Additional Information to the Paper

"Suicide Risk and Sexual Orientation. A Critical Review"

by Plöderl, M., Wagenmakers, E. J., Tremblay, P., Ramsay, R., Kralovec, K., Fartacek, C., Fartacek, R.

Published in the Archives of Sexual Behavior

An in depth analysis of the problems we address in the paper is on this webpage:

http://people.ucalgary.ca/~ramsay/adolescent-gay-lesbian-suicide-risk.htm

CONTENT

Ten Examples of How the Suicide Data has been Recited

These examples demonstrate how the statistically nonsignificant sexual orientation differences in the autopsy studies appear as "no-difference" or "lack of data" in the literature. Several reviews do not discuss potential problems with a "no-difference" interpretation. Not one review discusses the statistical problems addressed in our manuscript.

1.
One psychological autopsy study has examined the association between non-heterosexual sexual orientation and suicide, and **failed to find a difference** between cases and controls (Shaffer, Fisher, Hicks, Parides, & Gould,1995), although a psychological autopsy procedure may not be sensitive enough to detect issues of sexual orientation.

Jeffrey A. Bridge, Tina R. Goldstein, and David A. Brent (2006) Adolescent suicide and suicidal behavior. *Journal of Child Psychology and Psychiatry, 47*, 372–394.

2.
"Each of these studies has concluded that same-sex sexual orientation **is not disproportionately represented** among suicide victims." (p. 16)

Haas, A. P., Eliason, M., Mays, V. M., Mathy, R. M., Cochran, S. D., & D'Augelli, A. R.,…Clayton, P. J. (2011). Suicide and suicide risk in lesbian, gay, bisexual, and transgender populations: review and recommendation. *Journal of Homosexuality, 58*, 10-51.

3.
"However, studies examining linkages between sexual orientation and death by suicide **have not found this association** [127,129,130]. The failure to find linkages with suicide may be due to the fact that ascertainment of sexual orientation following suicide is more difficult than such assessment among those who survive suicide attempts." (p. 424).

Beautrais, A. L. (2000). Risk factors for suicide and attempted suicide among young people. *Australian and New Zealand Journal of Psychiatry, 34,* 420–436.

4.
"Both teams of researchers reporting population-based results concluded that the **rates of suicide are no higher for gay men and lesbians than for heterosexuals."** (p. 85). "Limited data from two community-based studies suggest that GLB people are not at increased risk for suicides compared with the general population" (p. 102)

McDaniel, J. S., Purcell, D., & D'Augelli, A. R. (2001). The relationship between sexual orientation and risk for suicide: Research findings and future directions for research and prevention. *Suicide and Life-Threatening Behavior*, *31*, 84-105

5.
"Completed suicides does not appear to be more common in gay men and lesbians than in heterosexual persons,..." (p. 303).

Catalan, J. (2000). Sexuality, reproductive cycle and suicidal behavior. In K. Hawton & K. van Heeringen (Eds). The international handbook of suicide and attempted suicide (p. 293-307).Chichester: Wiley.

6.

"However, based upon results of (scarce) studies conducted to date, completed suicide rates **do not appear to be increased** among the gay and lesbian populations. (p. 67)

Australian Government Department of Health and Ageing (2008). Research and evidence in suicide prevention. Retrieved May 9, 2011, from
http://www.livingisforeveryone.com.au/DownloadDocument.ashx?DocumentID=61

7.

"Gay, lesbian, or bisexual orientation" ..."Associated with increased rate of suicide attempts, but **no evidence is available on suicide rates per s**e." (p. 26)
"It **remains unclear** whether suicide rates in gay, lesbian, and bisexual individuals differ from the suicide rate among heterosexual individuals" (p. 78)

American Psychiatric Association (2003). Practice Guidelines for the assessment and treatment of patients with suicidal behaviors. Retrieved February 18, from
http://www.psychiatryonline.com/pracGuide/loadGuidelinePdf.aspx?file=SuicidalBehavior_05-15-06.

8.

"In addition, although it has been suggested that gay and lesbian youth may be at higher risk for suicidal behavior than heterosexual youth, research on this issue is complicated by a number of factors, including a **lack of accurate youth suicide rates** specific to the gay and lesbian population." (p. 157).

Miller, D. N., & Eckhert, T. L. (2009). Youth Suicidal Behavior: An Introduction and Overview. *School Psychology Review, 38*, 153–167

9.

"Rates of attempted suicide in homosexual and bisexual men and women are high, but **evidence is lacking for suicide"** (p. 1374).

Hawton, K., & van Heeringen, K. (2009). Suicide. Lancet, 373, 1372-1381

10.

"The difference was not statistically significant" (p. 24)

Gould, M. S., Shaffer, D., & Greenbert, T (2003). The epidemiology and youth suicide. From King, R. A., & Apter, A. (eds.). Suicide and Children in Adolescents (pp1-40). Cambridge: University Press

