

The use of nonregular fractional factorial designs in combination toxicity studies.

Frederick K.H. Phoa ^{a,*},^{1,2} Hongquan Xu ^{a,1,2}
Weng Kee Wong ^{b,2}

^a*Department of Statistics, University of California, Los Angeles, CA 90095,
United States*

^b*Department of Biostatistics, University of California, Los Angeles, CA 90095,
United States*

Abstract

When there is interest to study n chemicals using x dose levels each, factorial designs that require x^n treatment groups have been put forward as one of the valuable statistical approaches for hazard assessment of chemical mixtures. Exemplary applications and cost-efficiency comparisons of full factorial designs and regular fractional factorial designs in toxicity studies can be found in Nesnow et al. (1995), Narotsky et al. (1995), and Groten et al. (1996,1997). We introduce nonregular fractional factorial designs and show their benefits using two studies reported in Groten et al. (1996). Study 1 shows nonregular designs can provide the same amount of information using 75% of the experimental costs required in a regular design. Study 2 demonstrates nonregular designs can additionally estimate some partially aliased effects, which cannot be done using regular designs. We also provide a statistical method to evaluate the quality of an assumption made by experts in Study 2 of Groten et al. (1996).

Key words: Fractional Factorial Design; Orthogonal Array; Partially Aliasing; Plackett-Burman Design; Regression Analysis; Variable Selection.

* Corresponding author.

Email address: fredphoa@stat.ucla.edu (Frederick K.H. Phoa).

¹ Supported in part by National Science Foundation grant DMS-0806137.

² Supported in part by National Institutes of Health grant 5R01 GM072876.

1 Introduction

2 There is considerable scope for reducing resources used in research by de-
3 signing more efficient studies. Giles (2006) in a foreword in a recent issue in
4 the journal *Nature* observed that some toxicology studies seemed to lack so-
5 phisticated thinking in their designs and wondered whether that had led to
6 many inconclusive studies. The importance of a well designed study cannot be
7 over-emphasized. Experiments are increasingly complex, in addition to rising
8 experimental cost and competing resources. In the extreme case, a poorly-
9 designed study may not be able to answer the posited scientific hypotheses.
10 Careful design considerations even with only minor variation in traditional
11 designs can lead to a more efficient study in terms of more precise estimates
12 or able to estimate more effects in the study at the same cost.

13 A problem in the risk assessment of chemical mixtures is that the chemical
14 interactions hamper prediction of the toxicity of the mixture. It is impossible
15 to test each possible chemical interaction individually because of the multi-
16 tude of potential interactions. One way to overcome this problem is to treat
17 the mixture as a single compound and to test it as a whole. In this type
18 of study, the net combined effects of all components in the mixture are re-
19 flected. Factorial designs are used to detect interactions between two or more
20 chemicals in a chemical mixture. Such designs were suggested by the US En-
21 vironmental Protection Agency as one valuable statistical approach for risk
22 assessment of chemical mixtures (Svensgaard and Hertzberg 1994). A full fac-
23 torial experiment allows all factorial effects to be estimated independently and
24 is commonly used in practice (Nesnow et al. 1995, Narotsky et al. 1995). How-
25 ever, it is often too costly to perform a full factorial experiment. For example,
26 if we have 8 factors to investigate and each factor has two levels, we need to
27 have $2^8 = 256$ runs. Instead, a fractional factorial design, which is a subset or
28 fraction of a full factorial design, is often preferred because much fewer runs
29 are required. When this fraction is properly selected, the resulting design can
30 estimate the maximum number of factorial effects of interest with maximum
31 precision.

32 Fractional factorial designs are classified into two broad types: *regular* designs
33 and *nonregular* designs. Regular designs are constructed through defining re-
34 lations among factors and are described in many textbooks such as Wu and
35 Hamada (2000), Box, Hunter and Hunter (2005) and Montgomery (2009).
36 These designs are widely used in toxicity studies and other biochemical areas
37 because they are simple to construct and to analyze. The run sizes are always
38 a power of 2, 3 or the number of dose levels, and thus the “gaps” between
39 possible run sizes are getting wider as the power increases. Nonregular designs
40 such as Plackett-Burman (1946) designs and other orthogonal arrays are often
41 used in various screening experiments for their run size economy and flexibility

42 (Wu and Hamada 2000). They fill the gaps between regular designs in terms
43 of various run sizes and are flexible in accommodating various combinations of
44 factors with different numbers of levels. Compared to regular designs, nonreg-
45 ular designs have a more complex aliasing structure and thus is more difficult
46 to analyze because main effects may be partially aliased with some interac-
47 tions. Nevertheless, as we will demonstrate, the complex aliasing structure is
48 a benefit because partially aliased effects can be estimated together. A key
49 step is to disentangle the interactions from the estimates of the main effects.
50 As Hamada and Wu (1992) pointed out, ignoring non-negligible interactions
51 can lead to (i) important effects being missed, (ii) spurious effects being de-
52 tected, and (iii) estimated effects having reversed signs resulting in incorrectly
53 recommended factor levels.

54 This paper aims at demonstrating the advantages of nonregular designs over
55 regular designs in two subacute toxicity studies reported in the literature. In
56 particular, we use a 12-run Plackett-Burman design in the first study and a
57 16-run quaternary-code design in the second study. Both Plackett-Burman
58 and quaternary-code designs are special classes of nonregular designs. These
59 demonstrations show that nonregular designs are able to (i) further reduce
60 the cost of regular designs, (ii) estimate additional interactions besides those
61 that can be done with regular designs, and (iii) further reduce the biases in
62 the effect estimates.

