	0.07	0.05	0.21
F_2D_2	0.17	0.24	0.09	0.20
T_1D_1	-0.14	0.09		0.10
T_1D_2	-0.10	0.22	0.30	
T_2D_1	0.20	-0.14	-0.08	-0.13
T_2D_2	0.28	-0.17	0.12	

We then repeated the simulations generating the response values with the η-function given in (13). The following four scenarios for size of coefficients were investigated:

1. All coefficients are drawn uniformly from the interval [-1, 1].
2. All terms in the set $[\beta_1, \beta_2, \beta_3]$ are drawn uniformly from the interval [-1, 1] and all terms in $[\beta_{12}, \beta_{13}, \beta_{23}, \beta_{123}]$ from the interval [-3, 3].
3. All terms in the set $[\beta_1, \beta_2, \beta_3]$ are drawn uniformly from the interval [-3, 3] and all terms in $[\beta_{12}, \beta_{13}, \beta_{23}, \beta_{123}]$ from the interval [-1, 1].
4. All coefficients are drawn uniformly from the interval [-3, 3].

Table 15. Success proportions from the simulation study obtained with binomial distributions with n=10 and n=100, a gamma distribution and a Poisson distribution. The results are set out as a $2\times2\times3\times3$ design. Three factors are assumed active and $\eta(\mathbf{x}) = \beta_0 + \sum_{i=1}^{3}\beta_i x_i + \sum_{i<j}^{3}\beta_{ij} x_i x_j + \beta_{123} x_1 x_2 x_3$.

Run	F1	F2	T1	T2	D1	D2	$y \sim$ Bin n=10	$y \sim$ Bin n=100	$y \sim$ Gamma	$y \sim$ Poisson
1	-1	-1	-1	-1	-1	-1	0.241	0.424	0.266	0.395
2	1	-1	-1	-1	-1	-1	0.092	0.129	0.218	0.220
3	-1	1	-1	-1	-1	-1	0.590	0.689	0.544	0.615
4	1	1	-1	-1	-1	-1	0.333	0.375	0.346	0.357

5	-1	-1	-1	-1	1	0	0.410	0.864	0.340	0.787
6	1	-1	-1	-1	1	0	0.721	0.964	0.463	0.871
7	-1	1	-1	-1	1	0	0.586	0.930	0.541	0.944
8	1	1	-1	-1	1	0	0.780	0.956	0.561	0.933
9	-1	-1	-1	-1	0	1	0.386	0.906	0.258	0.787
10	1	-1	-1	-1	0	1	0.831	0.987	0.381	0.929
11	-1	1	-1	-1	0	1	0.578	0.937	0.238	0.892
12	1	1	-1	-1	0	1	0.793	0.973	0.329	0.887
13	-1	-1	1	0	-1	-1	0.221	0.427	0.270	0.386
14	1	-1	1	0	-1	-1	0.084	0.140	0.093	0.167
15	-1	1	1	0	-1	-1	0.611	0.736	0.702	0.690
16	1	1	1	0	-1	-1	0.373	0.411	0.460	0.408
17	-1	-1	1	0	1	0	0.412	0.892	0.507	0.864
18	1	-1	1	0	1	0	0.737	0.960	0.901	0.950
19	-1	1	1	0	1	0	0.602	0.951	0.938	0.946
20	1	1	1	0	1	0	0.773	0.956	0.941	0.964
21	-1	-1	1	0	0	1	0.367	0.933	0.504	0.911
22	1	-1	1	0	0	1	0.834	0.955	0.942	0.994
23	-1	1	1	0	0	1	0.616	0.959	0.783	0.970
24	1	1	1	0	0	1	0.770	0.984	0.939	0.988
25	-1	-1	0	1	-1	-1	0.112	0.410	0.274	0.360
26	1	-1	0	1	-1	-1	0.052	0.128	0.106	0.110
27	-1	1	0	1	-1	-1	0.321	0.739	0.780	0.630
28	1	1	0	1	-1	-1	0.195	0.408	0.444	0.282
29	-1	-1	0	1	1	0	0.409	0.887	0.463	0.801
30	1	-1	0	1	1	0	0.780	0.974	0.457	0.954
31	-1	1	0	1	1	0	0.633	0.942	0.909	0.954
32	1	1	0	1	1	0	0.781	0.960	0.910	0.960
33	-1	-1	0	1	0	1	0.407	0.911	0.470	0.870
34	1	-1	0	1	0	1	0.848	0.994	0.445	0.986
35	-1	1	0	1	0	1	0.639	0.952	0.970	0.948
36	1	1	0	1	0	1	0.819	0.983	0.980	0.990

