**Supplementary Materials for Venturing beyond the low-hanging fruit: Applying behavioural insights to child protection**

**Contents**

[SECTION ONE: POWER CALCULATIONS 2](#_Toc531183975)

[SECTION TWO: EMAIL TEXT ACCOMPANYING TRIAL LETTERS 3](#_Toc531183976)

[SECTION THREE: COMPOSITION OF BASELINE AND TRIAL BY RELATIVE DISTRIBUTION OF PROFESSION TYPE 5](#_Toc531183977)

[SECTION FOUR: RANDOMIZATION PROCEDURE 6](#_Toc531183978)

[SECTION FIVE: DATA ANALYSIS R CODE AND OUTCOMES 11](#_Toc531183979)

[A) Balance test: Is the *Baseline* sample comparable to our *Trial* sample? Comparison by profession type. 11](#_Toc531183980)

[B) Balance test for the randomization process: Proportion of professions between trial conditions 12](#_Toc531183981)

[C) Re report outcome GAM analysis: *New* letter (e.g. New:Gain and New:Loss combined) vs Control and Baseline 13](#_Toc531183982)

[D) Testing for the effect of the minor wording change in Response Guide 18](#_Toc531183983)

[E) Re report outcome GAM analysis: Comparison of all four letter conditions 19](#_Toc531183984)

[F) R Code for Weekly Report Outcome Proportions 24](#_Toc531183985)

# **SECTION ONE: POWER CALCULATIONS**

The standardized effect size for a chi-square test is calculated as

Where is the null hypothesis proportion for cell of cells in the contingency table, and is the alternative hypothesis value.  With a 4x2 contingency table, there are numerous combinations of proportions that could lead to the same . So, we calculated power for exemplar effect sizes, all considered small.  For example:

Nocall ROSH Closed Other Ntotal Effect Size

Control .62 .17 .13 .08 535 .15 (small)

New: Gain .68 .17 .1 .05

Control .62 .17 .13 .08 383 .18 (small)

New: Gain .68 .17 .07 .08

Control .62 .17 .13 .08 1532 .09 (small)

New: Gain .65 .17 .1 .08

A-priori sample size estimates were calculated assuming a small difference between the control condition and the trial condition expected to show the smaller change (*New:Gain*). This revealed the need for a total N of 1,532 to have at least .80 power to detect a small effect (.09):

χ² tests - Goodness-of-fit tests: Contingency tables

Analysis: A priori: Compute required sample size

Input: Effect size w = 0.0915133

α err prob = 0.05

Power (1-β err prob) = .8

Df = 5

Output: Noncentrality parameter λ = 12.830016

Critical χ² = 11.070498

Total sample size = 1532

Actual power = 0.800086

Thus, our actual estimated sample size of 2,400 gave us in excess of .95 power to detect an effect of similar magnitude (.09):