R-Code for Tian et al. (2009) Meta-Analysis

The code was provided by Lu Tian. The plotting options have been changed by the first author of the manuscript to create Figure 1 and to include the Bayesian results in the Figure. The changes are written in blue.

```
rm(list=ls())

##
=========================================================================
===== ##
##
=========================================================================
===== ##
## === meta.RD is the function for Obtaining the Combined Confidence Interval and P-value
=== ##
## ===                         for the Risk Difference based on exact study-specific intervals
======== ##
##
=========================================================================
==== =##
##=======================================================================
======== ##

poisson.confint=function(x, prob)
        {alpha=1-prob
         n=length(x)
         lower=upper=rep(0, n)
         for(i in 1:n)
           {x0=x[i]
            if(x0==0)
              {f=function(lambda){ppois(0, lambda)-alpha}
               fit=uniroot(f, c(0, 1000))
               upper[i]=fit$root
               }
            if(x0>0)
              {f1=function(lambda){ppois(x0, lambda)-dpois(x0, lambda)*0.5-alpha/2}
               f2=function(lambda){1-ppois(x0, lambda)+dpois(x0, lambda)*0.5-alpha/2}
               fit1=uniroot(f1, c(0, 1000))
               fit2=uniroot(f2, c(0, 1000))
               upper[i]=fit1$root
               lower[i]=fit2$root
               }
            }
        return(list(lower=lower, upper=upper))
        }


diff.pexact=function(x1,x2,e1,e2,delta.grd, n.grd=15, midp=T)
    {
     fit1=poisson.confint(x1,0.9995);
```

```
fit2=poisson.confint(x2,0.9995);

rev=F
if(fit2$upper/e2-fit2$lower/e2<fit1$upper/e1-fit1$lower/e1)
  {fit1=fit2
  e3=e2
  x3=x2

  e2=e1
  x2=x1
  e1=e3
  x1=x3

  rev=T}

l1=fit1$lower
u1=fit1$upper
N1=qpois(0.999999, u1)

lambda1.grd=seq(l1, u1, length=n.grd)/e1;
pnull1.tot=matrix(0, N1+1, n.grd)
for(b in 1:n.grd)
  {lambda1=lambda1.grd[b]
   pnull1.tot[,b]=dpois((0:N1),lambda1*e1)}

fit2=poisson.confint(x2,0.9995);
u2=max(fit2$upper, e2*(u1/e1+max(delta.grd)))
N2=qpois(0.999999, u2)


dfnull=matrix(0, N1+1,N2+1)
sdnull=matrix(0, N1+1,N2+1)
for(i in 0:N1)
  {lambda1=(i+0.5)/(e1+1)
   lambda2=(0:N2+0.5)/(e2+1)
   dfnull[i+1, ]=(0:N2)/e2-i/e1
   sdnull[i+1, ]=sqrt(lambda1/(e1+1)+lambda2/(e2+1))
  }

pv1=pv2=numeric(0)
for(theta in delta.grd)
  {lambda1=(x1+0.5)/(e1+1)
   lambda2=(x2+0.5)/(e2+1)
   t=(x2/e2-x1/e1-theta)/sqrt(lambda1/(e1+1)+lambda2/(e2+1))
   tnull=(dfnull-theta)/sdnull


   pvalue1=pvalue2=rep(0,n.grd); error=1e-6

   if(max(lambda1.grd)+theta<=0)
        { pv1=c(pv1, 0)
```

```
                pv2=c(pv2, 1) }
          if(max(lambda1.grd)+theta>0)
           {
           for(b in 1:n.grd)
             {lambda1=lambda1.grd[b]
              lambda2=lambda1+theta
              if(lambda2>=0)
                {pnull1=pnull1.tot[,b]
                 pnull2=dpois((0:N2), lambda2*e2)
                 n1.adj=N1+1-max(c(1,(1:(N1+1))[cumsum(sort(pnull1))<error]))
                 n2.adj=N2+1-max(c(1,(1:(N2+1))[cumsum(sort(pnull2))<error]))
                 id1=order(pnull1)
                 id2=order(pnull2)
                 id1=(id1[(N1+1):1])[1:n1.adj]
                 id2=(id2[(N2+1):1])[1:n2.adj]
                 pnull=pnull1[id1]%*%t(pnull2[id2])
                 if(midp==T)
                   {pvalue1[b]=sum(pnull[tnull[id1,id2]>t])+sum(pnull[tnull[id1,id2]==t])*0.5
                    pvalue2[b]=sum(pnull[tnull[id1,id2]<t])+sum(pnull[tnull[id1,id2]==t])*0.5
                   }else{
                      pvalue1[b]=sum(pnull[tnull[id1,id2]>t])+sum(pnull[tnull[id1,id2]==t])
                      pvalue2[b]=sum(pnull[tnull[id1,id2]<t])+sum(pnull[tnull[id1,id2]==t])
                      }
                 }

              }
           pv1=c(pv1, max(pvalue1)+(1-0.9995))
           pv2=c(pv2, max(pvalue2)+(1-0.9995))
          }
        }

      if(rev==T)
       {pv3=pv2
        pv2=rev(pv1)
        pv1=rev(pv3)
        }

      return(list(pv1=pv1, pv2=pv2))
    }


meta.pRD=function(data.mi, BB=1000, alpha=seq(0.01,0.99,length=100), B=100000,
cov.prob=0.95, MH.imputation=F)
  { set.seed(111)
   n=length(data.mi[,1])
   e1=data.mi[,3];        e2=data.mi[,4]
   lambda1=data.mi[,1]/data.mi[,3]; lambda2=data.mi[,2]/data.mi[,4]

   if(MH.imputation==T)
    {id=(1:n)[lambda1*lambda2==0]
```

```
    lambda1[id]=(data.mi[id,1]+0.5)/(data.mi[id,3]+1);
lambda2[id]=(data.mi[id,2]+0.5)/(data.mi[id,4]+1)
    e1[id]=data.mi[id,3]+1;              e2[id]=data.mi[id,4]+1
     }

  deltalambda=lambda2-lambda1
  varlambda=lambda2/data.mi[,4]+lambda1/data.mi[,3]
  weight=(e1*e2/(e1+e2))/sum(e1*e2/(e1+e2))
  mu.MH=sum(deltalambda*weight);  sd.MH=sqrt(sum(weight^2*varlambda))
  ci.MH=c(mu.MH-1.96*sd.MH, mu.MH+1.96*sd.MH)
  p.MH=1-pchisq(mu.MH^2/sd.MH^2,1)

  d0=max(abs(ci.MH))
  delta.grd=seq(-d0*5, d0*5,length=BB-1); delta.grd=sort(c(0,delta.grd))



  pv1.pool=pv2.pool=numeric(0)
  for(kk in 1:n)
   { xx1=data.mi[kk,1]
     xx2=data.mi[kk,2]
     ee1=data.mi[kk,3]
     ee2=data.mi[kk,4]
     fit=diff.pexact(xx1, xx2, ee1, ee2, delta.grd)
     pv1.pool=rbind(pv1.pool, fit$pv1); pv2.pool=rbind(pv2.pool, fit$pv2)
     print(kk)
    }


  for(i in 1:n)
   { for(j in 1:BB)
      { pv1.pool[i,(BB-j+1)]=max(pv1.pool[i,1:(BB-
j+1)]);pv2.pool[i,j]=max(pv2.pool[i,j:BB])
      }
    }



  K=length(alpha); sigma0=1/(data.mi[,3]+data.mi[,4])
  weight=matrix(0,K,4); weight[,1]=1; weight[,2]=1/(alpha*(1-alpha)); weight[,3]=1/(1-
alpha)

  t0=matrix(0,B,3); c=matrix(0,K,n)
  for(b in 1:B)
   { y=runif(n)
     for(k in 1:K){c[k,]=(y<alpha[k])-alpha[k]}
     t0[b,1]=sum(t(c*weight[,1])/sigma0); t0[b,2]=sum(t(c*weight[,2])/sigma0)
     t0[b,3]=sum(t(c*weight[,3])/sigma0)
    }

  t0=t0/n; alpha0=(1+cov.prob)/2; cut0=rep(0,3)
  for(b in 1:3){cut0[b]=quantile(t0[,b], 1-alpha0)}
```

```
t0.random=matrix(0,B,3); c=matrix(0,K,n)
for(b in 1:B)
  { y=runif(n)
    for(k in 1:K){c[k,]=0.5*(y<alpha[k])-alpha[k]}
    t0.random[b,1]=sum(t(c*weight[,1])/sigma0); t0[b,2]=sum(t(c*weight[,2])/sigma0)
    t0.random[b,3]=sum(t(c*weight[,3])/sigma0)
  }

t0.random=t0.random/n
cut0.random=rep(0,3)
for(b in 1:3){cut0.random[b]=quantile(t0.random[,b], 1-alpha0)}


t1=t2=matrix(0,BB,3); c1=c2=c

for(b in 1:BB)
  { for(k in 1:K)
      { c1[k,]=(pv1.pool[,b]>=(1-alpha[k]))-alpha[k]
        c2[k,]=(pv2.pool[,b]>=(1-alpha[k]))-alpha[k]
      }
    t1[b,1]=sum(t(c1*weight[,1])/sigma0);   t2[b,1]=sum(t(c2*weight[,1])/sigma0)
    t1[b,2]=sum(t(c1*weight[,2])/sigma0);   t2[b,2]=sum(t(c2*weight[,2])/sigma0)
    t1[b,3]=sum(t(c1*weight[,3])/sigma0);   t2[b,3]=sum(t(c2*weight[,3])/sigma0)
  }

t1=t1/n; t2=t2/n

ci.cons=c(min(delta.grd[t1[,1]>=cut0[1]]),max(delta.grd[t2[,1]>=cut0[1]]))
ci.iv=c(min(delta.grd[t1[,2]>=cut0[2]]),max(delta.grd[t2[,2]>=cut0[2]]))
ci.fisher=c(min(delta.grd[t1[,3]>=cut0[3]]),max(delta.grd[t2[,3]>=cut0[3]]))




n0=max((1:BB)[delta.grd==0])
c1=t1[n0,]; c2=t2[n0,]

p.cons=min(1, 2*min(c(1-mean(t0[,1]>=c1[1]), 1-mean(t0[,1]>=c2[1]))))
p.iv=min(1, 2*min(c(1-mean(t0[,2]>=c1[2]), 1-mean(t0[,2]>=c2[2]))))
p.fisher=min(1, 2*min(c(1-mean(t0[,3]>=c1[3]), 1-mean(t0[,3]>=c2[3]))))

pvalue=c(p.cons, p.iv, p.fisher, p.MH)
names(pvalue)=c("constant", "inverse-variance", "fisher", "asymptotical-MH")
ci=cbind(ci.cons, ci.iv, ci.fisher,ci.MH)
colnames(ci)=c("constant", "inverse-variance", "fisher", "asymptotical-MH")


cir.cons=c(min(delta.grd[t1[,1]>=cut0.random[1]]),max(delta.grd[t2[,1]>=cut0.random[1]]))
  cir.iv=c(min(delta.grd[t1[,2]>=cut0.random[2]]),max(delta.grd[t2[,2]>=cut0.random[2]]))
```

```
cir.fisher=c(min(delta.grd[t1[,3]>=cut0.random[3]]),max(delta.grd[t2[,3]>=cut0.random[3]]))
   cir=cbind(cir.cons, cir.iv, cir.fisher)
   colnames(cir)=c("constant", "inverse-variance", "fisher")




   study.ci=matrix(0, n, 2)
   for(i in 1:n)
     {study.ci[i,]=range(delta.grd[pv1.pool[i, ]>0.025 & pv2.pool[i, ]>0.025])}

   plot(range(delta.grd), c(0, n+5), xlab="risk difference", ylab="study", type="n")
   for(i in 1:n)
     {lines( study.ci[i,], rep(n+5-i+1, 2))}
   lines(ci.cons, rep(4, 2), lwd=2)
   lines(ci.iv, rep(3, 2), lwd=2)
   lines(ci.fisher, rep(2, 2), lwd=2)
   lines(ci.MH, rep(1, 2), lwd=2)
   lines(c(0, 0), c(0, n+6))

   text(0.15, 4, "const")
   text(0.15, 3, "iv")
   text(0.15, 2, "fisher")
   text(0.15, 1, "MH")

   return(list(ci.fixed=ci, ci.random=cir, study.ci=study.ci, accuracy=(max(delta.grd)-
min(delta.grd))/BB))
 }



###############################################################################