63 2 Methods

64 We first use two studies to demonstrate the differences in analyzing data from
65 regular designs and nonregular designs. In particular, we use Groten et al.
66 (1991, 1996) to demonstrate how nonregular designs can be more cost efficient
67 than regular designs. Our second study is taken from Groten et al. (1996, 1997)
68 and we show that nonregular designs can provide additional information on
69 the estimates of some effects that regular designs are unable to do.

70 **Example 1** *Interaction of eight minerals with the oral toxicity of cadmium*
71 *in rats: application of a 12-run Plackett-Burman design.*

72 Groten et al. (1991, 1996) performed an 8-week toxicity study in Wistar rats
73 to investigate the effect of several mineral supplements, all of which had been
74 suggested to interact with the accumulation and toxicity of cadmium chloride
75 (*CC*). The 8 minerals to be tested were calcium (*Ca*), phosphorus (*P*), man-
76 ganese (*Mn*), magnesium (*Mg*), selenium (*Se*), copper (*Cu*), zinc (*Zn*) and
77 iron (*Fe*). In their study, the researchers kept the ratio between *Ca* and *P* con-
78 stant to avoid the interactive effects of each other's bioavailability. Accordingly,
79 the two minerals *Ca* and *P* were always treated as one supplement resulting in

80 a total of 7 mineral supplements under investigation. The experiment used a
81 regular fractional factorial design with 8 test groups. The chemical Cadmium
82 (*Cd*) was present in all test groups and so we may ignore its contribution to
83 all statistical analyses. The responses included clinical chemistry parameters
84 and mineral content in liver and kidneys. Groten et al. (1996) analyzed the
85 main effects first and then further tested the significant main effects and their
86 aliased two-factor interactions in a subsequent experiment. Further details of
87 the experimental setting and conditions were given in Groten et al. (1996).

88 Although combining *Ca* and *P* as a single mineral supplement enabled the
89 researchers to study eight minerals in eight test groups, their design has two
90 major drawbacks. First, *Ca* and *P* were fully aliased and their effects could
91 not be separated. Two effects are fully aliasing if the correlation between them
92 is either -1 or $+1$. When the ratio of *Ca* and *P* was kept constant, one could
93 neither distinguish the effects between them nor discover how they would
94 interact with each other. This might not be a concern for Groten et al. (1996),
95 but this is not desirable in general. Second, the design with 8 test groups for
96 testing 7 mineral supplements is saturated, so there is no degree of freedom
97 left for estimating the error variance or interactions. In their design each main
98 effect is aliased with 3 two-factor interactions. The estimate of the main effect
99 was biased and could be misleading if any of the interactions were significant.
100 As a result, the researchers had to use follow-up experiments to resolve the
101 ambiguity of the interpretation of significant effects, adding the overall cost.
102 To overcome these drawbacks, one has to use a larger design with more test
103 groups.

104 One possible design would consist of 16 test groups shown in Table 1(a). For
105 instance, the first test group involves four mineral supplements *Ca*, *P*, *Mg*, *Cu*,
106 in addition to the common mineral *Cd*. In statistical design terminology, this
107 is a regular $1/16^{th}$ fraction of a 2^8 design or a 2^{8-4} design. In this design none
108 of the main effects is aliased with two-factor interactions; therefore, all of the
109 eight main effects can be estimated even if some two-factor interactions are
110 non-negligible. Furthermore, there are 7 degrees of freedom left for estimating
111 the error variance or potential significant interactions. One disadvantage of
112 this design is that it doubles the number of test groups. However, to study 8
113 minerals together (i.e. treat *Ca* and *P* separately) , a regular design requires
114 a minimum of 16 test groups.

115 To reduce the number of test groups, we suggest to use a nonregular design
116 with 12 test groups shown in Table 1(b). This design is an example of the
117 Plackett-Burman designs available from the large collection of orthogonal ar-
118 rays given by Plackett and Burman (1946). Since there are only 8 mineral
119 supplements in the study, we choose the first 8 columns in the design, and
120 treat the remaining 3 columns as dummy variables that are negligible. Table
121 2 gives the units, levels and level assignments of each factor.

122 An obvious advantage of the new plan is the cost efficiency. The Plackett-
 123 Burman design uses only 12 test groups, a 25% saving over the regular design
 124 with 16 test groups given in Table 1(a). Like the regular design, the Plackett-
 125 Burman design allows all eight main effects to be separately estimated. It
 126 also provides 3 degrees of freedom to estimate the error variance or potential
 127 interactions.

128 **Example 2** *Interactive effects of nine chemicals in a 4-week toxicity study:*
 129 *application of a 16-run quarternary-code design.*

130 Groten et al. (1996, 1997) performed a 4-week oral/inhalatory study in which
 131 the toxicity of combinations of nine compounds was examined in male Wis-
 132 tar rats. The nine chemicals tested were dichloromethane (*MC*), formalde-
 133 hyde (*For*), aspirin (*Asp*), di-(ethylhexyl) phthalate (*DEHP*), cadmium chlo-
 134 ride (*CC*), stannous chloride (*Sn*), butylated hydroxyanisole (*BHA*), lop-
 135 eramide (*Lop*) and spermine (*Sper*) at a concentration equal to the “minimum-
 136 observed-adverse-effect level” (MOAEL). Their experiment had 16 test groups
 137 (Table 3(a)), which is $1/32^{nd}$ fraction of a 2^9 design. Besides assuming that
 138 three-factor or higher-order interactions were negligible, Groten et al. (1996,
 139 1997) further assumed that there were no interactions between formaldehyde
 140 and other compounds in the study and so they deliberately chose a design
 141 such that the main effect of formaldehyde was fully aliased with four two-
 142 factor interactions. The aliasing pattern, experimental setting and conditions
 143 were reported in Groten et al. (1997).