χ² tests - Goodness-of-fit tests: Contingency tables

Analysis: Post hoc: Compute achieved power

Input: Effect size w = 0.0915133

α err prob = 0.05

Total sample size = 2400

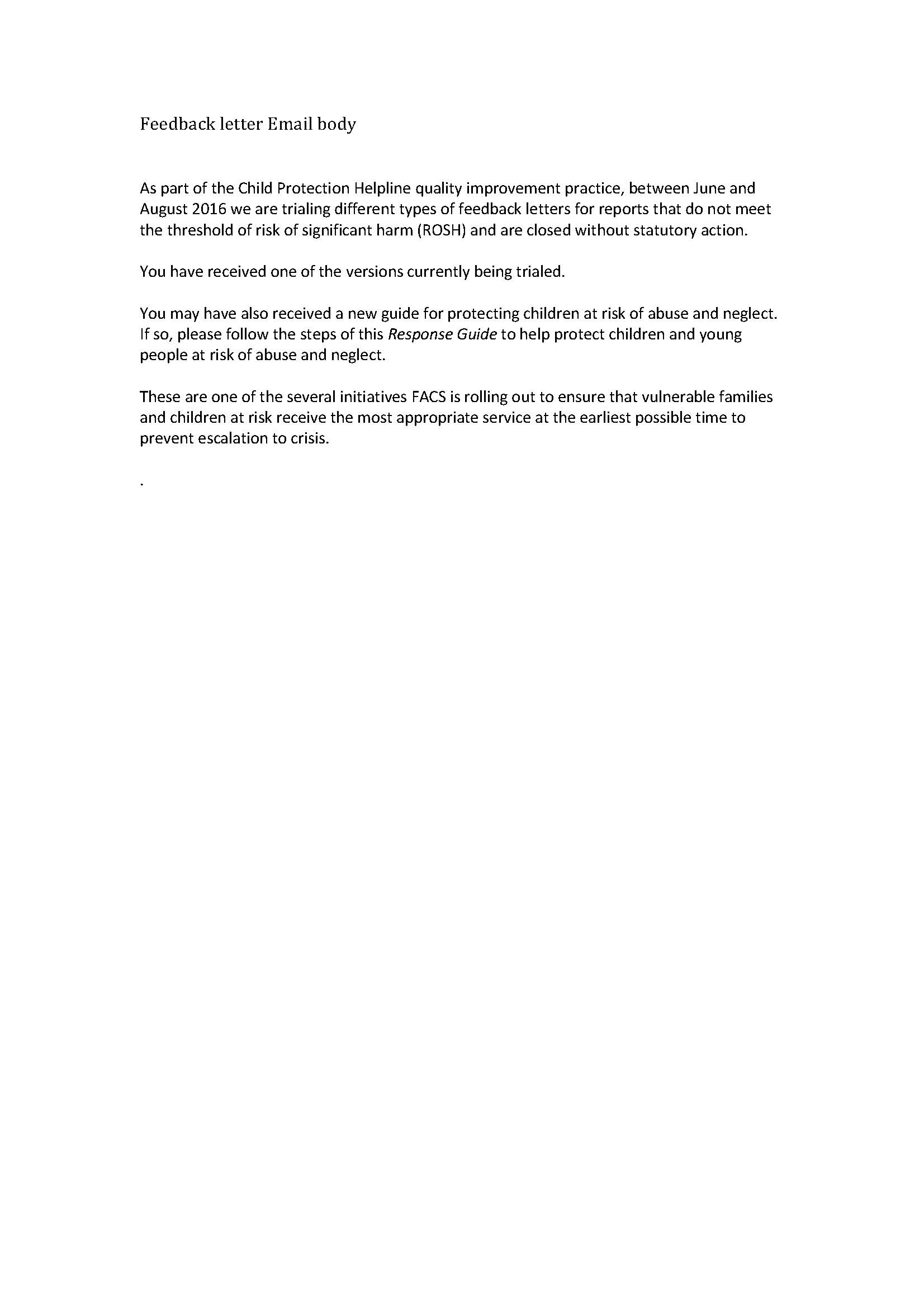
Df = 5

Output: Noncentrality parameter λ = 20.099242

Critical χ² = 11.070498

Power (1-β err prob) = 0.953362

# **SECTION TWO: EMAIL TEXT ACCOMPANYING TRIAL LETTERS**

Figure A

# **SECTION THREE: COMPOSITION OF BASELINE AND TRIAL BY RELATIVE DISTRIBUTION OF PROFESSION TYPE**

Table A and Figure B below portray the number and proportion of profession types captured in the *Baseline* and *Trial* samples. A chi-squared balance test of profession type revealed no significant difference in the proportion of profession type between the *Trial* and *Baseline* sample (GAM: F=17.02, p = 0.2323).

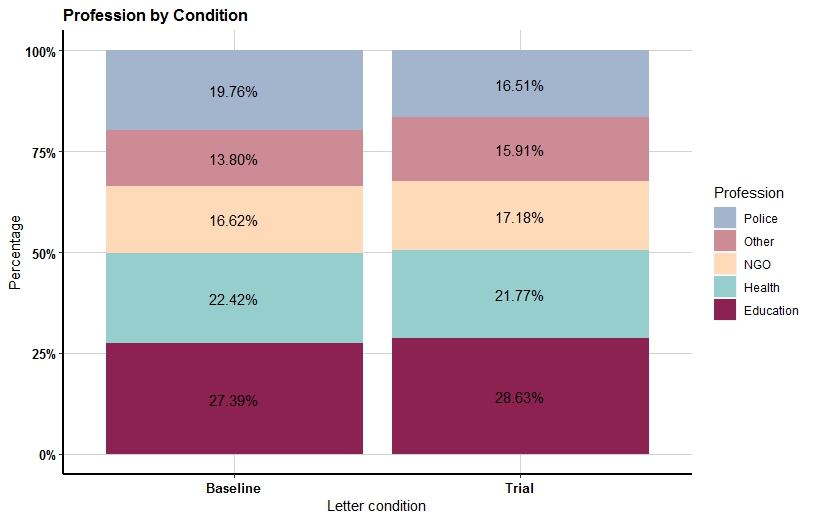
Table A

*Number of mandatory reporters by profession type within Baseline and Trial.*

|  |  |  |
| --- | --- | --- |
| **Profession** | **Baseline** | **Trial\*** |
| **Education** | 1020 | 943 |
| **Health** | 835 | 717 |
| **NGO** | 619 | 566 |
| **Other** | 514 | 524 |
| **Police** | 736 | 544 |
| *\*Combined Control, New:Gain, and New:Loss conditions* | | |

Figure B

Proportion of mandatory reporters by profession type within *Baseline* and *Trial*

****

# **SECTION FOUR: RANDOMIZATION PROCEDURE**

When a mandatory reporter makes a report, a caseworker processes the information they provide. Caseworkers place the processed reports in a shared electronic folder with sub-files for each report outcome (e.g. *Closed*, *ROSH*, etc.). This folder is accessed by Community Services Officers (CSOs) who generate, send, and save the feedback letters in another folder. As such, the randomization process was conducted by the CSOs.

During the trial, for each *Closed* feedback letter, CSOs used an on-line random number generator (as depicted in Figure C) to generate a number between 1-3 to determine which of the three trial letter templates (labelled “1”, “2”, or “3”) to use. Multiple steps were taken to encourage CSOs to engage appropriately in this process. Prior to the trial launch, the CSO manager discussed the trial with CSOs during multiple team meetings, researchers met with 14 of the 17 CSOs (3 were on leave during training) in small groups to answer any questions they had and train them as to the randomization process and its importance, each CSO was provided with a one-page instruction sheet regarding use of the random number generator (see Figure C), and a launch party was held for all CSOs involved. During the trial period, the structure of the shared electronic file was altered so that to save the generated feedback letters, CSOs had to open the folder labelled “Letter 1”, “Letter 2” or “Letter 3”. The letter templates contained a highlighted prompt (e.g. “Have you used the random generator? This is **letter 2!**”) in the field where CSOs insert the reporters name. A researcher was physically present to support CSOs on the first two days and then visited on a weekly basis throughout the trial period. In addition, researchers provided CSOs with visual aid prompts on their computers (see Figure D). Researchers instructed CSOs that there would be no repercussions for forgetting to use the random number generator, but that if they did, it was important to let researchers know. No CSOs raised issues with researchers.