###################################################

diff.exact=function(x1,x2,n1,n2,delta.grd, n.grd=15, midp=T)
    {
     fit1=binom.confint(x1,n1,0.9995,method="exact");
     fit2=binom.confint(x2,n2,0.9995,method="exact");

     rev=F
     if(fit2$upper-fit2$lower<fit1$upper-fit1$lower)
       {fit1=fit2
        n3=n2
        x3=x2
        n2=n1
        x2=x1
        n1=n3
        x1=x3

        rev=T}
```

```
l=fit1$lower;
u=fit1$upper


p1.grd=seq(l, u, length=n.grd);
pnull1.tot=matrix(0, n1+1, n.grd)
for(b in 1:n.grd)
  {p1=p1.grd[b]
   pnull1.tot[,b]=dbinom((0:n1),n1,p1)}
dfnull=matrix(0, n1+1,n2+1)
sdnull=matrix(0, n1+1,n2+1)
for(i in 0:n1)
  {p1=(i+0.5)/(n1+1)
   p2=(0:n2+0.5)/(n2+1)
   dfnull[i+1, ]=(0:n2)/n2-i/n1
   sdnull[i+1, ]=sqrt(p1*(1-p1)/n1+p2*(1-p2)/n2)
  }
pv1=pv2=numeric(0)
for(theta in delta.grd)
  {p1=(x1+0.5)/(n1+1)
   p2=(x2+0.5)/(n2+1)
   t=(x2/n2-x1/n1-theta)/sqrt(p1*(1-p1)/n1+p2*(1-p2)/n2)
   tnull=(dfnull-theta)/sdnull
   pvalue1=pvalue2=rep(0,n.grd); error=1e-6
   if(min(p1.grd)+theta>=1)
        { pv1=c(pv1, 1)
          pv2=c(pv2, 0) }
   if(max(p1.grd)+theta<=0)
        { pv1=c(pv1, 0)
          pv2=c(pv2, 1) }
   if(min(p1.grd)+theta<1 && max(p1.grd)+theta>0)
   {
   for(b in 1:n.grd)
     {p1=p1.grd[b]
      p2=p1+theta
      if(p2>=0 && p2<=1)
        {pnull1=pnull1.tot[,b]
         pnull2=dbinom((0:n2), n2, p2)
         n1.adj=n1+1-max(c(1,(1:(n1+1))[cumsum(sort(pnull1))<error]))
         n2.adj=n2+1-max(c(1,(1:(n2+1))[cumsum(sort(pnull2))<error]))
         id1=order(pnull1)
         id2=order(pnull2)
         id1=(id1[(n1+1):1])[1:n1.adj]
         id2=(id2[(n2+1):1])[1:n2.adj]
         pnull=pnull1[id1]%*%t(pnull2[id2])
         if(midp==T)
           { pvalue1[b]=sum(pnull[tnull[id1,id2]>t])+sum(pnull[tnull[id1,id2]==t])*0.5
             pvalue2[b]=sum(pnull[tnull[id1,id2]<t])+sum(pnull[tnull[id1,id2]==t])*0.5
           }else{
```

```
                pvalue1[b]=sum(pnull[tnull[id1,id2]>t])+sum(pnull[tnull[id1,id2]==t])
                pvalue2[b]=sum(pnull[tnull[id1,id2]<t])+sum(pnull[tnull[id1,id2]==t])
              }
          }

        }
      pv1=c(pv1, max(pvalue1)+(1-0.9995)); pv2=c(pv2, max(pvalue2)+(1-0.9995))
      }
    }

   if(rev==T)
    {pv3=pv2
     pv2=rev(pv1)
     pv1=rev(pv3)
     }

   return(list(pv1=pv1, pv2=pv2))
  }



meta.RD=function(data.mi, BB=1000, alpha=seq(0.01,0.99,length=100), B=100000,
cov.prob=0.95, MH.imputation=F)
 { set.seed(111)
   n=length(data.mi[,1])
   n1=data.mi[,3];          n2=data.mi[,4]
   p1=data.mi[,1]/data.mi[,3]; p2=data.mi[,2]/data.mi[,4]

   if(MH.imputation==T)
    {id=(1:n)[p1*p2==0]
     p1[id]=(data.mi[id,1]+0.5)/(data.mi[id,3]+1);
p2[id]=(data.mi[id,2]+0.5)/(data.mi[id,4]+1)
     n1[id]=data.mi[id,3]+1;                n2[id]=data.mi[id,4]+1
     }

   deltap=p2-p1
   varp=p1*(1-p1)/n1+p2*(1-p2)/n2
   weight=(n1*n2/(n1+n2))/sum(n1*n2/(n1+n2))
   mu.MH=sum(deltap*weight); sd.MH=sqrt(sum(weight^2*varp))
   ci.MH=c(mu.MH-1.96*sd.MH, mu.MH+1.96*sd.MH)
   p.MH=1-pchisq(mu.MH^2/sd.MH^2,1)

   d0=max(abs(ci.MH))
   delta.grd=seq(max(-1, -d0*5), min(1, d0*5),length=BB-1); delta.grd=sort(c(0,delta.grd))



   pv1.pool=pv2.pool=numeric(0)
   for(kk in 1:n)
    { x1=data.mi[kk,1]
      x2=data.mi[kk,2]
```

```
    n1=data.mi[kk,3]
    n2=data.mi[kk,4]
    fit=diff.exact(x1,x2,n1,n2, delta.grd)
    pv1.pool=rbind(pv1.pool, fit$pv1); pv2.pool=rbind(pv2.pool, fit$pv2)
    print(kk)
  }


  for(i in 1:n)
   { for(j in 1:BB)
      { pv1.pool[i,(BB-j+1)]=max(pv1.pool[i,1:(BB-
j+1)]);pv2.pool[i,j]=max(pv2.pool[i,j:BB])
      }
    }


   K=length(alpha); sigma0=1/(data.mi[,3]+data.mi[,4])
   weight=matrix(0,K,4); weight[,1]=1; weight[,2]=1/(alpha*(1-alpha)); weight[,3]=1/(1-
alpha)

   t0=matrix(0,B,3); c=matrix(0,K,n)
   for(b in 1:B)
    { y=runif(n)
     for(k in 1:K){c[k,]=(y<alpha[k])-alpha[k]}
     t0[b,1]=sum(t(c*weight[,1])/sigma0); t0[b,2]=sum(t(c*weight[,2])/sigma0)
     t0[b,3]=sum(t(c*weight[,3])/sigma0)
    }

   t0=t0/n; alpha0=(1+cov.prob)/2; cut0=rep(0,3)
   for(b in 1:3){cut0[b]=quantile(t0[,b], 1-alpha0)}


   t0.random=matrix(0,B,3); c=matrix(0,K,n)
   for(b in 1:B)
    { y=runif(n)
     for(k in 1:K){c[k,]=0.5*(y<alpha[k])-alpha[k]}
     t0.random[b,1]=sum(t(c*weight[,1])/sigma0); t0[b,2]=sum(t(c*weight[,2])/sigma0)
     t0.random[b,3]=sum(t(c*weight[,3])/sigma0)
    }

   t0.random=t0.random/n
   cut0.random=rep(0,3)
   for(b in 1:3){cut0.random[b]=quantile(t0.random[,b], 1-alpha0)}


   t1=t2=matrix(0,BB,3); c1=c2=c

   for(b in 1:BB)
    { for(k in 1:K)
      { c1[k,]=(pv1.pool[,b]>=(1-alpha[k]))-alpha[k]
        c2[k,]=(pv2.pool[,b]>=(1-alpha[k]))-alpha[k]
```

```
       }
     t1[b,1]=sum(t(c1*weight[,1])/sigma0);   t2[b,1]=sum(t(c2*weight[,1])/sigma0)
     t1[b,2]=sum(t(c1*weight[,2])/sigma0);   t2[b,2]=sum(t(c2*weight[,2])/sigma0)
     t1[b,3]=sum(t(c1*weight[,3])/sigma0);   t2[b,3]=sum(t(c2*weight[,3])/sigma0)
     }

   t1=t1/n; t2=t2/n

   ci.cons=c(min(delta.grd[t1[,1]>=cut0[1]]),max(delta.grd[t2[,1]>=cut0[1]]))
   ci.iv=c(min(delta.grd[t1[,2]>=cut0[2]]),max(delta.grd[t2[,2]>=cut0[2]]))
   ci.fisher=c(min(delta.grd[t1[,3]>=cut0[3]]),max(delta.grd[t2[,3]>=cut0[3]]))


   n0=max((1:BB)[delta.grd==0])
   c1=t1[n0,]; c2=t2[n0,]

   p.cons=min(1, 2*min(c(1-mean(t0[,1]>=c1[1]), 1-mean(t0[,1]>=c2[1]))))
   p.iv=min(1, 2*min(c(1-mean(t0[,2]>=c1[2]), 1-mean(t0[,2]>=c2[2]))))
   p.fisher=min(1, 2*min(c(1-mean(t0[,3]>=c1[3]), 1-mean(t0[,3]>=c2[3]))))

   pvalue=c(p.cons, p.iv, p.fisher, p.MH)
   names(pvalue)=c("constant", "inverse-variance", "fisher", "asymptotical-MH")
   ci=cbind(ci.cons, ci.iv, ci.fisher,ci.MH)
   colnames(ci)=c("constant", "inverse-variance", "fisher", "asymptotical-MH")


cir.cons=c(min(delta.grd[t1[,1]>=cut0.random[1]]),max(delta.grd[t2[,1]>=cut0.random[1]]))
   cir.iv=c(min(delta.grd[t1[,2]>=cut0.random[2]]),max(delta.grd[t2[,2]>=cut0.random[2]]))

cir.fisher=c(min(delta.grd[t1[,3]>=cut0.random[3]]),max(delta.grd[t2[,3]>=cut0.random[3]]))
   cir=cbind(cir.cons, cir.iv, cir.fisher)
   colnames(cir)=c("constant", "inverse-variance", "fisher")


   study.ci=matrix(0, n, 2)
   for(i in 1:n)
     {study.ci[i,]=range(delta.grd[pv1.pool[i, ]>0.025 & pv2.pool[i, ]>0.025])}

   plot(range(delta.grd), c(0, n+5), xlab="Difference in Proportions", ylab=" ", type="n", xlim
= c(-0.1,0.3), yaxt="n")
   box()
   for(i in 1:n)
     {lines( study.ci[i,], rep(n+5-i+1, 2))}
   lines(ci.cons, rep(4, 2), lwd=2)
   lines(ci.iv, rep(3, 2), lwd=2)
   lines(ci.fisher, rep(2, 2), lwd=2)
   lines(ci.MH, rep(1, 2), lwd=2)
   lines(c(0, 0), c(0, n+6))

   text(.22, 4, "Constant",adj = c(0.5, 0.5))
   text(.22 , 3, "Inverse variance",adj = c(0.5, 0.5))
```

```
   text(.22 , 2, "Fisher",adj = c(0.5, 0.5))
   text(.22 , 1, "Mantel Haenszel",adj = c(0.5, 0.5))

   return(list(ci.fixed=ci, ci.random=cir, study.ci=study.ci, pvalue=pvalue,
accuracy=(max(delta.grd)-min(delta.grd))/BB))
 }
```

```
##############################################################################
###
## Shaffer + Renaud - Data

library("binom")

data=matrix(0,2,4)
data[,2]=c(3,4)
data[,1]=c(0,0)
data[,4]=c(120,55)
data[,3]=c(147,55)
print(data)


##
## ====================================================================
##
## === The 1st column of the data matrix is # events   in group 1  === ##
## === The 2nd column of the data matrix is # events   in group 2  === ##
## === The 3rd column of the data matrix is # subjects in group 1  === ##
## === The 4th column of the data matrix is # subjects in group 2  === ##
##
## ====================================================================
##


fit=meta.RD(data, BB=10000)

##
## ====================================================================
===== ##
## === fit$ci      : lower and upper combined intervals  based on 4 types of weights     === ##
## === fit$pvalue  : p-values for testing H0 : Delta = 0 based on 4 types of weights     === ##
## === fit$accuracy: the grid size used for locating the end of the confidence interval  === ##
## ===              the grid size may be improved by increasing BB                       === ##
##
## ====================================================================
===== ##

print(fit)

text(.25,7, "Shaffer et al. (1995)",adj = c(0.5, 0.5))
text(.25,6, "Renaud et al. (2010)",adj = c(0.5, 0.5))
text(.25,5, "Bayesian combined",adj = c(0.5, 0.5))

# Bayesian Bugs-Result
# Shaffer
lines(c(-0.0036, 0.0601),c(6.8, 6.8), lwd=1, lty = 2)
points(0.0258,6.8,pch=19)
# Renaud
lines(c(-0.0023, 0.1703),c(5.8, 5.8), lwd=1, lty = 2)
```

```
points(0.0767,5.8,pch=19)
# Bayesian
lines(c(0.0141, 0.0763),c(5, 5), lwd=1, lty = 2)
points(0.0418,5,pch=19)
```