144 The responses in their study included body weights, organ weights, hematology,
 145 clinical chemistry and biochemistry values. They first analyzed the main
 146 effects, and then analyzed the significant main effects together with their two-
 147 factor interactions in a subsequent analysis. These analyses resulted in equa-
 148 tions that describe all hematological and clinical responses in terms of the
 149 variables tested. For example, using the aspartate aminotransferase (*ASAT*)
 150 activity (in Table 3(a)) as a response, the fitted regression equation is:

$$151 \quad ASAT(\text{units/liter}) = 75.31 + 3.44 * Asp + 5.19 * CC - 2.44 * Sn \\ + 2.56 * Lop + 2.19 * (For + CC \times Lop) - 2.56 * CC \times Sn$$

152 where $CC \times Lop$ is the interaction between *CC* and *Lop* and $CC \times Sn$ is the
 153 interaction between *CC* and *Sn*. Note that we have substituted the two-factor
 154 interaction $CC \times Lop$ in the original equation in Groten et al. (1996) by a term
 155 denoted by $(For + CC \times Lop)$ in the above equation, because the coefficient
 156 $+2.19$ is a mixed estimate from two fully aliased effects *For* and $CC \times Lop$.
 157 Because *For* and $CC \times Lop$ are fully aliased, it is impossible to distinguish
 158 between them in the analysis. Groten et al. (1996, 1997) ignored the main
 159 effect of *For* in this aliased pattern mainly because they assumed *For* was
 160 not active based on their expert opinion. However, as we will show below, by

161 using a nonregular design, we can estimate For and $CC \times Lop$ together and
162 question the validity of the expert opinion on statistical grounds.

163 For this study, we propose a nonregular design with 16 test groups displayed
164 in Table 3(b). This design is one of the quaternary-code designs constructed
165 by Xu and Wong (2007, design 9-5.ac in Table 2). The mixtures in all test
166 groups of the nonregular design are the same as those in the regular design,
167 except for test groups #2, #7, #10 and #15. In test groups #2 and #15, we
168 have added For into the original mixture, while in test groups #7 and #10,
169 we have deleted For from the original mixture. Table 4 gives the units, levels
170 and level assignments of each factor.

171 For illustrative purposes, we focus on the $ASAT$ activity as the only response
172 in this study. Data from Groten et al. (1996) for the study are shown in Table
173 3(a). To compare our proposed design with the design used in Groten et al.
174 (1996), we have to generate reasonable responses from runs in our design but
175 were not used in Groten’s design. Fortunately by construction, we can predict
176 how the set of responses will be for our design. Specifically, the only changes we
177 expect are shown in the column of $ASAT$ in Table 3(b), where there are now
178 “ $\pm a$ ” in test groups #2, #7, #10 and #15. Here the value of “ a ” represents
179 the hypothetical effect of For on the response $ASAT$ when we add For into
180 the original mixture.

181 Clearly the value of a is unknown without running a real study using our
182 design. We can however provide realistic guesses of likely values for a . In this
183 case, we consider likely values of a to be -4 , -2 , 0 , 2 and 4 . The rationale
184 for picking these values of a is consistent with the magnitude of the observed
185 effects from the real experiment. The values of a may be interpreted as follows:
186 for example, if $a = -2$, this reflects a significant negative effect, meaning
187 that when we add For into the mixture, the $ASAT$ is expected to decrease
188 significantly, other things being equal. Likewise, a value of $a = 2$ implies that
189 we can expect a significant increase in the mean $ASAT$ level when For is
190 included in the mixture. As an illustration, suppose $a = -2$. Our responses
191 in test groups #2, #7, #10 and #15 will change from 71, 96, 71, 72 to 69,
192 98, 73, 70 respectively, and other responses remain unchanged. Note that the
193 added effect “ $\pm a$ ” only changes the estimate of the main effect of For and its
194 aliased interactions including $CC \times Lop$, but it will not affect the estimates
195 of other main effects and interactions. For example, one can verify that the
196 estimate of Sn is always -2.44 for any choice of a .

197 The main reason that regular designs are incapable of estimating some in-
198 teractions is that these interactions are fully aliased with the main effects or
199 other interactions. This is a property of the regular design where fully aliasing
200 is the only possible kind of aliasing. In nonregular designs, partial aliasing is
201 possible, that is, the correlation between two effects is strictly between -1

202 and 0 or between 0 and +1. For example, the correlation between *For* and
 203 $CC \times Lop$ is 0.5 and they are partially aliased in the nonregular design. Since
 204 *For* is only partially aliased with other interactions including $CC \times Lop$, it is
 205 not necessary to assume that *For* is not active as Groten et al. (1996, 1997)
 206 did. In addition, partial aliasing reduces the bias of the estimation of main
 207 effects from non-negligible two-factor interactions.

208 3 Results

209 Groten et al. (1996) did a 4-week toxicity study with nine chemicals and
 210 showed that combined exposure to nine compounds at the “minimum-observed-
 211 adverse-effect level” (MOAEL) of the individual compounds resulted in a wide
 212 range of adverse effects. Their factorial analysis suggested that the main ef-
 213 fects of *Sn*, *CC*, *Lop*, *Asp* and the interactions between *CC* and *Lop* and be-
 214 tween *CC* and *Sn* were significant to the response aminotransferase (*ASAT*)
 215 activity. If the significant level were increased to 15%, the main effect of buty-
 216 lated hydroxyanisole (*BHA*) would also be significant to the response. They
 217 purposely designed their experiment such that formaldehyde (*For*) was fully
 218 aliased with four two-factor interactions, including the significant interaction
 219 between *CC* and *Lop*. Then they suggested choosing the interaction, rather
 220 than the main effect, as one of the significant effects based on their expert
 221 knowledge, even though the analysis failed to distinguish between them.