Post-trial, researchers interviewed nine CSOs (i.e. All CSOs present on the interview day; see Figure E). Seven CSOs reported having no difficulties remembering to randomise, the other two reported having difficulties in the first week if interrupted (e.g. by phone calls), but offered that they engaged in strategies to overcome this such as by recording the randomised value before answering the phone. Four stated they needed no assistance with the process as it was “easy”, and the other five indicated that they successfully got help from their teammates. The CSO manager reported that less than once or twice a week a CSO stated that they forgot to use the random number generator, but indicated that this was in relation to isolated reports not a string of reports. All nine indicated their involvement in the trial was a positive experience.

Table B reveals that trial letter allocation by data collection date intervals (approximately 1 week at a time) was roughly equal across time.

Table B

*Number of letters allocated per condition across data collection periods*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Letter** | **11 - 17 July** | **18 - 24 July** | **25 - 31 July** | **1 - 7 Aug** | **8 - 14 Aug** | **15 - 21 Aug** | **22 - 28 Aug** | **29 - 4 Aug** | **5 - 11 Sep** | **Total number sent** | **Proportion of total** |
| Control | 242 | 98 | 117 | 137 | 187 | 141 | 129 | 143 | 106 | **1300** | **34%** |
| New:Gain | 185 | 92 | 113 | 120 | 171 | 156 | 144 | 155 | 158 | **1294** | **33%** |
| New:Loss | 224 | 100 | 108 | 138 | 169 | 139 | 143 | 135 | 123 | **1279** | **33%** |
| Total | 651 | 290 | 338 | 395 | 527 | 436 | 416 | 433 | 387 | **3873** | **100%** |

Furthermore, balance checks on the three conditions revealed that there was no significant difference in the allocation of mandatory reporters to conditions in terms of the profession type proportions (GAM: F = 14.442, *p* = 0.51) – i.e. there was the same proportion of each profession type in all three trial conditions. Together, there was no evidence indicating that the CSOs failed to engage appropriately in the randomization process.