Bayesian Analysis in Manuscript – R- and Bugs-Code

R-Code

```
wd <- "..."              # Working Directory
bd <- "../WinBUGS"          # Where is WinBUGS

library(R2WinBUGS)

nnn <- 50000 # simulations


############################################
###### Shaffer
############################################


# # Analytical Bayes factor:
x1 = 3
x2 = 0
n1=120
n2=147

log.BF01 = lchoose(n1,x1) + lchoose(n2,x2) + log(n1+1) + log(n2+1) -
lchoose((n1+n2),(x1+x2)) - log(n1+n2+1)
BF01 = exp(log.BF01)
print(BF01)

data <- list(x1 = 3, x2 = 0, n1=120, n2=147)
inits <- function(){list(pi1 = 0.02, pi2 =0.02)}



samples  <- bugs (data, inits,  model.file="model.bug",
          working.directory= wd,
          parameters =  c("pi1", "pi2", "delta", "deltaprior"),
          n.chains=3, n.iter=nnn, n.burnin=10000, n.thin=1,
          # codaPkg = TRUE,
          bugs.directory= bd)

print(samples,digits=4)
#plot(samples,digits=4)
```

```
################################################################################
######
###### Renaud
data <- list(x1 = 4, x2 = 0, n1=51, n2=55)
inits <- function(){list(pi1 = 0.02, pi2 =0.02)}


samples  <- bugs (data, inits,  model.file="model.bug",
          working.directory= wd,
          parameters =  c("pi1", "pi2", "delta", "deltaprior"),
           n.chains=3, n.iter=nnn, n.burnin=10000, n.thin=1,
          # codaPkg = TRUE,
          bugs.directory=bd)


print(samples,digits=4)




############################################################################
### Renaud + Shaffer as prior
############################################################################

data <- list(x1 = 4+3, x2 = 0+0, n1=117+55, n2=147+55)
inits <- function(){list(pi1 = 0.02, pi2 =0.02)} #prior knowledge


samples   <- bugs (data, inits,  model.file="model.bug",
          working.directory= wd,
          parameters =  c("pi1", "pi2", "delta","deltaprior"),
          n.chains=3, n.iter=nnn, n.burnin=10000, n.thin=1,
          # codaPkg = TRUE,
          # debug=TRUE,
          bugs.directory=bd)


print(samples,digits=4)




############################################################################
### Cochran & Mays, 2011
############################################################################

data <- list(x1 = 0, x2 = 18, n1=85, n2=5292)
inits <- function(){list(pi1 = 0.02, pi2 =0.02)} #prior knowledge
```

```
samples   <- bugs (data, inits,  model.file="model.bug",
              working.directory= wd,
              parameters =  c("pi1", "pi2", "delta","deltaprior"),
              n.chains=3, n.iter=nnn, n.burnin=10000, n.thin=1,
              # codaPkg = TRUE,
              # debug=TRUE,
              bugs.directory=bd)


print(samples,digits=4)
#plot(samples,digits=4)
############################
# Bayes Factor Calculation
# To run after each BUGS-calculation
############################

# H1: delta is unrestricted

# Collect posterior samples across all chains:
delta.posterior  = samples$sims.list$delta
delta.prior      = samples$sims.list$deltaprior


library(polspline) # this package can be installed from within R
fit.prior     = logspline(delta.prior, lbound=-1, ubound=1) # note the bounds.
fit.posterior = logspline(delta.posterior, lbound=-1, ubound=1)

# 95% confidence interval:
x0=qlogspline(0.025,fit.posterior)
x1=qlogspline(0.975,fit.posterior)

posterior    = dlogspline(0, fit.posterior) # this gives the pdf at point delta = 0
prior        = dlogspline(0, fit.prior)     # based on the logspline fit
BF01         = posterior/prior
print(BF01)
# 1/BF01 gives 2.14 -- Exact solution: 2.223484
BF01         = posterior # because we know the height of the prior equals 1 at delta = 0
# 1/BF01 gives 2.17


########################################################
# Order-restriction. H1: delta  > 0


# height of the unrestricted posterior at delta=0:
# Method 1, renormalization
posterior    = dlogspline(0, fit.posterior) # this gives the pdf at point delta = 0
# renormalize:
area         = sum(delta.posterior > 0)/length(delta.posterior)
posterior.OR.M1 = posterior/area
prior.OR     = 2 # we know this exactly
```

```
BF02            = posterior.OR.M1/prior.OR # 0.1717235, BF20 = 5.823314
print (BF02)


# Method 2, estimate truncated posterior
delta.posterior.OR  = delta.posterior[which(delta.posterior>0)]
delta.prior.OR      = delta.prior[which(delta.prior>0)]
fit.posterior.OR    = logspline(delta.posterior.OR, lbound=0, ubound=1)
posterior.OR.M2     = dlogspline(0, fit.posterior.OR) # this gives the pdf at point delta = 0
BF02            = posterior.OR.M2/2 # 0.1679496, BF20 = 5.954168
print (BF02)
```

BUGS-Model-Code

1. With Uninformative Priors (all possible proportions of sexual minority members from 0-100% are equally likely)

```
model {
pi1 ~ dbeta(1,1)  # noninformative beta distrib.
pi2 ~ dbeta(1,1)  # noninformative beta distrib.
x1  ~ dbin(pi1,n1)
x2  ~ dbin(pi2,n2)

delta <- pi1-pi2 #the group-difference of interest
#quot <- pi1/pi2

 # Priors
 # Make "Dummy" Variables That Copy The Prior,
 # But Are Never Updated By Data
 pi1prior ~ dbeta(1,1)
 pi2prior ~ dbeta(1,1)
 deltaprior <- pi1prior-pi2prior
 }
```