222 The nonregular design has a distinct advantage over the regular design be-
 223 cause it allows the estimation of all of the main effects, even when they are
 224 partially aliased with some two-factor interactions. In our case, we were able to
 225 identify the significance of *For* and its partially aliased two-factor interactions
 226 together. For example, six compounds were found to affect the *ASAT* activity
 227 when we generated the response with $a = -2$: there was a decrease in *ASAT*
 228 activity due to *Sn* or *BHA* or *For*, and an increase in *ASAT* activity caused
 229 by *CC*, *Asp* or *Lop*. Two interactions ($CC \times Lop$ and $CC \times Sn$) included in
 230 the original analysis of the regular design were also found to be significant in
 231 our analysis.

232 Following Groten et al. (1996), we have a final equation to describe the value
 233 of the response in any particular mixture in terms of the compounds tested.
 234 The final equation for the *ASAT* activity with $a = -2$ is:

$$\begin{aligned}
 ASAT(\text{units/liter}) = & 75.31 + 3.44 * Asp + 5.19 * CC - 2.44 * Sn \\
 & + 2.56 * Lop - 1.94 * BHA - 3.54 * For \\
 & + 4.46 * CC \times Lop - 2.56 * CC \times Sn
 \end{aligned}$$

236 where *Asp*, *CC*, *Sn*, *Lop*, *BHA* and *For* are the level assignments of the

237 corresponding compounds in the mixture, having a value of either +1 (pres-
238 ence) or -1 (absence). For every random selection of mixtures from the nine
239 compounds tested, it is possible to predict the overall effect for the *ASAT*
240 activity with the final equation.

241 This equation can be interpreted as follows. When 5000mg acetyl salicylic acid
242 per 1kg diet is added and the exposure levels of other chemicals are fixed, the
243 *ASAT* activity increases by 6.88(= 3.44×2) units/liter. The interpretations for
244 *For* and *BHA* are similar. However, the interpretations for *Sn*, *Lop* and *CC*
245 are more complicated because of the existence of two-factor interactions. When
246 3000mg stannous chloride per 1kg diet without cadmium chloride is added and
247 the exposure levels of other chemicals are fixed, the *ASAT* activity increases by
248 0.24(= (-2.44+(-2.56)(-1))×2) units/liter. If cadmium chloride exists in the
249 diet, then the addition of stannous chloride leads to a decrease in the *ASAT*
250 activity by 10.00 units/liter because (-2.44 + (-2.56)(+1)) × 2 = -10.00.
251 Similarly, the interpretation for *Lop* depends the presence of *CC* while the
252 interpretation for *CC* depends the presence or absence of both *Sn* and *Lop*.

253 4 Discussion

254 Our first study illustrates the run size economy of nonregular designs without
255 sacrificing the estimation abilities of the designs. The number of test groups or
256 trials in an experiment using regular designs is always a power of the number
257 of dose levels. To study 8 mineral supplements, each with two dose levels, a
258 regular design requires at least 16 test groups while a nonregular design uses
259 only 12 test groups. Nonregular designs are also flexible in accommodating
260 various combinations of factors with different numbers of dose levels.

261 Our second study illustrates how a nonregular design provides additional in-
262 formation of the interactions through their partially aliasing with the main ef-
263 fects. Groten et al. (1996) noticed that the combined effects of two compounds
264 were not a simple summation of responses of the individual compounds. In
265 a regular design, independent estimates of a fully aliased pair of factorial ef-
266 fects are impossible without additional assumptions on the significance of the
267 aliased factorial effects. However, by proper choice of a nonregular design,
268 we were able to decouple the partial aliasing between main effects and two-
269 factor interactions and so able to estimate both effects simultaneously. This is
270 possible as long as there are enough degrees of freedom left in the model.

271 We demonstrate this advantage via Study 2. The analysis of Groten et al.
272 (1996) showed the significance of the *CC* × *Lop* interaction under the as-
273 sumption that *For* were negligible due to their expert knowledge. Figure 1
274 provides a test on the significance of the estimates of the main effect of *For*

275 and the $CC \times Lop$ interaction. We use the original equation from Groten et
 276 al. (1996) and vary different values of a . In Figure 1, For and $CC \times Lop$
 277 represent the estimates of the individual effects using the nonregular design,
 278 while $(For + CC \times Lop)$ represents the estimates of the fully aliased effects of
 279 For and $CC \times Lop$ using the regular design.

280 One of the most surprising results is that when $a = 0$, For has a negative effect,
 281 $CC \times Lop$ has a positive effect, and both For and $CC \times Lop$ are significant
 282 at 5% level while $(For + CC \times Lop)$ is not. Recall that the value of “ a ” is the
 283 *additional* hypothetical effect of For on the response $ASAT$ when we add For
 284 into the original mixture. Groten et al. (1996) assumed that the main effect of
 285 For was negligible in their analysis. If their assumption was correct, we would
 286 expect that For is not significant when $a = 0$. The contradiction provides
 287 statistical evidence to question their expert opinion on the insignificance of
 288 For . Our finding further suggests that the interaction $CC \times Lop$ could be
 289 underestimated by Groten et al. (1996) because For had a negative effect.

290 When we deliberately add a negative effect (like $a = -2$ or $a = -4$) to
 291 For , both For and $CC \times Lop$ are significant at 1% significance level while
 292 $(For + CC \times Lop)$ is not significant at 10% significance level. This shows
 293 how the nonregular design correctly identifies the significance of both For
 294 and $CC \times Lop$ individually but the regular design fails to do so. On the other
 295 hand, when we add a positive effect $a = 2$ to For , $CC \times Lop$ is significant at
 296 5% level but For and $For + CC \times Lop$ are not. This is not surprising because
 297 the additional positive effect cancels the original negative effect of For .