Figure C

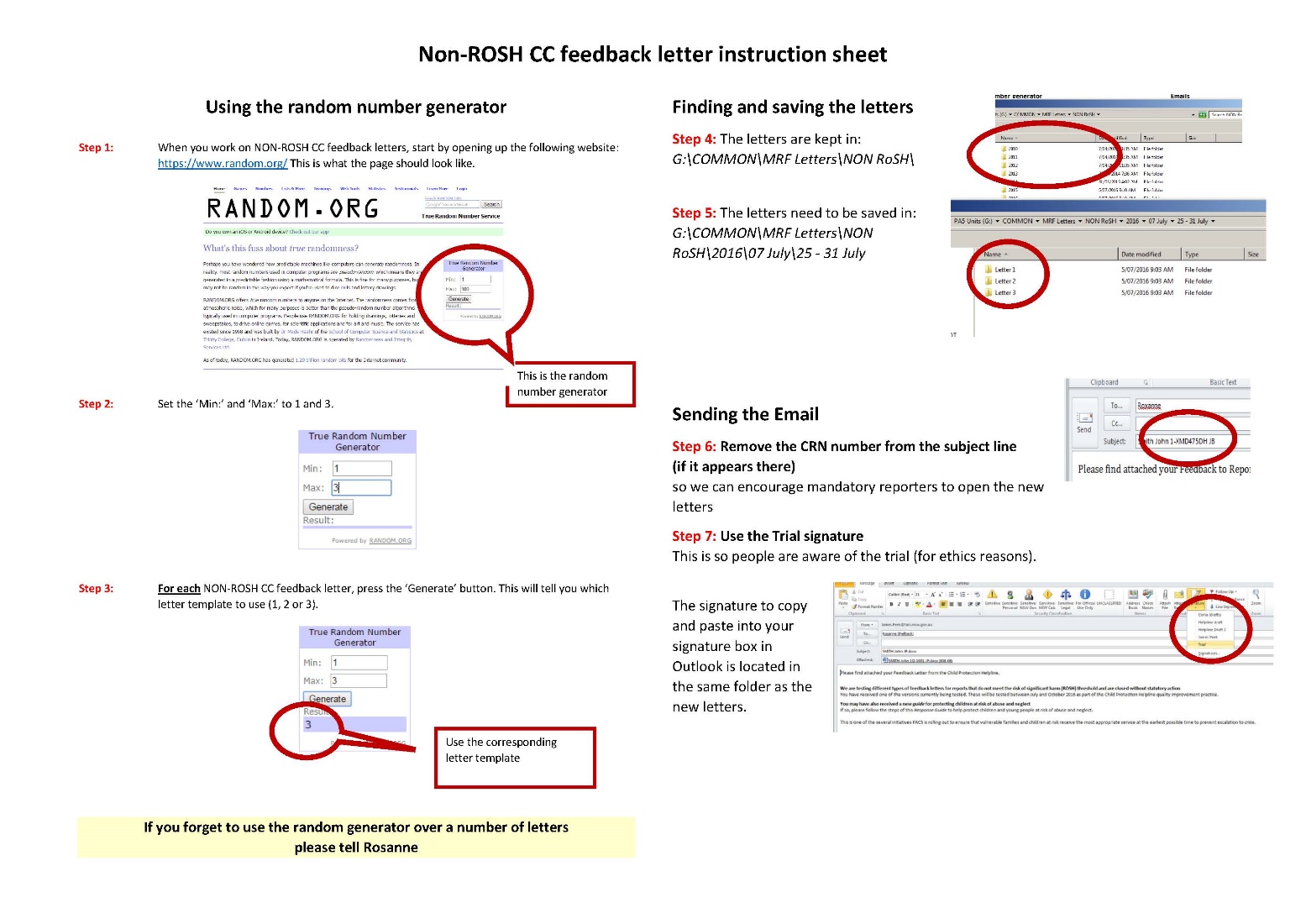


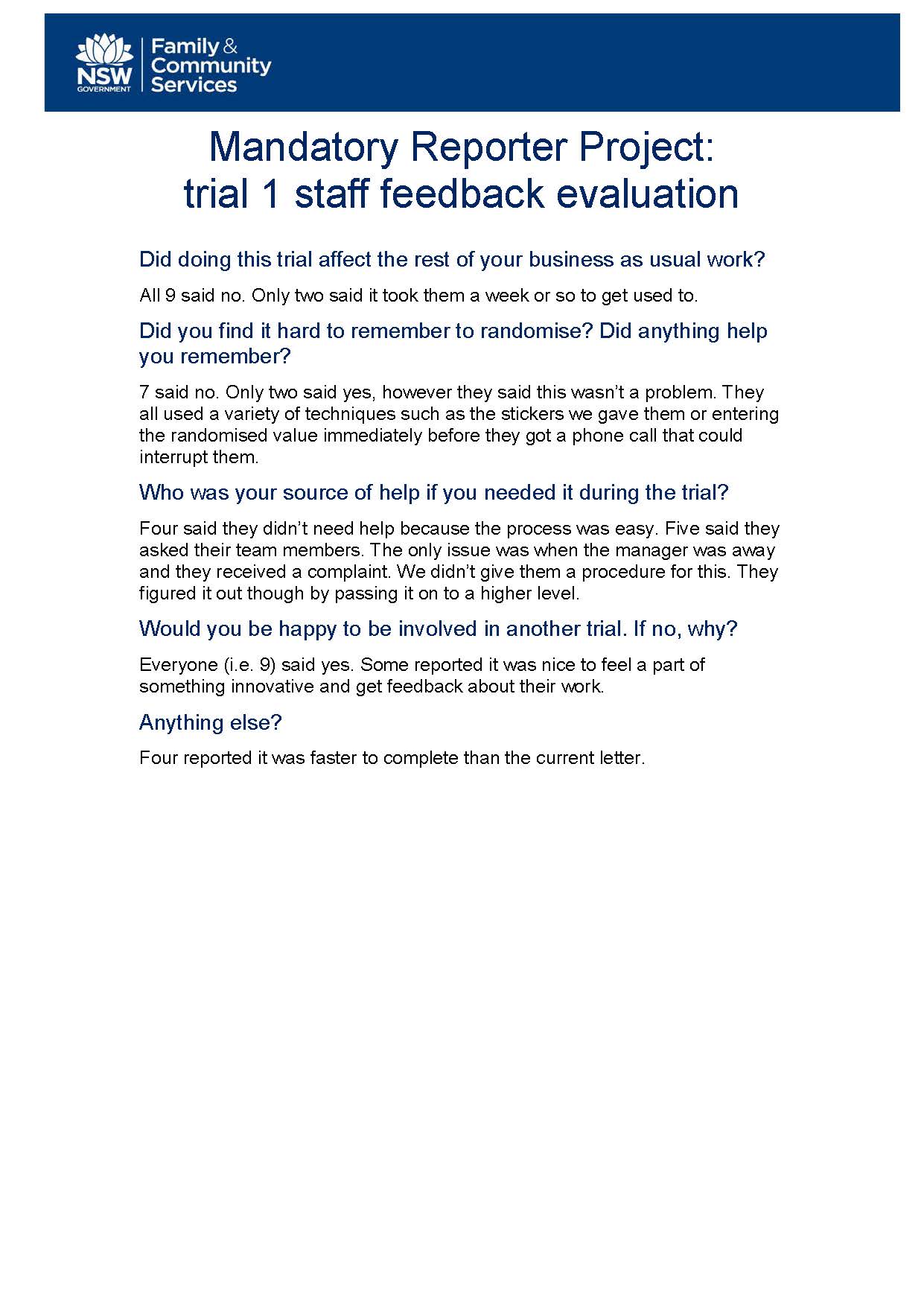
Figure D

Visual aid prompts on CSO’s computers.

**

Figure E

CSO trial process feedback.

**

# **SECTION FIVE: DATA ANALYSIS R CODE AND OUTCOMES**

## **A) Balance test: Is the *Baseline* sample comparable to our *Trial* sample? Comparison by profession type.**

> rm(list=ls())

> library(mgcv)

> library(multcomp)

> library(ggplot2)

> Alldata=read.csv("MRP.csv")

> Alldata$Letter=relevel(Alldata$Letter,ref="Baseline")

> Alldata$Profession=factor(Alldata$Profession)

> unique(Alldata$Letter)

[1] New:Loss Control New:Gain Baseline

Levels: Baseline Control New:Gain New:Loss

> levels(Alldata$Profession)

[1] "Education" "Health" "NGO" "Other" "Police"

> levels(Alldata$Letter)

[1] "Baseline" "Control" "New:Gain" "New:Loss"

> levels(Alldata$Letter) <- c ("Baseline", "Trial", "Trial", "Trial")

> Alldata$outcome.num=as.integer(Alldata$Letter)-1

> form.full=list(outcome.num~Profession,~Profession,~Profession, ~Profession,~Profession)

> full.mod <- gam(form.full,data=Alldata,family=multinom(K=5))

> new.dat=expand.grid(Profession=c("Education","Health", "NGO", "Other", "Police"))

> new.dat$pred=predict(full.mod,newdata=new.dat,type="response")

> form.null=list(outcome.num~1,~1,~1,~1,~1)

> null.mod <- gam(form.null,data=Alldata,family=multinom(K=5))

> anova(null.mod,full.mod,test="Chisq")

Analysis of Deviance Table

Model 1: outcome.num ~ 1

Model 2: outcome.num ~ Profession

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 7014.6 2254.7

1. 7001.0 2237.6 13.607 17.015 0.2323

## **B) Balance test for the randomization process: Proportion of professions between trial conditions**

> rm(list=ls())

> library(mgcv)

> library(multcomp)

> library(ggplot2)

> Alldata=read.csv("MRP.csv")