2. With restricted, more plausible Priors (assuming a maximum of 50% sexual minority individuals among those who died by suicide and 15% among the living controls)

```
model {
pi1 ~ dbeta(1,1)I(0.0,0.50)  # restricted beta distrib.
pi2 ~ dbeta(1,1)I(0.0,0.15)  # restricted beta distrib.
x1  ~ dbin(pi1,n1)
x2  ~ dbin(pi2,n2)

delta <- pi1-pi2 #the group-difference of interest
#quot <- pi1/pi2

 # Priors
 # Make "Dummy" Variables That Copy The Prior,
 # But Are Never Updated By Data
 pi1prior ~ dbeta(1,1)I(0.0,0.50)
 pi2prior ~ dbeta(1,1)I(0.0,0.15)
 deltaprior <- pi1prior-pi2prior
 }
```
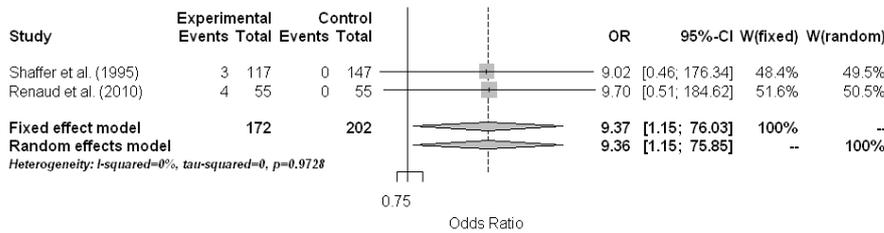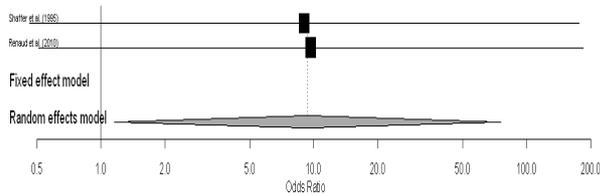
Meta-Analysis with Continuity Correction (Adding 0.5 to all Cell Entries)
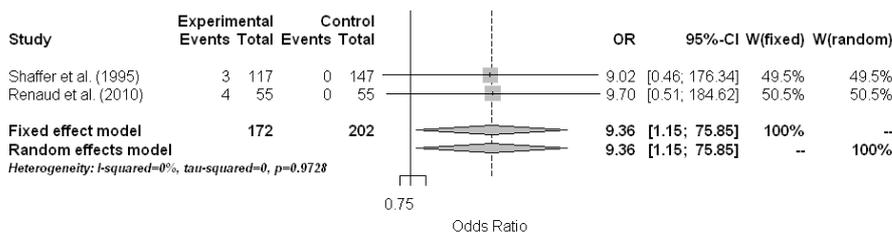
#1. Mantel Haenszel Method

Study	Experimental Events	Total	Control Events	Total		OR	95%-CI	W(fixed)	W(random)
Shaffer et al. (1995)	3	117	0	147		9.02	[0.46; 176.34]	48.4%	49.5%
Renaud et al. (2010)	4	55	0	55		9.70	[0.51; 184.62]	51.6%	50.5%
Fixed effect model		172		202		9.37	[1.15; 76.03]	100%	--
Random effects model						9.36	[1.15; 75.85]	--	100%

Heterogeneity: I-squared=0%, tau-squared=0, p=0.9728

0.75

Odds Ratio

#2. Mantel Haenszel Method – Exact (Only Random Effect Model Possible)

#3. Peto Method

Study	Experimental Events	Total	Control Events	Total		OR	95%-CI	W(fixed)	W(random)
Shaffer et al. (1995)	3	117	0	147		9.72	[0.99; 95.62]	43%	43%
Renaud et al. (2010)	4	55	0	55		7.82	[1.07; 57.06]	57%	57%
Fixed effect model		172		202		8.59	[1.92; 38.48]	100%	--
Random effects model						8.59	[1.92; 38.48]	--	100%

Heterogeneity: I-squared=0%, tau-squared=0, p=0.8883

1

Odds Ratio
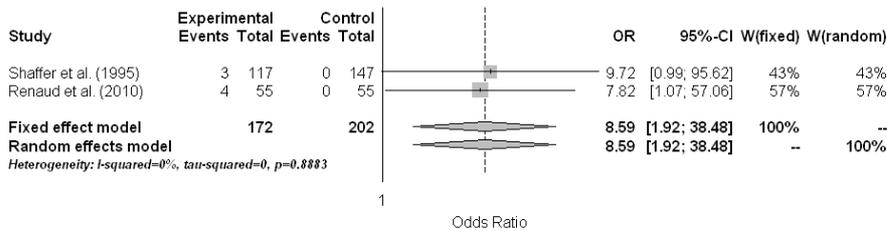
#4. Inverse Variance Method

Study	Experimental Events	Total	Control Events	Total		OR	95%-CI	W(fixed)	W(random)
Shaffer et al. (1995)	3	117	0	147		9.02	[0.46; 176.34]	49.5%	49.5%
Renaud et al. (2010)	4	55	0	55		9.70	[0.51; 184.62]	50.5%	50.5%
Fixed effect model		172		202		9.36	[1.15; 75.85]	100%	--
Random effects model						9.36	[1.15; 75.85]	--	100%

Heterogeneity: I-squared=0%, tau-squared=0, p=0.9728

0.75

Odds Ratio

R-Code for Meta-Analysis

```
library(meta)
x <- read.table("data.meta.txt",sep="\t",header=T)
epsilon <- .5

#1
meta1 <- metabin(homo.suic, n.suic, homo.control, n.control, data=x, allstudies = T, incr =
epsilon, sm="OR", meth="MH",MH.exact = FALSE, studlab=x$Studie)
summary(meta1)

#2
meta1.exact <- metabin(homo.suic, n.suic, homo.control, n.control, data=x, allstudies = T,
incr = epsilon, sm="OR", meth="MH",MH.exact = TRUE,studlab=x$Studie)
summary(meta1.exact)

#3
meta1.p <- metabin(homo.suic, n.suic, homo.control, n.control, data = x, allstudies = T, incr =
epsilon, sm="OR", meth="Peto",studlab=x$Studie)
summary(meta1.p)

#4
meta1.inv <- metabin(homo.suic, n.suic, homo.control, n.control, data = x, allstudies = T, incr
= epsilon, sm="OR", meth="Inverse", studlab=x$Studie)
summary(meta1.inv)
```

Bayesian Results for the Cochran & Mays (2011) study

The gray-shaded distribution is the posterior distribution of the difference suicide rates (MSM vs. heterosexual men). It can be seen that the 95% credible interval of the posterior distribution centers around the zero-proportion difference, but most of the distribution and its mean is on the positive side, thus actually suggesting a nonzero proportion-difference. This is at clearly odds to the interpretation in Cochran and Mathy (2011). The BUGS- and R-Code are given above.
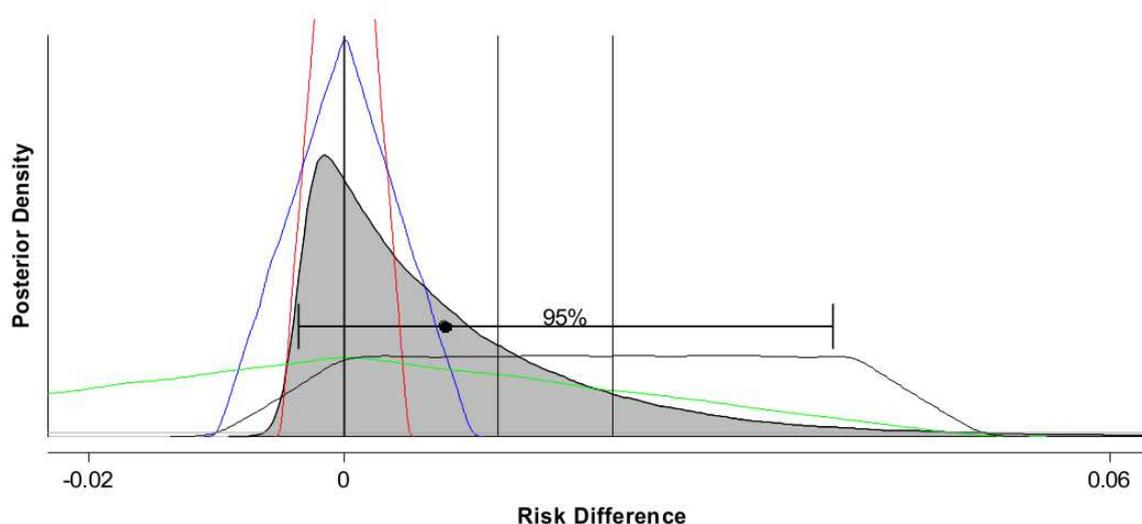
Different Prior Distributions are depicted with different colors in the Figure
Red: 0-0.5% suicides in both groups   BF01 = 0.90
Blue: 0-1% suicides in both groups   BF01 = 1.13
Green: 0-5% suicides in both groups   BF01 = 3.27
Black: 0-5% in homos, 0-1% in heteros   BF01 = 3.27
Not drawn: 0-3% in homo, 0-1 in het.   BF01 = 2.03
Grey: uniformative   BF01 = 66.07

For this data set, the Bayes Factor (BF) is extremely sensitive to the prior-distributions. A noninformative prior distribution is highly implausible, because a suicide rates of 3% in the study period is already a 10-fold increase of risk, given the observed 0.33% among the total sample. The BF is given above.
In contrast to the BF, the credible interval of the posterior distribution is not sensitive to the selection of prior-distributions.

The two black vertical lines at the right of the zero risk difference depict an assumed 5-fold and 8-fold increase of risk, which correspond to 1.2% and 2.1% of risk-difference.

Note: the results are based on a two-independent binomial model. We will replace the analysis with a more appropriate multinomial model, as soon as we find the related BUGS code. However, we expect the resulting credible interval to be only marginally different from the one presented here, and the interpretation of the results will not change.

Note: The information on this page is new, because it was obtained during the review process.