298 Furthermore, the nonregular design can reduce the bias in the estimates of
 299 the main effects when not all two-factor interactions are negligible. If it is
 300 not known in advance which interactions can be considered as negligible, a
 301 conservative approach is to minimize the maximum possible bias arising from
 302 the existence of two-factor interactions in the true model. Because main effects
 303 are partially aliased with two-factor interactions in nonregular designs but not
 304 in regular designs, it follows that the maximum value of the bias could be
 305 relatively small in nonregular designs. This implies that the estimates of the
 306 main effects suffer a smaller bias in nonregular designs than in regular designs.

307 To fix ideas, consider the bias of the estimate of a main effect from both the
 308 regular design and the nonregular design. In the regular design, the expected
 309 value of the estimate of the main effect of For is

$$310 \quad E(\hat{\beta}_{For}) = \beta_{For} + \beta_{MC \times DEHP} + \beta_{Asp \times BHA} + \beta_{CC \times Lop} + \beta_{Sn \times Sper}$$

311 This expression includes the main effect of For and four two-factor interac-
 312 tions with coefficients all equal to 1. The aliasing structure of the nonregular
 313 design is more complicated than that of the regular design. Table 5 gives the
 314 expected value of the estimate of each main effect when two-factor interac-

315 tions are present. All the expressions include some two-factor interactions with
316 coefficients all equal to $\pm 1/2$. Therefore, if there is no prior information on
317 which interactions can be considered as negligible, a conservative approach in
318 minimizing the coefficients is to minimize their maximum value, which is 1 in
319 the case of the regular design and $1/2$ in the case of the nonregular design.
320 This shows that there is a larger bias in the regular design than in the non-
321 regular design. Further details on bias reduction are given in Wu and Hamada
322 (2000) and Deng and Tang (2002).

323 The second study shows a potential drawback of a nonregular design is that
324 its aliasing pattern can be more complicated than that from a regular design.
325 However, we feel that the advantages of nonregular designs outweigh their
326 disadvantages.

327 As a final note, all the designs discussed here are two-level designs. While
328 two-level designs are cost-effective in screening variables, they cannot identify
329 nonlinear relationship between the response and factors. A linear relationship
330 is good approximation when the high and low dose levels are close enough. The
331 approximation becomes worse when the distance between two levels increases.
332 One way to cope with this concern is to add a few (3–5) runs at the center.
333 Adding center points to a two-level design can not only provide a check on a
334 curvature effect but also provide an unbiased estimate of the error variance. If a
335 curvature effect is present, the researchers should conduct further experiments
336 to investigate the nonlinear relationship.

337 **Acknowledgments**

338 The authors thank the Editor and two referees for their constructive com-
339 ments, which led to improvements in the article.

340 **Appendix: Statistical Analysis Strategy**

341 We provide more details on how we perform analysis in study 2. We adopt
342 one of the analysis strategies suggested by Hamada and Wu (2000, p. 356).
343 The procedure is as follows.

344 **Step 1** *For each factor X , consider X and all its two-factor interactions XY*
345 *with other factors. Use a stepwise regression procedure to identify significant*
346 *effects from the candidate variables and denote the selected model by M_X .*
347 *Repeat this for each of the factors and then choose the best model.*

348 **Step 2** Use a stepwise regression procedure to identify significant effects among
349 the effects identified in the previous step as well as all the main effects.

350 **Step 3** Consider (i) the effects identified in step 2 and (ii) the two-factor
351 interactions that have at least one component factor appearing among the main
352 effects in (i). Use a stepwise regression procedure to identify significant effects
353 among effects in (i) and (ii).

354 We iterate between steps 2 and 3 until the selected model does not change. We
355 may have an over-parameterized model, i.e., more variables than the number
356 of runs, in steps 2 and 3. In such a case we replace stepwise regression with
357 forward selection.

358 In step 1 we compare nine different models, each consisting of a main effect and
359 some two-factor interactions selected via stepwise regression. Guided by the
360 prior information that *For* does not interact with other compounds, we choose
361 a model consisting of the main effect of *CC* and three two-factor interactions
362 $CC \times Lop$, $CC \times Sn$ and $CC \times Asp$. In step 2 we consider all main effects and
363 the three interactions suggested in step 1. When stepwise regression is applied,
364 there are eight significant effects at the 5% significance level. They are *Asp*,
365 *CC*, *Sn*, *Lop*, *BHA*, *For*, $CC \times Lop$ and $CC \times Sn$. Note that $CC \times Asp$ is no
366 longer significant. In step 3 we consider the eight significant effects identified in
367 step 2 together with two-factor interactions that have at least one component
368 factor appearing among the six main effects in step 2. Forward selection does
369 not find any additional significant effects and thus there is no need to iterate
370 between steps 2 and 3. The final model consisting of the eight effects has a
371 multiple R-squared of 0.97, indicating a good fit.

372 The analysis strategy works well under the following two conditions: (1) only
373 a few effects are statistically significant and (2) when a two-factor interaction
374 is significant, at least one of the corresponding factor main effects is also
375 significant. In practice it is possible to obtain uninterpretable models that
376 consist of an interaction term without any of its parent main effects. It is also
377 possible that the analysis procedure finds several incompatible models that
378 are equally plausible. When these happen, it is a strong indication that the
379 information provided in the data and design is limited and no analysis method
380 can rescue. One solution is to conduct follow-up experiments using additional
381 runs. See Wu and Hadamard (2000, Section 4.4) and Box, Hunter and Hunter
382 (2005, Section 7.2) for choosing follow-up runs.

