How to ensure that policy experiments are credible and actionable

Jake Bowers

February 28, 2018

Overall

1.
2.
3.
4.
5.

1. A proposed research integrity process
2. Standard Operating Procedures

The Current DRAFT OES Process

1. **Project Initiation**: early ideas about the project are discussed to ensure general feasibility, proper planning, and wise investment of team resources prior to formally initiating a project and committing to a collaboration with agency partners
2. **Design Review**: the project design is peer reviewed and then presented to the team, to ensure a sound design that effectively addresses research objectives before we invest resources in fielding a study
3. **Analysis Plan Commitment**: an analysis plan (also known as a "pre-analysis plan" or "pre-specification plan") is finalized, date-stamped, and posted publicly on our website before data are received and analyzed
4. **Findings Review**: an initial analysis of results is presented to the team, to ensure that tentative findings are consistent with a sound analysis of the data, that important limitations on the study's findings have been identified, and that alternative explanations have been addressed to the greatest extent possible
5. **Reanalysis**: an internal replication of the initial analysis, to ensure that results and conclusions are sound, reliable, and reproducible
6. **Pre-Publication Review**: to ensure OES maintains transparency, retains materials necessary for reproducibility, and meets all legal and administrative requirements in disseminating knowledge for the whole of government and the public

The Current DRAFT OES Process

1. **Project Initiation**: early ideas about the project are discussed to ensure general feasibility, proper planning, and wise investment of team resources prior to formally initiating a project and committing to a collaboration with agency partners
2. **Design Review**: the project design is peer reviewed and then presented to the team, to ensure a sound design that effectively addresses research objectives before we invest resources in fielding a study
3. **Analysis Plan Commitment**: an analysis plan (also known as a "pre-analysis plan" or "pre-specification plan") is finalized, date-stamped, and posted publicly on our website before data are received and analyzed
4. **Findings Review**: an initial analysis of results is presented to the team, to ensure that tentative findings are consistent with a sound analysis of the data, that important limitations on the study's findings have been identified, and that alternative explanations have been addressed to the greatest extent possible
5. **Reanalysis**: an internal replication of the initial analysis, to ensure that results and conclusions are sound, reliable, and reproducible
6. **Pre-Publication Review**: to ensure OES maintains transparency, retains materials necessary for reproducibility, and meets all legal and administrative requirements in disseminating knowledge for the whole of government and the public

The Current DRAFT OES Process

1. **Project Initiation**: early ideas about the project are discussed to ensure general feasibility, proper planning, and wise investment of team resources prior to formally initiating a project and committing to a collaboration with agency partners
2. **Design Review**: the project design is peer reviewed and then presented to the team, to ensure a sound design that effectively addresses research objectives before we invest resources in fielding a study
3. **Analysis Plan Commitment**: an analysis plan (also known as a "pre-analysis plan" or "pre-specification plan") is finalized, date-stamped, and posted publicly on our website before data are received and analyzed
4. **Findings Review**: an initial analysis of results is presented to the team, to ensure that tentative findings are consistent with a sound analysis of the data, that important limitations on the study's findings have been identified, and that alternative explanations have been addressed to the greatest extent possible
5. **Reanalysis**: an internal replication of the initial analysis, to ensure that results and conclusions are sound, reliable, and reproducible
6. **Pre-Publication Review**: to ensure OES maintains transparency, retains materials necessary for reproducibility, and meets all legal and administrative requirements in disseminating knowledge for the whole of government and the public

The Current DRAFT OES Process

1. **Project Initiation**: early ideas about the project are discussed to ensure general feasibility, proper planning, and wise investment of team resources prior to formally initiating a project and committing to a collaboration with agency partners
2. **Design Review**: the project design is peer reviewed and then presented to the team, to ensure a sound design that effectively addresses research objectives before we invest resources in fielding a study
3. **Analysis Plan Commitment**: an analysis plan (also known as a "pre-analysis plan" or "pre-specification plan") is finalized, date-stamped, and posted publicly on our website before data are received and analyzed
4. **Findings Review**: an initial analysis of results is presented to the team, to ensure that tentative findings are consistent with a sound analysis of the data, that important limitations on the study's findings have been identified, and that alternative explanations have been addressed to the greatest extent possible
5. **Reanalysis**: an internal replication of the initial analysis, to ensure that results and conclusions are sound, reliable, and reproducible
6. **Pre-Publication Review**: to ensure OES maintains transparency, retains materials necessary for reproducibility, and meets all legal and administrative requirements in disseminating knowledge for the whole of government and the public