Thanks to the help of Prof. Simon Jackman, an expert on Bayesian Statistics from Stanford University, we could run the calculations with the proper model, namely multinomial sampling. Prof. Jackman provided the R-code. We added the calculation of the rate-differences (delta).

```
M <- 1E6 # Number of resamples
n <- c(0,85,18,5274) # The data
#pi <- MCMCpack::rdirichlet(n=M,alpha=n+rep(.5,4))   #Jeffreys prior (.5,.5,.5,.5)
pi <- MCMCpack::rdirichlet(n=M,alpha=n+rep(1,4))   #Uninformative prior (1,1,1,1)
OR <- (pi[,1]*pi[,4])/(pi[,2]*pi[,3])
LOR <- log(OR)
delta <- pi[,1]/(pi[,1]+pi[,2]) - pi[,3]/(pi[,3]+pi[,4])
plot(density(delta))
summary(delta)
quantile(delta, probs=c(0.025,.5,0.975))
summary(OR)
quantile(OR, probs=c(0.025,.5,0.975))
summary(LOR)
mean(OR>1)
mean(LOR>0)
hist(LOR,
    xlab="Log of the Odds Ratio",
    151,
    main="Histogram, Posterior Density of Log-Odds Ratio,\nDirichlet(1,1,1,1) prior for
multinomial model")
```

Results (Mean and 95%-Credible Interval):

Uninformative Dirichlet prior (1,1,1,1):	M = 0.01, CI = 0.00 – 0.04
Jeffreys Prior (.5, .5, .5, .5):	M = 0.00, CI = 0.00 – 0.03
Binomial Model (see above):	M = 0.01, CI = 0.00 – 0.04

Thus, using the proper model leads to nearly identical results and therefore to the same conclusions than those we had originally made in the manuscript.

An Expanded Homosexuality Factor in Adolescent Suicide

The Shaffer et al. (1995) & Renaud et al. (2010) Psychological Autopsy Studies of adolescents only explored the suicide results for adolescents deemed to be homosexually oriented, but the evidence suggests that, in addition to these adolescents, adolescent not known to be homosexually oriented - but targeted for harassment based on other adolescents' beliefs that they are gay or lesbian (related information was solicited in both studies) - would likely also be at greater risk for dying by suicide. Hence, a new analysis is carried out with the data from the two studies to explore the suicide risk of adolescents who are homosexually oriented combined with those targeted for anti-gay harassment but were not known to be homosexually oriented. Using more advance statistical methods - Bayesian Analyses & Meta-analysis - it is shown that these adolescents - as a group - are most likely at a greater risk for suicide. The analyses are also carried out only for the males in both studies who formed the great majority of suicide deaths for those homosexually oriented combined with those targeted for anti-gay harassment.

An Expanded Homosexuality Factor In Adolescent Suicide
Combining: Those Deemed to be Homosexually Oriented, Plus
Those Harassed/Abused* Because They Were Gender Nonconformable, or
Likely Suspected to be Homosexual and Treated/Abused* Accordingly

Study	Homosexuality As Suicide Risk Factor	Study Information
Shaffer et al. (1995)	Odds Ratio Calculation = Not Possible Fisher Test, One Sided: $p = 0.008$ Fisher-Boschloo: $p = 0.005$ Mid-P: $p = 0.004$	**Males & Females** **(See Above Table for Study Information)** 3 Males - Deemed Homosexual - out of 95 Male Suicide Victims. 0 Females - Deemed Homosexual - out of 25 Female Suicide Victims Total: 3 Homosexual Suicide Victims out of 120 3 Additional Male & 0 Female Suicide Victims Known by Informants to have been 'Teased' for Gender Nonconforming Reasons. 0 Male & 0 Female Controls Known by Informants to have been 'Teased' for Gender Nonconforming Reasons. 0 Male & 0 Female Controls Deemed to be Homosexual. Therefore: 6 Homosexuality-Related Suicide Victims. **For 2 X 2 Statistical Calculations:** 6 Homosexuality-Related Suicide Victims & 114 Non-Homosexuality Related Suicide Victims (6 / 120 = 5.0%) - vs - 0 Homosexuality Related Controls & 147 Non-Homosexuality Related Controls (0 / 147 = 0.0%).
Renaud et al. (2010)	OR = 0.86>3.9<17.1 Fisher Test, One Sided: $p = 0.081$ Fisher-Boschloo: $p = 0.054$ Mid-P: $p = 0.048$	**Males & Females** **(See Above Table for Study Information)** 3 Males - Deemed Homosexual - out of 43 Male Suicide Victims. 1 Female - Deemed Homosexual - out of 12 Female Suicide Victims Total: 4 Homosexual Suicide Victims out of 55 2 Additional Males & 1 Female Suicide Victims Known by Informants to have been 'Teased' for Gender Nonconforming Reasons. 0 Male & 2 Female Controls Known by Informants to have been 'Teased' for Gender Nonconforming Reasons. 3 Males & 1 Female Controls Deemed to be Homosexual. Therefore: 7 Homosexuality-Related Suicide Victims. **For 2 X 2 Statistical Calculations:** 7 Homosexuality-Related Suicide Victims & 48 Non-Homosexuality Related Suicide Victims (7 / 55 = 12.7%) - vs - 2 Homosexuality-Related Controls & 53 Non-Homosexuality Related Controls (2 / 55 = 3.6%).
Combining		Results from Bayesian Analysis, Including Combined Meta-

the Above Shaffer et al. (1995) & Renaud et al. (2010) Study Results		Analysis Results. See Figure 1

Shaffer et al. (1995)	Odds Ratio Calculation = Not Possible

Fisher Test, One Sided: $p = 0.008$ Fisher-Boschloo: $p = 0.005$ Mid-P: $p = 0.004$ | **Males Only** **(See Above Table for Study Information)** 3 Males - Deemed Homosexual - out of 95 Male Suicide Victims.

3 Additional Males Suicide Victims Known by Informants to have been 'Teased' for Gender Nonconforming Reasons. 0 Males Known by Informants to have been 'Teased' for Gender Nonconforming Reasons. 0 Male Controls Deemed to be Homosexual. Therefore: 6 Homosexuality-Related Suicide Victims. **For 2 X 2 Statistical Calculations**: 6 Male Homosexuality-Related Suicide Victims & 89 Male Non-Homosexuality Related Suicide Victims (6 / 95 = 6.3%) - vs - 0 Homosexuality-Related Controls & 116 Non-Homosexuality Related Controls (0 / 116 = 0.0%). |
| Renaud et al. (2010) | Odds Ratio Calculation = Not Possible

Fisher Test, One Sided: $p = 0.028$ Fisher-Boschloo: $p = 0.016$ Mid-P: $p = 0.014$ | **Males Only** **(See Above Table for Study Information)** 3 Males - Deemed Homosexual - out of 43 Male Suicide Victims.

  2 Additional Male Suicide Victims Known by Informants to have been 'Teased' for Gender Nonconforming Reasons. 0 Males Known by Informants to have been 'Teased' for Gender Nonconforming Reasons. 0 Male Controls Deemed to be Homosexual. Therefore: 5 Homosexuality-Related Suicide Victims. **For 2 X 2 Statistical Calculations**: 5 Male Homosexuality-Related Suicide Victims & 38 Male Non-Homosexuality Related Suicide Victims (5 / 43 = 11.6%) - vs - 0 Homosexuality-Related Controls & 43 Non-Homosexuality Related Controls (0 / 43 = 0.0%). |
| Combining the Above Shaffer et al. (1995) & Renaud et al. (2010) Study Results | | Results from Bayesian Analysis, Including Combined Meta-Analysis Results. See Figure 2 |

Statistical Calculations for Fisher Exact Test, One Sided, Carried Out At:
http://statpages.org/ctab2x2.html. At: http://www.stat.ncsu.edu/exact/ - to also produce the Fisher-Boschloo Test (Confidence Interval method). Fisher Mid-P values calculated st: http://www.quantitativeskills.com/sisa/statistics/fisher.htm.
*In both studies, the word "teased" is used for those most likely harassed or abused because adolescents were being gender nonconformable (generally a proxy for being gay or lesbian): feminine or effeminate for boys and masculine for girls. However, a related Canadian study result

was highlighted with: **Gay teens 'terrorized' in Canada's schools** (2009, Study). Harland (2009) reports on Irish study results that would generally apply for boys in the English speaking western world: "Homophobic attitudes are common in contemporary male youth culture in Northern Ireland (Beattie, 2004), as young males believe they are expected to refute any behaviour construed as feminine or that which contravenes traditional masculine stereotypes. In Harland's (2000) study the young men poured scorn on the thought of other males displaying traits they perceived as feminine. This association was always linked directly to derogatory remarks such as "fruits" and "queers". Such comments reveal the extent of repulsion and prejudice that many young men hold towards effeminate behaviour, and why they feel so much pressure to be perceived by others as matching up to the masculine ideal." The expression "teasing" may thus downplay what is more likely to be "abuse" or "violence".