> Trialdf <- subset(Alldata, Letter != "Baseline")

> Trialdf <- droplevels(Trialdf)

> levels(Trialdf$Profession)

[1] "Education" "Health" "NGO" "Other" "Police"

> levels(Trialdf$Letter)

[1] "Control" "New:Gain" "New:Loss"

> Table1 <-table(Trialdf$Profession, Trialdf$Letter)

> Table1

Control New:Gain New:Loss

Education 308 329 306

Health 239 251 227

NGO 211 158 197

Other 176 182 166

Police 185 161 198

> Trialdf$outcome.num=as.integer(Trialdf$Letter)-1

> form.full=list(outcome.num~Profession,~Profession,~Profession, ~Profession,~Profession)

> full.mod <- gam(form.full,data=Trialdf,family=multinom(K=5))

> new.dat=expand.grid(Profession=c("Education","Health", "NGO", "Other", "Police"))

> new.dat$pred=predict(full.mod,newdata=new.dat,type="response")

> form.null=list(outcome.num~1,~1,~1,~1,~1)

> null.mod <- gam(form.null,data=Trialdf,family=multinom(K=5))

> anova(null.mod,full.mod,test="Chisq")

Analysis of Deviance Table

Model 1: outcome.num ~ 1

Model 2: outcome.num ~ Profession

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 3290.2 4999.0

2 3275.0 4984.6 15.205 14.422 0.5091

## **C) Re report outcome GAM analysis: *New* letter (e.g. New:Gain and New:Loss combined) vs Control and Baseline**

>rm(list=ls()), library(mgcv), library(multcomp, library(ggplot2)

>Alldata=read.csv("MRP.csv")

>Alldata$Re.Report.Outcome=relevel(Alldata$Re.Report.Outcome,ref="NoCall")

>Alldata$Letter=factor(Alldata$Letter)

>unique(Alldata$Re.Report.Outcome)

[1] ROSH Other Closed NoCall

Levels: NoCall Closed Other ROSH

>levels(Alldata$Letter)

[1] "Baseline" "Control" "New:Gain" "New:Loss"

>levels(Alldata$Letter)<-c("Baseline","Control","New","New")

>levels(Alldata$Letter)

[1] "Baseline" "Control" "New"

>Alldata$outcome.num=as.integer(Alldata$Re.Report.Outcome)-1

>form.full=list(outcome.num~Letter,~Letter,~Letter)

>full.mod <-gam(form.full,data=Alldata,family=multinom(K=3))

>new.dat=expand.grid(Letter=c("Baseline","Control", "New"))

>new.dat$pred=predict(full.mod,newdata=new.dat,type="response")

>form.null=list(outcome.num~1,~1,~1)

>null.mod <-gam(form.null,data=Alldata,family=multinom(K=3))

>anova(null.mod,full.mod,test="Chisq")

Analysis of Deviance Table

Model 1: outcome.num ~ 1

Model 2: outcome.num ~ Letter

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 7015 5309.3

2 7009 5285.0 6 24.361 0.000448 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

To answer whether each groups proportion differed by treatment we fit each outcome against all the others. First we set up a matrix to store the p values, and then we fit a model for each outcome against everything else and saved the p values.

>p.val=matrix(NA,3,4)

>colnames(p.val)=unique(Alldata$Re.Report.Outcome)

>rownames(p.val)=c("BaselineControl","BaselineNew","ControlNew")

>p.val

ROSH Other Closed NoCall

BaselineControl NA NA NA NA

BaselineNew NA NA NA NA

ControlNew NA NA NA NA

> Alldata\_ROSH=Alldata

> levels(Alldata\_ROSH$Re.Report.Outcome)=c(levels(Alldata\_ROSH$Re.Report.Outcome),"Other")

> Alldata\_ROSH$Re.Report.Outcome[Alldata\_ROSH$Re.Report.Outcome!="ROSH"]="Other"

> ROSH\_mod=glm(Re.Report.Outcome~Letter,family = "binomial",data=Alldata\_ROSH)

> res=glht(ROSH\_mod,linfct=mcp(Letter="Tukey"))[[1]](#footnote-1)

> out=summary(res, test = adjusted("none"))

> out

Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Fit: glm(formula = Re.Report.Outcome ~ Letter, family = "binomial",

data = Alldata\_ROSH)

Linear Hypotheses:

Estimate Std. Error z value Pr(>|z|)

Control - Baseline == 0 0.09200 0.08808 1.044 0.296

New - Baseline == 0 -0.03150 0.07178 -0.439 0.661

New - Control == 0 -0.12350 0.09572 -1.290 0.197

(Adjusted p values reported -- none method)

> out$test$pvalues[1:3]

Control - Baseline New - Baseline New - Control

0.2962754 0.6607491 0.1969376

> p.val[,1]=out$test$pvalues[1:3]

> Alldata\_Other=Alldata

> levels(Alldata\_Other$Re.Report.Outcome)=c(levels(Alldata\_Other$Re.Report.Outcome),"Alternative")

> Alldata\_Other$Re.Report.Outcome[Alldata\_Other$Re.Report.Outcome!="Other"]="Alternative"

> Other\_mod=glm(Re.Report.Outcome~Letter,family = "binomial",data=Alldata\_Other)

> out=summary(Other\_mod)

> res=glht(Other\_mod,linfct=mcp(Letter="Tukey"))

> out=summary(res, test = adjusted("none"))

> out

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: glm(formula = Re.Report.Outcome ~ Letter, family = "binomial",

data = Alldata\_Other)

Linear Hypotheses:

Estimate Std. Error z value Pr(>|z|)

Control - Baseline == 0 0.1534 0.1397 1.099 0.2720

New - Baseline == 0 0.2627 0.1132 2.321 0.0203 \*

New - Control == 0 0.1093 0.1556 0.703 0.4822

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- none method)