The Current DRAFT OES Process

1. **Project Initiation**: early ideas about the project are discussed to ensure general feasibility, proper planning, and wise investment of team resources prior to formally initiating a project and committing to a collaboration with agency partners
2. **Design Review**: the project design is peer reviewed and then presented to the team, to ensure a sound design that effectively addresses research objectives before we invest resources in fielding a study
3. **Analysis Plan Commitment**: an analysis plan (also known as a "pre-analysis plan" or "pre-specification plan") is finalized, date-stamped, and posted publicly on our website before data are received and analyzed
4. **Findings Review**: an initial analysis of results is presented to the team, to ensure that tentative findings are consistent with a sound analysis of the data, that important limitations on the study's findings have been identified, and that alternative explanations have been addressed to the greatest extent possible
5. **Reanalysis**: an internal replication of the initial analysis, to ensure that results and conclusions are sound, reliable, and reproducible
6. **Pre-Publication Review**: to ensure OES maintains transparency, retains materials necessary for reproducibility, and meets all legal and administrative requirements in disseminating knowledge for the whole of government and the public

The Current DRAFT OES Process

1. **Project Initiation**: early ideas about the project are discussed to ensure general feasibility, proper planning, and wise investment of team resources prior to formally initiating a project and committing to a collaboration with agency partners
2. **Design Review**: the project design is peer reviewed and then presented to the team, to ensure a sound design that effectively addresses research objectives before we invest resources in fielding a study
3. **Analysis Plan Commitment**: an analysis plan (also known as a "pre-analysis plan" or "pre-specification plan") is finalized, date-stamped, and posted publicly on our website before data are received and analyzed
4. **Findings Review**: an initial analysis of results is presented to the team, to ensure that tentative findings are consistent with a sound analysis of the data, that important limitations on the study's findings have been identified, and that alternative explanations have been addressed to the greatest extent possible
5. **Reanalysis**: an internal replication of the initial analysis, to ensure that results and conclusions are sound, reliable, and reproducible
6. **Pre-Publication Review**: to ensure OES maintains transparency, retains materials necessary for reproducibility, and meets all legal and administrative requirements in disseminating knowledge for the whole of government and the public

1. A proposed research integrity process

2. Standard Operating Procedures

Current DRAFT OES SOP Principles

1. Collaborate. Serve the agency partner. Practice humility and listening.
2. Work in public as much as possible. (Post code, Post pre-analysis plans, Post results)
3. Randomization is a reasoned basis for statistical inference (i.e. p-values should refer to distributions generated by design). (Ramdom sampling, likelihood functions, Bayesian posterior (likelihood+prior) are all reasoned bases as well. But we can much more easily test hypotheses about alternative randomizations than we can justify the other claims.)

Current DRAFT OES SOP Principles

1. Collaborate. Serve the agency partner. Practice humility and listening.
2. Work in public as much as possible. (Post code, Post pre-analysis plans, Post results)
3. Randomization is a reasoned basis for statistical inference (i.e. p-values should refer to distributions generated by design). (Ramdom sampling, likelihood functions, Bayesian posterior (likelihood+prior) are all reasoned bases as well. But we can much more easily test hypotheses about alternative randomizations than we can justify the other claims.)

Current DRAFT OES SOP Principles

1. Collaborate. Serve the agency partner. Practice humility and listening.
2. Work in public as much as possible. (Post code, Post pre-analysis plans, Post results)
3. Randomization is a reasoned basis for statistical inference (i.e. p-values should refer to distributions generated by design). (Ramdom sampling, likelihood functions, Bayesian posterior (likelihood+prior) are all reasoned bases as well. But we can much more easily test hypotheses about alternative randomizations than we can justify the other claims.)

Example 1: Randomization Assessment

Setup an experiment:

```
set.seed(20180225)
N <- 50
K <- 30
Xdat <- as.data.frame(replicate(K,rnorm(N)))
names(Xdat) <- paste0("X",1:K)
y0 <- Xdat$X1+Xdat$X2+rchisq(N,df=1)
y1 <- y0 + rnorm(N,mean=0,sd=sd(y0))## No treatment effect but different variance
Z <- complete_ra(N,m=15) ## very simple randomization
Y <- Z*y1 + (1-Z) * y0 ## randomization reveals a potential outcome
dat <- data.frame(cbind(Xdat,Y=Y,Z=Z,y0=y0,y1=y1))
```