Table 16 reports the results of a GLM analysis with logistic link after a backward elimination has been performed with α to remove set to 0.05. It is important to be aware that F_1 now represents the product terms, and F_2 represents the first order terms.

Table 16. Estimated coefficients for contrasts and interactions obtained by a GLM analysis with logistic link on the success proportions in Table 15.

Distr/Effects	Binomial, n=10	Binomial, n=100	Gamma	Poisson
Const	0.05	1.91	0.33	1,63
F_1	0.30	0.11	0.03	0.10
F_2	0.40	0.37	0.77	0.42
T_1	0.09	0.06	0.62	0.38
T_2	-0.15		0.27	
D_1	0.55	0.91	0.63	0.89
D_2	0.70	1.38	0.32	1.27
F_1F_2	-0.16		-0.10	-0.12
F_1T_1		-0.05	0.33	0.05
F_1T_2	0.03		-0.34	
F_1D_1	0.27	0.34	0.21	0.23
F_1D_2	0.45	0.50	0.34	0.45
F_2T_1		0.05		0.08
F_2T_2			0.53	
F_2D_1	-0.15	-0.15	0.06	0.09
F_2D_2	-0.21	-0.16	-0.14	-0.17
T_1D_1	-0.11		0.26	
T_1D_2	-0.14		0.32	0.26
T_2D_1	0.22		-0.08	
T_2D_2	0.26		0.31	0.19

Also in this case, GLM performed badly on data from a binomial distribution with n=10 obtained from a DSD_{13}. The main conclusions are the same. As expected there is an overall positive effect of increasing both linear terms and product terms but it is reduced somewhat when both are large and it is design dependent.

Table 17 summarizes which combination of design and way of handling data that obtained the highest success proportion.

Table 17. The combination of design and way of handling data that gave the highest success proportion for the four distributions and the two link transformed expectation functions.

	$\eta(\boldsymbol{x}) = \beta_0 + \sum_{i=1}^{3}\beta_i x_i + \sum_{i<j}^{3}\beta_{ij} x_i x_j + \sum_{i=1}^{3}\beta_{ii} x_i^2$		$\eta(\mathbf{x}) = \beta_0 + \sum_{i=1}^{3}\beta_i x_i + \sum_{i<j}^{3}\beta_{ij} x_i x_j + \beta_{123} x_1 x_2 x_3$	
Distribution	Design	Way of handling data	Design	Way of handling data
Binomial, n=10	$PB_{12}(6.1)$	GLM	$PB_{12}(6.1)$	GLM
Binomial, n=100	$PB_{12}(6.1)$	Transform	$PB_{12}(6.1)$	Transform or GLM
Gamma	$PB_{12}(6.1)$	Transform	$MinResIV_{12}$	Transform
Poisson	$PB_{12}(6.1)$ $MinResIV_{12}$	GLM Transform	$PB_{12}(6.1)$	Transform

The overall main conclusion from the study, if only screening is the purpose, seems to be that the largest increase in success proportions comes from using two-level design and the $PB_{12}(6.1)$ design performs a little better than the $MinresIV_{12}$ design. In addition, variance stabilizing transformations perform equally good or better than GLM, except for the binomial distribution with n=10.