> out$test$pvalues[1:3]

Control - Baseline New - Baseline New - Control

0.27196439 0.02027832 0.48220351

> p.val[,2]=out$test$pvalues[1:3]

> Alldata\_Closed=Alldata

> levels(Alldata\_Closed$Re.Report.Outcome)=c(levels(Alldata\_Closed$Re.Report.Outcome),"Other")

> Alldata\_Closed$Re.Report.Outcome[Alldata\_Closed$Re.Report.Outcome!="Closed"]="Other"

> Closed\_mod=glm(Re.Report.Outcome~Letter,family = "binomial",data=Alldata\_Closed)

> out=summary(Closed\_mod)

> res=glht(Closed\_mod,linfct=mcp(Letter="Tukey"))

> out=summary(res, test = adjusted("none"))

> out

Simultaneous Tests for General Linear Hypotheses, Multiple Comparisons of Means: Tukey Contrasts. Fit: glm(formula = Re.Report.Outcome ~ Letter, family = "binomial",

data = Alldata\_Closed)

Linear Hypotheses:

Estimate Std. Error z value Pr(>|z|)

Control - Baseline == 0 0.09221 0.10385 0.888 0.374594

New - Baseline == 0 0.31077 0.08661 3.588 0.000333 \*\*\*

New - Control == 0 0.21856 0.11646 1.877 0.060558 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- none method)

> out$test$pvalues[1:3]

Control - Baseline New - Baseline New - Control

0.374593703 0.000333047 0.060558024

> p.val[,3]=out$test$pvalues[1:3]

> Alldata\_NoCall=Alldata

> levels(Alldata\_NoCall$Re.Report.Outcome)=c(levels(Alldata\_NoCall$Re.Report.Outcome),"Other")

> Alldata\_NoCall$Re.Report.Outcome[Alldata\_NoCall$Re.Report.Outcome!="NoCall"]="Other"

> NoCall\_mod=glm(Re.Report.Outcome~Letter,family = "binomial",data=Alldata\_NoCall)

> out=summary(NoCall\_mod)

> res=glht(NoCall\_mod,linfct=mcp(Letter="Tukey"))

> out=summary(res, test = adjusted("none"))

> out

Simultaneous Tests for General Linear Hypotheses, Multiple Comparisons of Means: Tukey Contrasts. Fit: glm(formula = Re.Report.Outcome ~ Letter, family = "binomial",

data = Alldata\_NoCall)

Linear Hypotheses:

Estimate Std. Error z value Pr(>|z|)

Control - Baseline == 0 -0.02591 0.07059 -0.367 0.71354

New - Baseline == 0 -0.22667 0.05697 -3.979 6.93e-05 \*\*\*

New - Control == 0 -0.20075 0.07706 -2.605 0.00918 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- none method)

> out$test$pvalues[1:3]

Control - Baseline New - Baseline New - Control

7.135407e-01 6.933581e-05 9.182374e-03

> p.val[,4]=out$test$pvalues[1:3]

> p.val

Raw P values:

ROSH Other Closed NoCall

BaselineControl 0.2962754 0.27196439 0.374593703 7.135407e-01

BaselineNew 0.6607491 0.02027832 0.000333047 6.933581e-05

ControlNew 0.1969376 0.48220351 0.060558024 9.182374e-03

> p.val.adj=matrix(NA,3,4)

> p.val.adj[1,]=p.adjust(p.val[1,], method = "holm")

> p.val.adj[2,]=p.adjust(p.val[2,], method = "holm")

> p.val.adj[3,]=p.adjust(p.val[3,], method = "holm")

> colnames(p.val.adj)=unique(Alldata$Re.Report.Outcome)

> rownames(p.val.adj)=c("BaselineControl", "BaselineNew","ControlNew")

> p.val.adj

Adjusted P values:

ROSH Other Closed NoCall

BaselineControl 1.0000000 1.00000000 1.0000000000 1.0000000000

BaselineNew 0.6607491 0.04055664 0.0009991409 0.0002773432

ControlNew 0.3938752 0.48220351 0.1816740735 0.0367294962

## **D) Testing for the effect of the minor wording change in Response Guide**

> #OVERALL TEST

> rm(list=ls())

> library(mgcv)

> library(multcomp)

> #setwd

> Alldata=read.csv("MRP.csv")

> #make NoCall the reference

> Alldata$Re.Report.Outcome=relevel(Alldata$Re.Report.Outcome,ref="NoCall")

> Alldata$Letter=factor(Alldata$Letter)

> unique(Alldata$Re.Report.Outcome)

[1] ROSH Other Closed NoCall

Levels: NoCall Closed Other ROSH

> #create subset data without Baseline

> Trialdf = subset(Alldata, Letter != "Baseline")

> #test for effect of letter change

> Trialdf$outcome.num=as.integer(Trialdf$Re.Report.Outcome)-1

> form.noint=list(outcome.num~Letter+Before.date1,~Letter+Before.date1,

+ ~Letter+Before.date1)

> full.noint <- gam(form.noint,data=Trialdf,family=multinom(K=3))

> form.int=list(outcome.num~Letter\*Before.date1,~Letter\*Before.date1,

+ ~Letter\*Before.date1)

> full.int <- gam(form.int,data=Trialdf,family=multinom(K=3))

> anova(full.noint,full.int,test = "Chisq")

Analysis of Deviance Table

Model 1: outcome.num ~ Letter + Before.date1

Model 2: outcome.num ~ Letter \* Before.date1

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 3282 2065.3

2 3276 2057.9 6 7.4679 0.2797

> #if there is a significant effect - run the analysis on the history data

## **E) Re report outcome GAM analysis: Comparison of all four letter conditions**

> rm(list=ls())

> library(mgcv)

> library(multcomp)