Figure 1: Proportion (Risk) difference for male and female sexual minority members in suicide group and living controls.
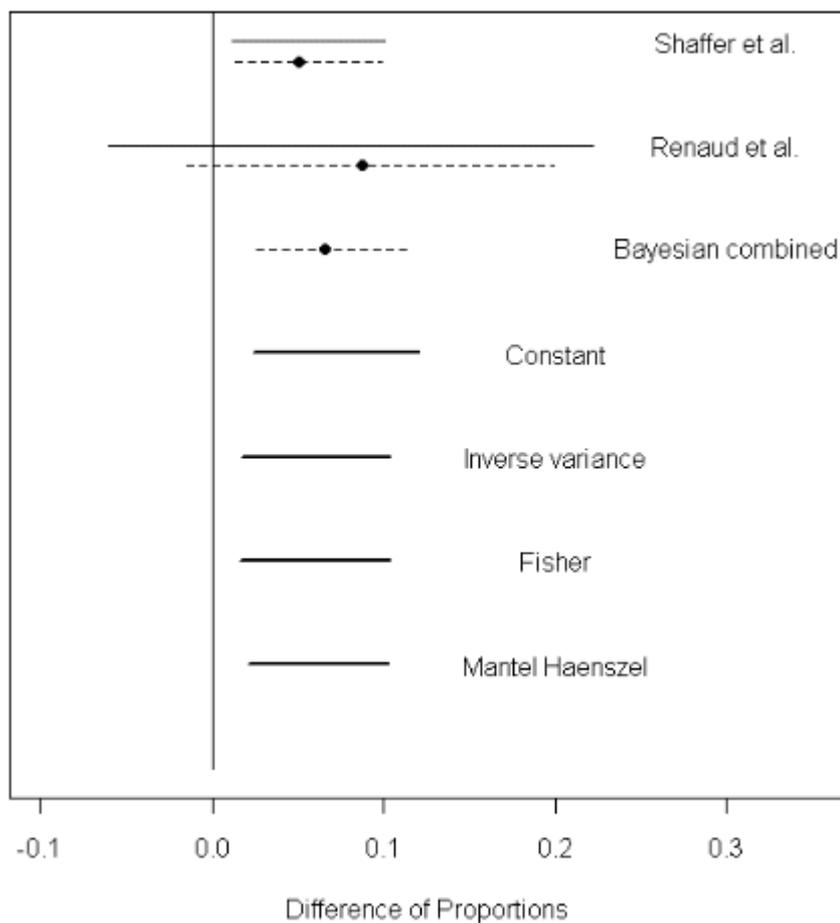
See: Legend & Related Information & Tabled Results.

Figure 2: Proportion (Risk) difference for male sexual minority members in suicide group and living controls.
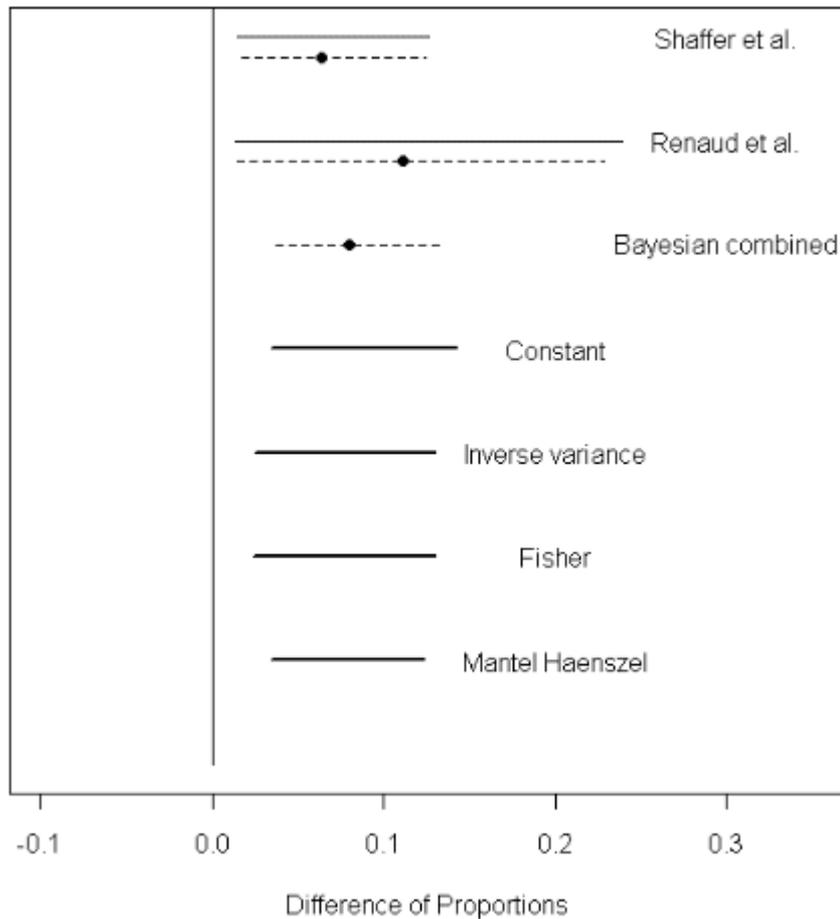
See: Legend & Related Information & Tabled Results.

Legend & Related Information: Solid slim horizontal lines denote the 95% confidence intervals for proportion differences for the two individual studies (Shaffer et al., 1995; Renaud et al., 2010), calculated with the Tian et al. (2009) procedure. Dotted horizontal lines depict the 95% Bayesian credible intervals, with the black dot indicating the mean of the distribution. Solid thick horizontal lines depict the 95% confidence intervals of the meta-analytically combined proportion differences, with four different methods, based on the Tian et al. (2009) procedure. All combined risk differences are statistically significant at $p < 0.05$. For exact confidence intervals, see Print-Out Results for Figure 1 & Figure 2.

Bayesian Methodology. WinBugs Version 1.4.3 (Lunn, Thomas, Best, & Spiegelhalter, 2000) was used in combination with R Version 2.12.1 (R Development Core Team, 2010) and the necessary packages (R2WinBUGS, polspline) for the Bayesian analysis.

Table 1. Metaanalytical Results for Proportion (Risk) difference for male and female sexual minority members in suicide group and living controls. Output from R-calculations with "meta.RD" function.

	Method	Fixed Effect Model		Random Effect Model		p
		Lower-Limit	Upper - Limit	Lower-Limit	Upper - Limit	
Shaffer et al. (1995)	-	0.01	0.10			
Renaud et al. (2010)	-	-0.06	0.22			
Metaanalytic Combination	Constant	0.02	0.12	0.02	0.15	.003
	Inverse-Variance	0.02	0.10	0.06	0.06	-[a]
	Fisher	0.02	0.10	0.00	0.14	.001
	Asymptotical-MH	0.02	0.10	-	-	.003

Note. The distinction between fixed effect and random effect model is obsolete for the CI of the individual study. Sampling was set to BB = 10,000.
[a] problem with *p*-value calculation, will be provided later when the bug is fixed in the meta.RD function.

Table 2. Metaanalytical Results for Proportion (Risk) difference for male sexual minority members in suicide group and living controls. Output from R-calculations with "meta.RD" function.

	Method	Fixed Effect Model		Random Effect Model		p
		Lower-Limit	Upper - Limit	Lower-Limit	Upper - Limit	
Shaffer et al. (1995)	-	0.01	0.13			
Renaud et al. (2010)	-	0.01	0.24			
Metaanalytic Combination	Constant	0.03	0.14	0.02	0.17	.0002
	Inverse-Variance	0.02	0.13	0.07	0.08	-[a]
	Fisher	0.02	0.13	0.01	0.16	.0002
	Asymptotical-MH	0.03	0.12	-	-	.0005

Note. The distinction between fixed effect and random effect model is obsolete for the CI of the individual study. Sampling was set to BB = 10,000.
[a] problem with *p*-value calculation, will be provided later when the bug is fixed in the meta.RD function

Bibliography

Bailey JM, Zucker KJ (1995). Childhood sex-typed behavior and sexual orientation: A conceptual analysis and quantitative review. *Developmental Psychology*, 31(1): 43-55. Abstract.

Bell AP, and Weinberg MS (1978). *Homosexualities: a study of diversity among men & women*. New York: Simon and Schuster.

Bering J (2010). Is your child a "prehomosexual"? Forecasting adult sexual orientation. Scientific American. Full Text.

Gordon AR, Meyer IH (2007). Gender nonconformity as a target of prejudice, discrimination, and violence against LGB individuals. *Journal of LGBT Health Research*, 3(3): 55-71.

Lippa RA (2008). The Relation Between Childhood Gender Nonconformity and Adult Masculinity–Femininity and Anxiety in Heterosexual and Homosexual Men and Women. *Sex Roles*, 59: 684–693. Abstract.

Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337.

Månsdotter A, Lundin A, Falkstedt D, Hemmingsson T (2009). The association between masculinity rank and mortality patterns: a prospective study based on the Swedish 1969 conscript cohort. *Journal of Epidemiology and Community Health*, 63(5): 408-13. Abstract.

Menon M (2011). Does Felt Gender Compatibility Mediate Influences of Self-Perceived Gender Nonconformity on Early Adolescents' Psychosocial Adjustment? *Child Development*. Article first published online. Abstract.

Plöderl M, Fartacek R (2009). Childhood Gender Nonconformity and Harassment as Predictors of Suicidality among Gay, Lesbian, Bisexual, and Heterosexual Austrians. *Archives of Sexual Behavior*, 38: 400–410. PubMed Abstract.

R Development Core Team. (2010). *R: A language and environment for statistical computing [Computer software manual]*. Vienna, Austria. Available from http://www.r-project.org/ (ISBN 3-900051-07-0).