383 References

- 384 [1] Box G.E.P., Hunter W.G., Hunter J.S., 2005. *Statistics for Experimenters:*
385 *Design, Innovation, and Discovery*, 2nd ed. New York: Wiley.

- 386 [2] Deng L.Y., Tang B., 2002. Design selection and classification for
387 Hadamard matrices using generalized minimum aberration criteria. *Tech-*
388 *nometrics* 44, 173-184.
- 389 [3] Giles, J., 2006. Animal experiments under fire for poor design. *Nature*
390 444, 981.
- 391 [4] Gorten J.P., Schoen E.D., Feron V.J., 1996. Use of factorial designs in
392 combination toxicity studies. *Food and Chemical Toxicology* 34, 1083-
393 1089.
- 394 [5] Groten J.P., Schoen E.D., Kuper C.F., van Bladeren P.J., Van Zorge J.A.,
395 Feron V.J., 1997. Subacute toxicity of a mixture of nine chemicals in rats:
396 detecting interactive effects with a fractionated two-level factorial design.
397 *Fundamental and Applied Toxicology* 36, 13-29.
- 398 [6] Groten J.P., Sinkeldam E.J., Muys T., Luten J.B., van Bladeren P.J.,
399 1991. Interaction of dietary Ca, P, Mg, Mn, Cu, Fe, Zn and Se with the
400 accumulation and oral toxicity of cadmium in rats. *Food and Chemical*
401 *Toxicology* 29, 249-258.
- 402 [7] Hamada M., Wu C.F.J., 1992. Analysis of designed experiments with
403 complex aliasing. *Journal of Quality Technology* 24, 130-137.
- 404 [8] Montgomery, D.C., 2009. *Design and analysis of experiments*. 7th ed.
405 New York: Wiley.
- 406 [9] Narotsky M.G., Weller E.A., Chinchilli V.M., Kevlock R.J., 1995. Non-
407 additive developmental toxicity in mixtures of trichloroethylene, di(2-
408 ethylhexyl)phthalate and heptachlor in a 5x5x5 design. *Fundamental and*
409 *Applied Toxicology* 27, 203-216.
- 410 [10] Nesnow S., Ross J.A., Stoner G.D., Mass M.J., 1995. Mechanistic linkage
411 between DNA adducts, mutations in oncogenes and tumorigenesis of car-
412 cinogenic environmental polycyclic aromatic hydrocarbons in strain A/J
413 mice. *Toxicology* 105, 403-413.
- 414 [11] Plackett R.L., Burman J.P., 1946. The design of optimum multifactorial
415 experiments. *Biometrika* 33, 305-325.
- 416 [12] Svendsgaard D.J., Hertzberg R.C., 1994. Statistical methods for the tox-
417 icological evaluation of the additivity assumption as used in the environ-
418 mental protection agency chemical mixture risk assessment guideline. In
419 *Toxicology of Chemical Mixtures*. Edited by R. S. H. Yang. pp.599-640.
420 Academic Press, San Diego, CA.
- 421 [13] Wu C.F.J., Hamada M., 2000. *Experiments: Planning, Analysis, and Pa-*
422 *rameter Design Optimization*. New York: Wiley.
- 423 [14] Xu H., Wong A., 2007. Two-level nonregular designs from quaternary
424 codes. *Statistica Sinica* 17, 1191-1213.

425 Table 1: Test groups in Study 1: Interaction of mineral supplements with the
 426 toxicity of *CC*: (a) Test groups of a regular design and (b) Test groups of a
 427 nonregular design.

Table (a): $1/16^{th}$ fraction of a 2^8 design (Regular Design)

- | | |
|---------------------------|--|
| 1. + Cd + Ca, P, Mg, Cu | 2. + Cd + Ca, P, Fe, Zn |
| 3. + Cd + Ca, P, Se, Mn | 4. + Cd + Ca, Mg, Fe, Se |
| 5. + Cd + Ca, Mg, Zn, Mn | 6. + Cd + Ca, Fe, Cu, Mn |
| 7. + Cd + Ca, Cu, Zn, Se | 8. + Cd + all minerals at a high level |
| 9. + Cd + Mn, Se, Zn, Fe | 10. + Cd + Mn, Mg, Se, Cu |
| 11. + Cd + Mg, Cu, Zn, Fe | 12. + Cd + P, Mn, Cu, Zn |
| 13. + Cd + P, Se, Cu, Fe | 14. + Cd + P, Mg, Se, Zn |
| 15. + Cd + P, Mn, Mg, Cu | 16. + Cd + all minerals at a low level |

Table (b): 12-run Plackett-Burman design (Nonregular Design)

- | | |
|-----------------------------|---|
| 1. + Cd + Mn, Zn, Fe | 2. + Cd + P, Cu, Zn, Fe |
| 3. + Cd + Ca, Se, Cu, Zn | 4. + Cd + Mg, Se, Cu, Fe |
| 5. + Cd + Mn, Mg, Se, Zn | 6. + Cd + P, Mn, Mg, Cu |
| 7. + Cd + Ca, P, Mn, Se, Fe | 8. + Cd + Ca, P, Mg, Zn |
| 9. + Cd + Ca, Mn, Cu | 10. + Cd + P, Se |
| 11. + Cd + Ca, Mg, Fe | 12. + Cd + all minerals at a high level |
-

429 Table 2: Study 1: (a) Factors and levels; (b) Test groups and exposure levels (D1,
 430 D2, D3 are dummy).

431

(a)