5. Some reflections and comparison to normally distributed responses

The success of the two level designs compared to the DSD_{13} is not too surprising. Both the $PB_{12}(6.1)$ design and the $MinResIV_{12}$ design are projectivity P=3 designs, and the PB_{12} may be considered to have the best projections onto three dimensions. For instance, the PB_{12} design projects onto a complete 2^3 + a half replicate in any three dimensions. Assuming a constant variance for the response variables and m, $m \le 3$, active factors, a model independent estimate of the variance is obtained from the MSE using a search model of the form given in (11) with m factors included. When the variance is not constant, the performance of such a procedure is expected to be inferior. The deviance too has the same property as can be seen as follows. For a binomial distribution, let

$d_i = y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) + (n - y_i)\log\left(\frac{n - y_i}{n - \hat{\mu}_i}\right)$. Then, since $d_i \geq 0$, i=1,2,…,N, its minimum value is 0 for $\hat{\mu}_i = y_i$. If run number *i* and *j* are two replicated runs, than the minimum value of $d_i + d_j$ is obtained for $\hat{\mu}_i = \hat{\mu}_j = \frac{y_i + y_j}{2}$. Hence, the minimum deviance can be found without knowing the functional relationship. This can be generalised to the case when a run has more than one replicate, and it also applies to the gamma and the Poisson distribution. Therefore, a screening may be performed without knowing the true functional relationship between the response and the factors.

The DSD_{13} does not possess the possibility of providing a model independent estimate of the variance if three factors are active, and except for the logarithmic transformed gamma distributed responses obtained from the link transformed expectation function given in (12), it is in this case, as shown below, not able to take proper advantage of its ability to also estimate quadratic effects. It is, however, possible that other ways of analysing the data will improve the performance of the DSD_{13}. Jones and Nachtsheim[25] provide interesting work in that direction. They also show how it is possible to get a model independent estimate of the variance, though at the expense of increasing the number of runs.

To have a better understanding of the performance of the variance stabilizing transformations, a first order Taylor expansion around the expected value of the responses is presented below. For simplicity, we first consider the standard variance stabilizing transformations for the binomial and Poisson distributed random variables since these provide the necessary insight. In the approximate models, the distribution is indicated with an index in the transformation, *B* for the binomial, *G* for the gamma and *P* for the Poisson.

$$h_B(Y_i) \approx \mathrm{Arcsin}\left(\sqrt{\frac{1}{1+e^{-\eta_i}}}\right) + V_i \quad \text{where } V_i = \frac{1}{2n\sqrt{p_i(1-p_i)}}(Y_i - np_i) \text{ and } p_i = \frac{1}{1+e^{-\eta_i}}.$$

$$h_G(Y_i) \approx \eta_i + U_i \text{ where } U_i = \frac{1}{\mu_i}(Y_i - \mu_i) \text{ and } \mu_i = e^{\eta_i}.$$

$$h_P(Y_i) \approx e^{\frac{\eta_i}{2}} + W_i \text{ where } W_i = \frac{1}{2\sqrt{\mu_i}}(Y_i - \mu_i) \text{ and } \mu_i = e^{\eta_i}.$$

Further the transformed responses can be characterized having approximate variances and skewness of $\frac{1}{4n}$ and $\frac{1-2p_i}{\sqrt{np_i(1-p_i)}}$ for the binomial distributed data, $\frac{1}{2}$ and $\sqrt{2}$ for the gamma distributed data and $\frac{1}{4}$ and $\frac{1}{\mu_i^{1/2}}$ for the Poisson distributed data. An approximate variance for the Freeman Tukey transformation of the responses is obtained by multiplying by 4.

We notice that the transformed binomial distributed response variables have small variances that decrease with n and small skewness for most values of $p \in \left[\frac{1}{n},\ 1\text{-}\frac{1}{\mathrm{n}}\right]$, the transformed Poisson distributed response variables have approximate constant variances and skewness that decreases when μ increases, while the transformed gamma distributed response variables have approximate constant variances and skewness. Further, it illustrates that the chosen models and transformations cannot provide response variables with comparable properties with respect to both variance and skewness.

It might be reasonable to believe that if a variance stabilizing transformation is used, the screening performance is comparable to what can be obtained having normally distributed responses with the same variance. The results obtained show that a direct comparison is far from trivial. Both for the binomial distribution and for the Poisson distribution the approximate expectations for the transformed responses differ considerably from $\eta_i(\boldsymbol{x})$. This concerns the range, but also the nature of the transformations that treat large values of $\eta_i(\boldsymbol{x})$ differently. Above a certain limit, increasing $\eta_i(\boldsymbol{x})$ almost means nothing to an arcsine transformed response, while the effect may be substantial if a square root transformation is used. Direct comparison of the results obtained from the PB_{12} (6.1) with the results from the simulation study in Tyssedal and Chaudhry[14] also shows that the success proportions for the transformed binomial distributed responses with a variance of about $0.01\ (n=100)$ are comparable to those obtained from a normal distribution with variance 0.4. The success proportions for the variance stabilized Poisson distributed responses are higher than for a normal distribution with variance equal to 1.