Example 1: Randomization Assessment

Hansen and Bowers (2008) developed an omnibus balance test that refers to a Normal distribution that approximates the randomization-based reference distribution in large-samples. (see the help page for `balanceTest` for `strata()` and `cluster()` arguments for block and/or clustered designs)

```
balfmla <- reformulate(names(Xdat),response="Z")
## See balanceTest help for block and cluster randomized designs
randTest1 <- balanceTest(balfmla,data=dat,report="all",p.adjust.method="none")
signif(randTest1$results[,"p",],3)[1:5]
```

```
    X1     X2     X3     X4     X5
0.1430 0.0924 0.3980 0.5700 0.7510
```

```
randTest1$overall[1,]
```

```
chisquare        df   p.value
  34.0239   30.0000    0.2799
```

```
sum(randTest1$results[,"p",]<.05,na.rm=TRUE) ## How many false positives
```

```
[1] 3
```

Example 1: Randomization Assessment

A common approach with two problems (separation/problems in high dimensions and not-randomization based reference distribution)

```
balfmla <- reformulate(names(Xdat),response="Z")
## See balanceTest help for block and cluster randomized designs
glm1 <- glm(balfmla,data=dat,family=binomial())
```

```
Warning: glm.fit: algorithm did not converge
```

```
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
glm0 <- glm(Z~1,data=dat,family=binomial())
anova(glm0,glm1,test="Chisq")
```

```
Analysis of Deviance Table

Model 1: Z ~ 1
Model 2: Z ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 +
    X12 + X13 + X14 + X15 + X16 + X17 + X18 + X19 + X20 + X21 +
    X22 + X23 + X24 + X25 + X26 + X27 + X28 + X29 + X30
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        49       61.1
2        19        0.0 30     61.1  0.00068 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example 1: Randomization Assessment

Which to use? How should we choose? (One idea: False Positive Rate)

```
newExp <- function(N){ complete_ra(N,m=15) }

getPvalues <- function(){
  dat$newZ <- newExp(N)
  newfmla <- reformulate(names(Xdat),response="newZ")

  bt1 <- balanceTest(newfmla,data=dat,report="all",p.adjust.method="none")
  btp<-bt1$overall[1,"p.value"]

  theglm1 <- glm(newfmla,data=dat,family=binomial())
  theglm0 <- glm(newZ~1,data=dat,family=binomial())
  theanova <- anova(theglm0,theglm1,test="Chisq")
  anovap <- theanova[2,"Pr(>Chi)"]

  return(c(btp = btp, anovap = anovap))
}
```

```
theps <- replicate(1000,getPvalues())
theps[,1:5]
```

```
            [,1]      [,2]      [,3]      [,4]      [,5]
btp    0.1384167 0.2308195 0.5880233 0.2045699 0.1915891
anovap 0.0006787 0.0006787 0.0006787 0.0006787 0.0006787
```

Example 1: Randomization Assessment

```
apply(theps,1,function(x){ mean(x<=.05) })
```

```
  btp anovap
    0     NA
```

Example 2: Pre-registration

We pre-register that we will look at the effect of the treatment moderated by X1:

```
lm1 <- lm(Y~Z*X1,data=dat)
lm1ci <- coefci(lm1,vcov=vcovHC(lm1,type="HC2"))
lm1p <- coeftest(lm1,vcov=vcovHC(lm1,type="HC2"))[4,4]
```

vs we hunt for a statistically significant moderating effect:

```
theres <- sapply(dat[,names(Xdat)],function(thex){
                  thelm<-lm(Y~Z*thex,data=dat)
                  lm1p <- coeftest(thelm,vcov=vcovHC(thelm,type="HC2"))[4,4]
})
anx <- names(theres[theres==min(theres)])
afmla <- as.formula(paste0("Y~Z*",anx))
alm <- lm(afmla,data=dat)
almp <- coeftest(alm,vcov=vcovHC(alm,type="HC2"))[4,4]
```

Example 2: Pre-registration

How often will we make a false positive error in each case?