Factor	Unit	Low (-)	High (+)
Ca	%	0.64 – 0.66	1.28 – 1.30
P	%	0.59 – 0.60	1.30 – 1.35
Zn	mg/kg	28 – 29	125 – 140
Cu	mg/kg	7 – 11	46 – 70
Fe	mg/kg	35 – 46	185 – 245
Mg	%	0.046 – 0.047	0.24 – 0.26
Mn	mg/kg	45 – 60	235 – 270
Se	mg/kg	0.09 – 0.11	0.62 – 0.88

(b)

432

Compounds	Ca	P	Mn	Mg	Se	Cu	Zn	Fe	D1	D2	D3
+Cd+Mn, Zn, Fe	-	-	+	-	-	-	+	+	+	-	+
+Cd+P, Cu, Zn, Fe	-	+	-	-	-	+	+	+	-	+	-
+Cd+Ca, Se, Cu, Zn	+	-	-	-	+	+	+	-	+	-	-
+Cd+Mg, Se, Cu, Fe	-	-	-	+	+	+	-	+	-	-	+
+Cd+Mn, Mg, Se, Zn	-	-	+	+	+	-	+	-	-	+	-
+Cd+P, Mn, Mg, Cu	-	+	+	+	-	+	-	-	+	-	-
+Cd+Ca, P, Mn, Se, Fe	+	+	+	-	+	-	-	+	-	-	-
+Cd+Ca, P, Mg, Zn	+	+	-	+	-	-	+	-	-	-	+
+Cd+Ca, Mn, Cu	+	-	+	-	-	+	-	-	-	+	+
+Cd+P, Se	-	+	-	-	+	-	-	-	+	+	+
+Cd+Ca, Mg, Fe	+	-	-	+	-	-	-	+	+	+	-
+all minerals at a high level	+	+	+	+	+	+	+	+	+	+	+

433 Table 3: Test groups in Study 2: Interactive effects between nine chemicals in 4-week
 434 toxicity study with the response *ASAT*: (a) Test groups of a regular design and (b)
 435 Test groups of a nonregular design.

436

Table (a): $1/32^{nd}$ fraction of a 2^9 design (Regular Design)

Mixture Components	ASAT	Mixture Components	ASAT
1. +For	70	2. +Sn, MC, Lop, Asp	71
3. +CC, MC, Sper, Asp	86	4. +Sn, CC, Sper, Lop, For	75
5. +BHA, MC, Sper, Lop	65	6. +Sn, BHA, Sper, Asp, For	70
7. +CC, BHA, Lop, Asp, For	96	8. +Sn, CC, BHA, MC	65
9. +DEHP, Sper, Lop, Asp	77	10. +Sn, DEHP, MC, Sper, For	71
11. +CC, DEHP, MC, Lop, For	88	12. +Sn, CC, DEHP, Asp	80
13. +BHA, DEHP, MC, Asp, For	68	14. +Sn, BHA, DEHP, Lop	69
15. +CC, BHA, DEHP, Sper	72	16. +All nine compounds at MOAEL	82

437

Table (b): $1/32^{nd}$ fraction of a 2^9 design (Nonregular Design)

Mixture Components	ASAT	Mixture Components	ASAT
1. +For	70	2. +Sn, MC, Lop, Asp, For	$71 + a$
3. +CC, MC, Sper, Asp	86	4. +Sn, CC, Sper, Lop, For	75
5. +BHA, MC, Sper, Lop	65	6. +Sn, BHA, Sper, Asp, For	70
7. +CC, BHA, Lop, Asp	$96 - a$	8. +Sn, CC, BHA, MC	65
9. +DEHP, Sper, Lop, Asp	77	10. +Sn, DEHP, MC, Sper	$71 - a$
11. +CC, DEHP, MC, Lop, For	88	12. +Sn, CC, DEHP, Asp	80
13. +BHA, DEHP, MC, Asp, For	68	14. +Sn, BHA, DEHP, Lop	69
15. +CC, BHA, DEHP, Sper, For	$72 + a$	16. +All nine compounds at MOAEL	82

438 Table 4: Study 2: (a) Factors and levels; (b) Test groups and exposure levels.

439

(a)

Factor (Symbol)	Unit	Low (-)	High (+)
Aspirin (Asp)	mg/kg	0	5000
Cadmium Chloride (CC)	mg/kg	0	50
Stannous Chloride (Sn)	mg/kg	0	3000
Loperamine (Lop)	mg/kg	0	30
Spermine (Sper)	mg/kg	0	2000
Butyl hydroxyanisol (BHA)	mg/kg	0	3000
di(2-ethylhexyl)phthalate (DEHP)	mg/kg	0	1000
Dichloromethane (MC)	ppm	0	500
Formaldehyde (For)	ppm	0	3

(b)

Compounds	For	MC	Asp	CC	Sn	Lop	Sper	BHA	DEHP	ASAT
+For	+	-	-	-	-	-	-	-	-	70
+Sn,MC,Lop,Asp,For	+	+	+	-	+	+	-	-	-	71 + a
+CC,MC,Sper,Asp	-	+	+	+	-	-	+	-	-	86
+Sn,CC,Sper,Lop,For	+	-	-	+	+	+	+	-	-	75
+BHA,MC,Sper,Lop	-	+	-	-	-	+	+	+	-	65
+Sn,BHA,Sper,Asp,For	+	-	+	-	+	-	+	+	-	70
+CC,BHA,Lop,Asp	-	-	+	+	-	+	-	+	-	96 - a
+Sn,CC,BHA,MC	-	+	-	+	+	-	-	+	-	65
+DEHP,Sper,Lop,Asp	-	-	+	-	-	+	+	-	+	77
+Sn,DEHP,MC,Sper	-	+	-	-	+	-	+	-	+	71 - a
+CC,DEHP,MC,Lop,For	+	+	-	+	-	+	-	-	+	88
+Sn,CC,DEHP,Asp	-	-	+	+	+	-	-	-	+	80
+BHA,DEHP,MC,Asp,For	+	+	+	-	-	-	-	+	+	68
+Sn,BHA,DEHP,Lop	-	-	-	-	+	+	-	+	+	69
+CC,BHA,DEHP,Sper,For	+	-	-	+	-	-	+	+	+	72 + a
+all compounds at MOAEL	+	+	+	+	+	+	+	+	+	82