The gamma distributed response values have approximate expectations equal to $\eta_i(\boldsymbol{x})$, and Table 18 shows a comparison between success frequencies for normally distributed responses with variances $\sigma^2 = 1/2$ taken from Tyssedal and Chaudhry[14], and logarithmic transformed gamma-distributed responses using the η-function in (12).

Table 18. Comparison of success proportions for all three designs using a normal distribution and a transformed gamma distribution with approximately the same variance $\sigma^2 = 1/2$. Case 1 and Case 2 correspond to small main effects and interactions, while the others correspond to large.

Design	Distribution	Case 1	Case 2	Case 3	Case 4
DSD_{13}	Normal, $\sigma^2 = 1/2$	0.504	0.504	0.816	0.805
	Transformed gamma	0.333	0.353	0.695	0.714
$MinResIV_{12}$	Normal, $\sigma^2 = 1/2$	0.720	0.724	0.961	0.955
	Transformed gamma	0.475	0.465	0.903	0.895
PB_{12} (6.1)	Normal, $\sigma^2 = 1/2$	0.757	0.736	0.988	0.996
	Transformed gamma	0.474	0.465	0.962	0.955

All three designs have an approximate decrease in success frequencies of 30-35 % for the transformed gamma compared to the normal distribution with the same variance, when the absolute value of the main effects and interactions are small. When the main effects and interactions are large, the decrease is considerable worse for the DSD_{13} than for the two-level designs where it only amounts to about 6 % or less. The decline in the success frequencies may be due to that the transformed responses still have some skewness.

6. Concluding remarks

In this work, we have assessed some aspects of factor screening when the responses are non-normal using designs with 12 and 13 runs. We started out reanalysing some examples from Aguirre[7] where we instead of using his sequential procedure, used a factor-based procedure based on factor sparsity and projection properties of the designs. In addition to the DSD_{13} used by Aguirre[7], we also tested two two-level designs, the $MinResIV_{12}$ and the PB_{12} (6.1) designs. To test out the overall performance of the designs we used four types of distributions, two binomial, one gamma and one Poisson, and two types of link transformed expectation functions. One with quadratic terms (12) and one containing main effects and interactions (13). The two level designs had a clear overall better performance, and the PB_{12} (6.1) design performed the best. Even though we knew the correct link function, transforming

the data gave, in most cases, equal or higher success proportions, although compared to a GLM modelling the differences are small. Both performed better than using the raw data except for the binomial distribution with n=10, where GLM performed badly on data obtained from a DSD_{13}.

It should be pointed out that our investigation is about how to do the best screening in the sense of finding the subset of factors that explains most of the variation in the data, and this should not be confused with how to best estimate the possible effects of the factors in this subset. Only the DSD_{13} can estimate quadratic effects if these are present in the η-function. It might well be that GLM performs better than transforming the data in estimating the effects of the active factors once they are identified. This is, of course, a natural procedure to perform in order to plan what to do next. However, these valuable properties may be of little value if one has the wrong subset of active factors.

Although we have tested out only a few cases of non-normal responses, the diversity of their distribution properties gives us reasons to have faith in what was observed. Two-level designs with good projection properties used together with a variance stabilizing transformation of the response values may be valuable for factor screening also when the distribution of the responses is non-normal.

References

[1] Box, G. and Tyssedal, J., (1996). Projective properties of certain orthogonal arrays. *Biometrika*, 83(4), pp. 950-955.

[2] Box, G. and Tyssedal, J., (2001). Sixteen run designs of high projectivity for screening. *Communications in Statistics-Simulations and Computation*, 30(2), pp. 217-228.