```
preRegProcedure <- function(newZ){
  lm1 <- lm(Y~newZ*X1,data=dat)
  lm1p <- coeftest(lm1,vcov=vcovHC(lm1,type="HC2"))[4,4]
  return(lm1p)
}
```

```
pHuntProcedure <- function(newZ){
  theres <- sapply(dat[,names(Xdat)],function(thex){
                    thelm<-lm(Y~newZ*thex,data=dat)
                    lm1p <- coeftest(thelm,vcov=vcovHC(thelm,type="HC2"))[4,4]
                    return(lm1p)
})
  anx <- names(theres[theres==min(theres)])
  afmla <- as.formula(paste0("Y~newZ*",anx))
  alm <- lm(afmla,data=dat)
  almp <- coeftest(alm,vcov=vcovHC(alm,type="HC2"))[4,4]
  return(almp)
}

assessPs <- function(newZ){
  preRegP <- preRegProcedure(newZ = newZ)
  pHuntP <- pHuntProcedure(newZ = newZ)
  return(c(preregp=preRegP, phuntp = pHuntP))
}
```

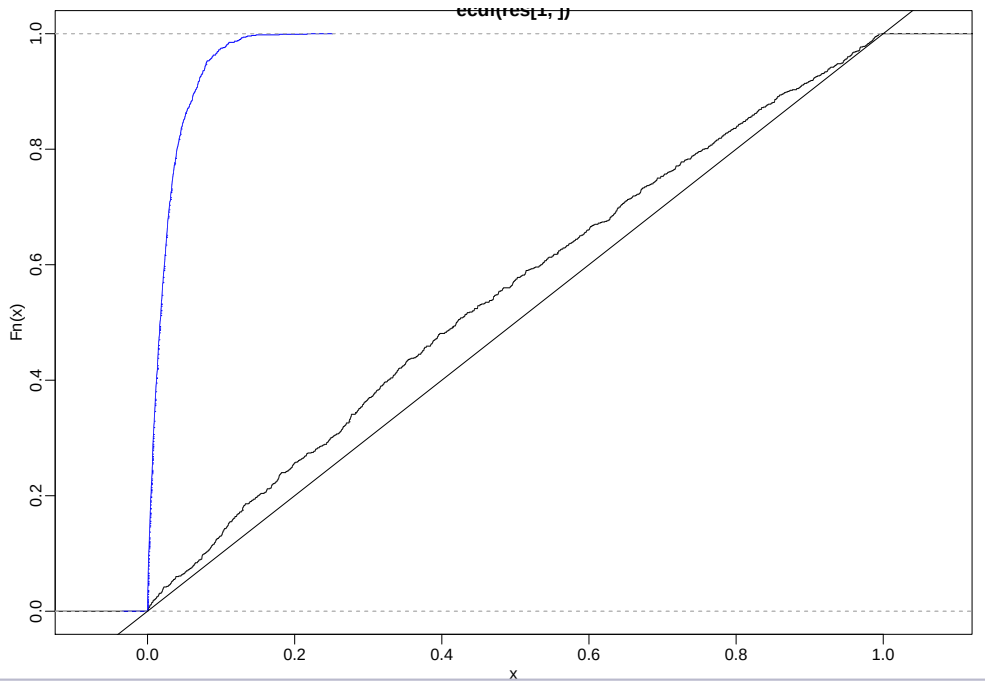
Example 2: Pre-registration

Notice that we could have similar kinds of issues with covariance adjustment and not moderating effects.

```
res <- replicate(1000,assessPs(newZ=complete_ra(N=50,m=15)))
apply(res,1,function(x){ mean(x <=.05) })
```

```
preregp  phuntp
  0.065   0.852
```

```
plot(ecdf(res[1,]))
plot(ecdf(res[2,]),add=TRUE,col="blue")
abline(0,1)
```



Example 3: Covariance Adjustment Avoiding Bias

We often use OLS to estimate the ATE using β_1 (below).

$$Y_i = \beta_0 + \beta_1 Z_i$$

$$\hat{\beta}_1 = Y|\bar{Z} = 1 - Y|\bar{Z} = 0 = \frac{\text{cov(Y,Z)}}{\text{var(Z)}}$$

.

And we know:

$$E_R(\hat{\beta}_1) = \beta_1 \equiv \text{ATE}$$

.

Example 3: Covariance Adjustment Avoiding Bias

Now, what about when we have a covariate X_i and we use it as would be normal in the analysis of non-experimental data:

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 X_i$$

What is β_1 in this case? Well, we all know the matrix representation here $(X^T X)^{-1} X^T y$, but here is the scalar formula for this case:

$$\hat{\beta}_1 = \frac{var(X)cov(Z, Y) - cov(X, Z)cov(X, Y)}{var(Z)var(X) - cov(Z, X)^2}$$

In very large experiments $cov(X, Z) \approx 0$ however in any given finite sized experiment $cov(X, Z) \neq 0$ so this does not reduce to the unbiased estimator of the bivariate case. Is it itself unbiased?

References