441 Table 5: Aliasing structure between each main effect and two-factor interactions in
 442 the quaternary-code design used in Study 2.

$$\begin{aligned}
 E(\hat{\beta}_{For}) = & \beta_{For} + \frac{1}{2}(\beta_{MC \times Asp} + \beta_{MC \times Lop} - \beta_{MC \times Sper} + \beta_{MC \times DEHP} \\
 & - \beta_{Asp \times CC} + \beta_{Asp \times Sn} + \beta_{Asp \times BHA} + \beta_{CC \times Lop} \\
 & + \beta_{CC \times Sper} + \beta_{CC \times DEHP} + \beta_{Sn \times Lop} + \beta_{Sn \times Sper} \\
 & - \beta_{Sn \times DEHP} - \beta_{Lop \times BHA} + \beta_{Sper \times BHA} + \beta_{BHA \times DEHP})
 \end{aligned}$$

$$E(\hat{\beta}_{MC}) = \beta_{MC} + \frac{1}{2}(\beta_{For \times Asp} + \beta_{For \times Lop} - \beta_{For \times Sper} + \beta_{For \times DEHP})$$

$$E(\hat{\beta}_{Asp}) = \beta_{Asp} + \frac{1}{2}(\beta_{For \times MC} - \beta_{For \times CC} + \beta_{For \times Sn} + \beta_{For \times BHA})$$

$$E(\hat{\beta}_{CC}) = \beta_{CC} + \frac{1}{2}(-\beta_{For \times Asp} + \beta_{For \times Lop} + \beta_{For \times Sper} + \beta_{For \times DEHP})$$

$$E(\hat{\beta}_{Sn}) = \beta_{Sn} + \frac{1}{2}(\beta_{For \times Asp} + \beta_{For \times Lop} + \beta_{For \times Sper} - \beta_{For \times DEHP})$$

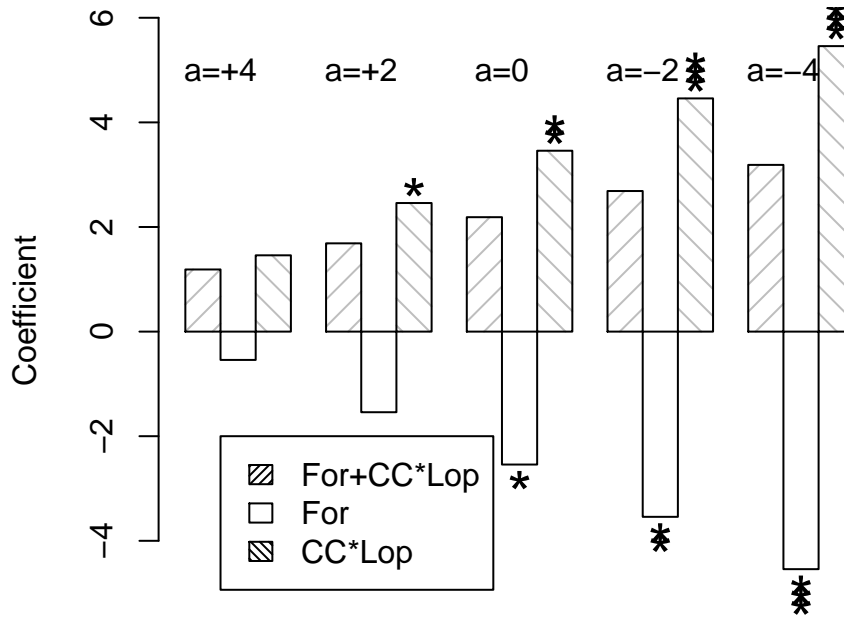
$$E(\hat{\beta}_{Lop}) = \beta_{Lop} + \frac{1}{2}(\beta_{For \times MC} + \beta_{For \times CC} + \beta_{For \times Sn} - \beta_{For \times BHA})$$

$$E(\hat{\beta}_{Sper}) = \beta_{Sper} + \frac{1}{2}(-\beta_{For \times MC} + \beta_{For \times CC} + \beta_{For \times Sn} + \beta_{For \times BHA})$$

$$E(\hat{\beta}_{BHA}) = \beta_{BHA} + \frac{1}{2}(\beta_{For \times Asp} - \beta_{For \times Lop} + \beta_{For \times Sper} + \beta_{For \times DEHP})$$

$$E(\hat{\beta}_{DEHP}) = \beta_{DEHP} + \frac{1}{2}(\beta_{For \times MC} + \beta_{For \times CC} - \beta_{For \times Sn} + \beta_{For \times BHA})$$

452 Figure 1. A comparison of the magnitudes and the significance of the estimated
 453 coefficients of, ($For + CC \times Lop$) in the final equation of $ASAT$ using the regular
 454 design with the corresponding magnitudes and coefficients for For and $CC \times Lop$
 455 in the final equation of $ASAT$ using the nonregular design when the value of “ a ”
 456 varies from +4, +2, 0, -2 to -4. The height of a bar represents the magnitude of
 457 the estimate and the number of asterisks represents the significance level ($0.01 <^* P < 0.05$, $0.001 <^{**} P < 0.01$ and $^{***} P < 0.001$).



458