> library(ggplot2)

> Alldata=read.csv("MRP.csv")

> Alldata$Re.Report.Outcome=relevel(Alldata$Re.Report.Outcome,ref="NoCall")

> Alldata$Letter=factor(Alldata$Letter)

> unique(Alldata$Re.Report.Outcome)

[1] ROSH Other Closed NoCall

Levels: NoCall Closed Other ROSH

> Alldata$outcome.num=as.integer(Alldata$Re.Report.Outcome)-1

> form.full=list(outcome.num~Letter,~Letter,~Letter,~Letter)

> full.mod <- gam(form.full,data=Alldata,family=multinom(K=4))

> new.dat=expand.grid(Letter=c("Baseline","Control","New:Gain", "New:Loss"))

> new.dat$pred=predict(full.mod,newdata=new.dat,type="response")

> form.null=list(outcome.num~1,~1,~1,~1)

> null.mod <- gam(form.null,data=Alldata,family=multinom(K=4))

> anova(null.mod,full.mod,test="Chisq")

Analysis of Deviance Table

Model 1: outcome.num ~ 1

Model 2: outcome.num ~ Letter

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 7014.4 5309.3

2 7003.6 5281.0 10.801 28.272 0.002627 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> p.val=matrix(NA,6,4)

> colnames(p.val)=unique(Alldata$Re.Report.Outcome)

> rownames(p.val)=c("BaselineControl", "BaselineNew:Gain","BaselineNew:Loss", "ControlNew:Gain", "ControlNew:Loss","New:GainNew:Loss")

> Alldata\_ROSH=Alldata

> levels(Alldata\_ROSH$Re.Report.Outcome)=c(levels(Alldata\_ROSH$Re.Report.Outcome),"Other")

> Alldata\_ROSH$Re.Report.Outcome[Alldata\_ROSH$Re.Report.Outcome!="ROSH"]="Other"

> ROSH\_mod=glm(Re.Report.Outcome~Letter,family = "binomial",data=Alldata\_ROSH)

> res=glht(ROSH\_mod,linfct=mcp(Letter= "Tukey"))

> out=summary(res, test = adjusted("none"))

> out

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: glm(formula = Re.Report.Outcome ~ Letter, family = "binomial",

data = Alldata\_ROSH)

Linear Hypotheses:

Estimate Std. Error z value Pr(>|z|)

Control - Baseline == 0 0.091999 0.088084 1.044 0.296

New:Gain - Baseline == 0 -0.060829 0.092739 -0.656 0.512

New:Loss - Baseline == 0 -0.003078 0.090946 -0.034 0.973

New:Gain - Control == 0 -0.152829 0.112290 -1.361 0.174

New:Loss - Control == 0 -0.095078 0.110813 -0.858 0.391

New:Loss - New:Gain == 0 0.057751 0.114549 0.504 0.614

(Adjusted p values reported -- none method)

> out$test$pvalues[1:6]

Control - Baseline New:Gain - Baseline New:Loss - Baseline New:Gain - Control New:Loss - Control New:Loss - New:Gain

0.2962754 0.5118780 0.9730004 0.1735072 0.3908932 0.6141473

> p.val[,1]=out$test$pvalues[1:6]

> Alldata\_Other=Alldata

> levels(Alldata\_Other$Re.Report.Outcome)=c(levels(Alldata\_Other$Re.Report.Outcome),"Alternative")

> Alldata\_Other$Re.Report.Outcome[Alldata\_Other$Re.Report.Outcome!="Other"]="Alternative"

> Other\_mod=glm(Re.Report.Outcome~Letter,family = "binomial",data=Alldata\_Other)

> out=summary(Other\_mod)

> res=glht(Other\_mod, linfct=mcp(Letter= "Tukey"))

> out=summary(res, test = adjusted("none"))

> out

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: glm(formula = Re.Report.Outcome ~ Letter, family = "binomial",

data = Alldata\_Other)

Linear Hypotheses:

Estimate Std. Error z value Pr(>|z|)

Control - Baseline == 0 0.15343 0.13967 1.099 0.27196

New:Gain - Baseline == 0 0.10119 0.13906 0.728 0.46679

New:Loss - Baseline == 0 0.44903 0.15693 2.861 0.00422 \*\*

New:Gain - Control == 0 -0.05224 0.17527 -0.298 0.76567

New:Loss - Control == 0 0.29559 0.18976 1.558 0.11930

New:Loss - New:Gain == 0 0.34783 0.18931 1.837 0.06615 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- none method)

> out$test$pvalues[1:6]

Control - Baseline New:Gain - Baseline New:Loss - Baseline New:Gain - Control New:Loss - Control New:Loss - New:Gain

0.27196439 0.46678910 0.00421781 0.76566624 0.11929575 0.06615417

> p.val[,2]=out$test$pvalues[1:6]

> Alldata\_Closed=Alldata

> levels(Alldata\_Closed$Re.Report.Outcome)=c(levels(Alldata\_Closed$Re.Report.Outcome),"Other")

> Alldata\_Closed$Re.Report.Outcome[Alldata\_Closed$Re.Report.Outcome!="Closed"]="Other"

> Closed\_mod=glm(Re.Report.Outcome~Letter,family = "binomial",data=Alldata\_Closed)

> out=summary(Closed\_mod)

> res=glht(Closed\_mod, linfct=mcp(Letter= "Tukey"))

> out=summary(res, test = adjusted("none"))

> out

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: glm(formula = Re.Report.Outcome ~ Letter, family = "binomial",

data = Alldata\_Closed)

Linear Hypotheses:

Estimate Std. Error z value Pr(>|z|)

Control - Baseline == 0 0.09221 0.10385 0.888 0.37459

New:Gain - Baseline == 0 0.27363 0.11136 2.457 0.01400 \*

New:Loss - Baseline == 0 0.34860 0.11355 3.070 0.00214 \*\*

New:Gain - Control == 0 0.18141 0.13588 1.335 0.18183

New:Loss - Control == 0 0.25638 0.13768 1.862 0.06258 .