[3] Box, G. E. P. and Meyer, R. D. (1993). Finding the Active factors in Fractionated Screening Experiments. *Journal of Quaølity Technology*, 25(2), pp 94-105.

[4] Yu, G., 2009. Variance stabilizing transformations of poisson, binomial and negative binomial distributions. *Statistics and Probability Letters*, 79(14), pp. 1621-1629.

[5] Mee, R. W. (2009). A Comprehensive Guide to Factorial two-Level Experimentation. Springer, New,York.

[6] Freeman, M. and Tukey, J., (1950). Transformation related to the angular and the square root. *Annals of Mathematical Statistics*, 21(4), pp. 607-611.

[7] Aguirre, V. M., (2016). Bayesian analysis of definitive screening designs when the response is nonnormal. *Applied Staochastic Models in Bussiness and Industry*, 32(4), pp. 440-452.

[8] Jones, B. and Nachtsheim, C., (2011). A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects. *Journal of Quality Technology*, Volume 43(1), pp. 1-15.

[9] Plackett, R. and Burman, J., (1946). The design of optimum multifactorial experiments. *Biometrika*, 33(4), pp. 305-325.

[10] Webb, S., 1968. Non-orthogonal designs of even resolution. *Technometrics*, 10(2), pp. 291- 299.

[11] Tyssedal, J. and Hussain, S., (2016). Factor screening in nonregular two-level designs based on projection-based variable selection. *Journal of Applied Statistics*, 43(3), pp. 490-508.

[12] Tyssedal, J and Niemi, R. (2014) Graphical Aids for the Analysis of Two-level Non-regular Designs. *Journal of Computational and Graphical Statistics*, vol 23, 3, 678-699

[13] Tyssedal, J., Grinde, H. and Røstad, C.C. (2006). The Use of a 12-run Plackett-Burman Design in the Injection Moulding of a Technical Plastic Component. *Quality and Reliability Engineering International*, 22(6), pp 651-657.

[14] Tyssedal, J. S. and Chaudhry, A. M., 2017. The Choice of screeeing design. *Applied Stochastic Models in Business and Industry*, 33(6), pp. 662-673.

[15] Wolters, M. and Bingham, D., 2011. Simulated annealing model search for subset selection in screening experiments. *Technometrics*, 53(3), pp. 225-237.

[16] Aguirre-Torres, V. and Vars, R. D. L (2014). A Bayesian Analysis of Very Small Unreplicated Experiments. *Quality and Reliability Engineering International*, 30, pp 413-426.

[17] Lewis, S. L., Montgomery, D. C. and Myers, R. H. (2001a). Examples of designed experiments with non-normal responses. *Journal of Quality Technology*, 33, pp. 265-278.

[18] Myers, R. H., Montgomery, D. C and Vining, G. G. (2002). Generalized Linear Models With Applications in Engineering and the Sciences. Wiley, New York.

[19] Woods, D. C., Mcgree, J.M. and Lewis, S.M. (2017). Model Selection via Bayesian information capacity designs for generalized linear models. *Computational Statistics &*

Data Analysis, 113, 226-238.

[20] Khuri, A. I., Mukherjee, B., Sinha, B. K. and Ghosh, M., (2006). Design Issues for Generalized Linear Models: A Review. *Statistical Science*, 21(3), pp 376-399.

[21] Atkinson, A. C. and Woods, D. C. (2015). Designs for generalized linear models. In " *Handbook of design and Analysis of Experiments"*. Chapman & Hall/CRC Press, Boca raton.

[22] Lin, D. and Draper, N., (1992). Projection properties of Plackett and Burman designs. *Technometrics*, 34(4), pp. 423-428.

[23] Wang, J. C. and Wu, C. F., 1995. A hidden projection property of Plackett-Burman and Related Designs. 5, *Statistica Sinica*, 5(1), pp. 235-250.

[24] Hussain, S. (2017). Analysis and Exploration of Two-Level Nonregular Designs by Using Projection Properties. PhD-thesis, Department of Mathematical Sciences, NTNU.

[25] Jones, B. and Nachtsheim, C., (2017). Effective Design-Based Model Selection for Definite Screening Designs. *Technometrics*, 59(3), pp. 319-329.