New:Loss - New:Gain == 0 0.07497 0.14343 0.523 0.60119

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- none method)

> out$test$pvalues[1:6]

Control - Baseline New:Gain - Baseline New:Loss - Baseline New:Gain - Control New:Loss - Control New:Loss - New:Gain

0.374593703 0.014004397 0.002141627 0.181826885 0.062580412 0.601186242

> p.val[,3]=out$test$pvalues[1:6]

> Alldata\_NoCall=Alldata

> Alldata\_NoCall$Re.Report.Outcome[Alldata\_NoCall$Re.Report.Outcome!="NoCall"]="Other"

> levels(Alldata\_NoCall$Re.Report.Outcome)=c(levels(Alldata\_NoCall$Re.Report.Outcome),"Other")

> NoCall\_mod=glm(Re.Report.Outcome~Letter,family = "binomial",data=Alldata\_NoCall)

> out=summary(NoCall\_mod)

> res=glht(NoCall\_mod, linfct=mcp(Letter= "Tukey"))

> out=summary(res, test = adjusted("none"))

> out

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: glm(formula = Re.Report.Outcome ~ Letter, family = "binomial",

data = Alldata\_NoCall)

Linear Hypotheses:

Estimate Std. Error z value Pr(>|z|)

Control - Baseline == 0 -0.02591 0.07059 -0.367 0.713541

New:Gain - Baseline == 0 -0.18845 0.07291 -2.585 0.009748 \*\*

New:Loss - Baseline == 0 -0.26494 0.07334 -3.612 0.000303 \*\*\*

New:Gain - Control == 0 -0.16254 0.08949 -1.816 0.069331 .

New:Loss - Control == 0 -0.23903 0.08984 -2.661 0.007801 \*\*

New:Loss - New:Gain == 0 -0.07649 0.09167 -0.834 0.404064

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- none method)

> out$test$pvalues[1:6]

Control - Baseline New:Gain - Baseline New:Loss - Baseline New:Gain - Control New:Loss - Control New:Loss - New:Gain

0.713540741 0.009748262 0.000303425 0.069331023 0.007801175 0.404064216

> p.val[,4]=out$test$pvalues[1:6]

> p.val

**RAW P VALUES**

ROSH Other Closed NoCall

BaselineControl 0.2962754 0.27196439 0.374593703 0.713540741

BaselineNew:Gain 0.5118780 0.46678910 0.014004397 0.009748262

BaselineNew:Loss 0.9730004 0.00421781 0.002141627 0.000303425

ControlNew:Gain 0.1735072 0.76566624 0.181826885 0.069331023

ControlNew:Loss 0.3908932 0.11929575 0.062580412 0.007801175

New:GainNew:Loss 0.6141473 0.06615417 0.601186242 0.404064216

> p.val.adj=matrix(NA,6,4)

> p.val.adj[1,]=p.adjust(p.val[1,], method = "holm")

> p.val.adj[2,]=p.adjust(p.val[2,], method = "holm")

> p.val.adj[3,]=p.adjust(p.val[3,], method = "holm")

> p.val.adj[4,]=p.adjust(p.val[4,], method = "holm")

> p.val.adj[5,]=p.adjust(p.val[5,], method = "holm")

> p.val.adj[6,]=p.adjust(p.val[6,], method = "holm")

> colnames(p.val.adj)=unique(Alldata$Re.Report.Outcome)

> rownames(p.val.adj)=c("BaselineControl", "BaselineNew:Gain","BaselineNew:Loss", "ControlNew:Gain", "ControlNew:Loss","New:GainNew:Loss")

> p.val.adj

**ADJUSTED P VALUES**

ROSH Other Closed NoCall

BaselineControl 1.0000000 1.00000000 1.00000000 1.00000000

BaselineNew:Gain 0.9335782 0.93357820 0.04201319 0.03899305

BaselineNew:Loss 0.9730004 0.00843562 0.00642488 0.00121370

ControlNew:Gain 0.5205215 0.76566624 0.52052152 0.27732409

ControlNew:Loss 0.3908932 0.23859150 0.18774124 0.03120470

New:GainNew:Loss 1.0000000 0.26461668 1.00000000 1.00000000

## **F) R Code for Weekly Report Outcome Proportions**

install.packages("qcc")

library(qcc)

>OTHER <- read.csv("Other weekly data.csv")

# makes the test data the default dataset to use

>attach(OTHER)

# Uses the "training" data (indicated in the test dataset as Train=TRUE) to calculate the "in control" limits and plots them. “n[Train]” says use the Number of reports in the Training phase. “sizes=total[Train]” tells R that the total number of reports in that time period is included in the variable total (and only use the Training data). “type="p"” is telling R it is a proportion that we are plotting. “ylim” is a command to scale the y-axis to between 0 - .80. “newdata” is the rest of the data, with Train=False (n[!Train]). “newsizes” are the total for the non-Training time periods.

>q1 <- qcc(n[Train], sizes=total[Train], type="p", ylim=c(0.10,.35))

>q2 <- qcc(n[Train], sizes=total[Train], type="p", ylim=c(0.10,.35),

newdata=n[!Train], newsizes=total[!Train],

xtitle=("label"))

>detach(OTHER)

1. R coding is slightly misleading. The “Tukey” option for the glht function in the multcomp package does not actually use the Tukey correction, it just sets up all pairwise comparisons. P-value adjustments actually occur in the “summary.glht” function. To extract the raw values without the step-wise adjustment being applied we used the following function (out=summary(res, test = adjusted("none")). [↑](#footnote-ref-1)