Rees-Turyn AM, Doyle C, Holland A, Root S (2008). Sexism and sexual prejudice (homophobia): The impact of the gender belief system and inversion theory on sexual orientation research and attitudes toward sexual minorities. *Journal of LGBT Issues in Counseling*, 2(1): 2-25. Related Information in: Rees AM, Doyle C, Miesch J (2006). Sexual Orientation, Gender Role Expression, and Stereotyping: The Intersection Between Sexism and Sexual Prejudice (Homophobia). *VISTAS Online*: Full Text.

Remafedi G, Farrow JA, Deisher RW (1991). Risk factors for attempted suicide in gay and bisexual youth. *Pediatrics*, 87(6): 869-75. PubMed Abstract. Tabulated Results.

Renaud J, Berlim MT, Begolli M, McGirr A, Turecki G (2010). Sexual orientation and gender identity in youth suicide victims: an exploratory study. *Canadian Jourrnal of Psychiatry*, 55(1): 29-34. PubMed Abstract. Full Text .

Rieger G, Linsenmeier JA, Gygax L, Bailey JM (2008). Sexual orientation and childhood gender nonconformity: evidence from home videos. *Developmental Psychology*, 44(1): 46-58. Abstract.

Sandfort TG, Melendez RM, Diaz RM (2007). Gender nonconformity, homophobia, and mental distress in Latino gay and bisexual men. *Journal of Sex Research*, 44(2): 181–189.

Shaffer, D., Fisher, P., Parides, M., Gould, M (1995). Sexual orientation in adolescents who commit suicide. *Suicide and Life-Threatening Behavior*, Supplement 25: 64-71.

Skidmore WC, Linsenmeier JAW, Bailey JM (2006). Gender nonconformity and psychological distress in lesbians and gay men. *Archives of Sexual Behavior*, 35(6): 685-697.

Tian L, Cai T, Pfeffer MA, Piankov N, Cremieux PY, Wei LJ (2009). Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2 x 2 tables with all available data but without artificial continuity correction. *Biostatistics*, 10: 275-281.

Tremblay P, Ramsay R (2002). Statistical results generated via SPSS-10 by Thomas G. Albright, Systems Analyst/Programmer at the Kinsey Institute for Research in Sex, Gender and Research. January 9 and 11, 2002. Tabulated Results.

Wade L (2011). If We're Born Gay, How Would We Know? Full Text.

Yunger JL, Carver PR, Perry DG (2004). Does gender identity influence children's psychological well-being? *Developmental Psychology*, 40(4): 572-82. PubMed Abstract.

Considering Different Degrees of Underreporting of Sexual Minority Individuals through Bayesian Analysis

The Screenshot gives the results for different amounts of underreporting among those who died by suicide and living controls. Critical regions where the 95%-credible interval of the proportion difference delta overlaps with zero are marked in yellow, and are depicted graphically on the screenshot. It can be seen that a nonzero proportion difference is observed even with substantial underreporting in either group.

Microsoft Excel - underreporting-results.xls

Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster ?

A16 = 10

	A	B	C	D	E	H
1	Underreporting (%)					
2	Suiciders	Controls	LL	Mean	UL	
3	0	0	0.0126	0.0403	0.0754	
4	0	10				
5	0	20	0.0099	0.0391	0.0759	
6	0	30				
7	0	40	0.0045	0.0369	0.0741	
8	0	50	0.0001	0.0353	0.0731	
9	0	60	-0.0076	0.0328	0.0710	
10	0	70				
11	0	80				
12	0	90				
13	0	100				
14	10	10	0.0138	0.0447	0.0846	
15	10	20				
16	10	30				
17	10	40	0.0076	0.0418	0.0830	
18	10	50	0.0034	0.0404	0.0820	Underreporting
19	10	60	-0.0041	0.0380	0.0809	Controls (%)
20	10	70				100
21	10	80				90 x
22	10	90				80 x x x
23	10	100				70 x x x x
24	20	10	0.0172	0.0511	0.0953	60 x x x x
25	20	20				50 x x
26	20	30				40
27	20	40	0.0113	0.0482	0.0936	30
28	20	50				20
29	20	60	0.0004	0.0441	0.0912	10
30	20	70	-0.0133	0.0399	0.0897	0
31	20	80	-0.0372	0.0323	0.0862	0 10 20 30 40 50 60 70 80 90 100 Underreporting
32	20	90				Suicides (%)
33	20	100				
34	30	10				

	A	B	C	D	E
38	30	50			
39	30	60	0.0050	0.0521	0.1059
40	30	70	-0.0073	0.0480	0.1051
41	30	80			
42	30	90			
43	30	100			
44	40	10			
45	40	20			
46	40	30			
47	40	40			
48	40	50			
49	40	60	0.0111	0.0628	0.1244
50	40	70	0.0000	0.0589	0.1228
51	40	80	-0.0248	0.0510	0.1186
52	40	90			
53	40	100			
54	50	10	0.0327	0.0850	0.1566
55	50	20			
56	50	30			
57	50	40			
58	50	50	0.0255	0.0808	0.1533
59	50	60			
60	50	70	0.0100	0.0743	0.1490
61	50	80	-0.0118	0.0668	0.1454
62	50	90			
63	50	100			
64	60	10			

R- Code for BUGS calculation with underreporting factor

```
# clears workspace:
rm(list=ls(all=TRUE))

# sets working directories:
setwd(wd)

bd <- "C:/Programme/WinBUGS14/"

library(R2WinBUGS)
bugsdir = bd

# Shaffer et al. (1995)
#k_sui = 3
#k_con = 0
#n_sui = 120
#n_con = 147

# Renaud et al. (2010)
#k_sui = 4
#k_con = 0
#n_sui = 55
#n_con = 55

# Combined
k_sui = 7
k_con = 0
n_sui = 175
n_con = 202

# gamma is rate of underreporting
gamma1 = .40
gamma2 = .80
# Woh even with gamma = 0.99 delta is still not overlapping with 0
# But note that the restrictions of 0.15 and 0.5 do come into play there.
# Nevertheless, changing gamma does not seems to lead to different
# inference,at least not on the level of the distribution for delta

# This has to be the case:
# delta = theta_sui_true - theta_con_true =
#        (theta_sui - theta_con) / (1 - gamma)

data  = list("gamma1", "gamma2", "n_sui", "n_con", "k_sui", "k_con") # to be passed on to
WinBUGS

# parameters to be monitored:
parameters = c("delta", "theta_sui_true", "theta_con_true")

# The following command calls WinBUGS with specific options.
# For a detailed description see Sturtz, Ligges, & Gelman (2005).
```

```
samples = bugs(data, inits=NULL, parameters,
                      model.file ="model-underreporting.bug",
                      n.chains=3, n.iter=10000, n.burnin=10, n.thin=1,
                      DIC=T, bugs.directory=bd,
                      codaPkg=F, debug=F)
print(samples,digits=4)
# Now the values for the monitored parameters are in the "samples" object,
# ready for inspection.
```

BUGS-Model

```
# Difference Between Two Rates
# Probability of Being Homosexual
# for Suicide Completers vs. Controls
# With Underreporting Probability gamma

model
{
# Prior on True Rates
  theta_sui_true ~ dbeta(1,1)I(0,0.5)
  theta_con_true ~ dbeta(1,1)I(0,0.15)

# Observed Rates
# gamma is Probability of Underreporting
# (from R code)
  theta_sui <- theta_sui_true * (1-gamma1)
  theta_con <- theta_con_true * (1-gamma2)

# Data Follow a Binomial Distribution:
  k_sui ~ dbin(theta_sui,n_sui)
  k_con ~ dbin(theta_con,n_con)

# Difference between Rates
  delta <- theta_sui_true-theta_con_true
}
```

Results from Unconditional Tests

Problems With The Conditional Fisher Exact Test: The Alternative Use of Unconditional Tests and Mid-*p* Values

By 1990, an important fact related to the conditional Fisher Exact Test was highlighted in the Hirji et al. (1991) abstract: "The use of the Fisher exact test for comparing two independent binomial proportions has spawned an extensive controversy in the statistical literature. Many critics have faulted this test for being highly conservative."

By being "highly conservative," in a situation where the "conditional" Fisher Exact Test has produced a statistical significance of $p = 0.070$, for example, this would mean that the Null Hypothesis is not rejected, also meaning that the difference between the 2 compared groups is not statistically significant, but these conclusions might be in error. That is, if more appropriate statistical tests had been carried out - such as "unconditional exact tests" - or other tests for statistical significance that approximate unconditional test results such as Mid-*p* - most "*p*" results might have been in the "statistically significant" category. That is, the conservatism of the Fisher Exact test would have been removed as noted by Agresti & Gottard (2007). With respect to the Shaffer (1995) and Renaud (2010) studies, the one-sided Fisher Exact Test was used, but Lyderson et al. (2009) emphasized the following in both the abstract and conclusion of their paper:

"The traditional Fisher's exact test should practically never be used." The reason for this is given: "Unconditional tests preserve the significance level and generally are more powerful than [the conditional] Fisher's exact test for moderate to small samples, but previously were disadvantaged by being computationally demanding. This disadvantage is now moot, as software to facilitate unconditional tests has been available for years. Moreover, Fisher's exact test with mid-*p* adjustment gives about the same results as an unconditional test. Consequently, several better tests are available, and the choice of a test should depend only on its merits for the application involved. Unconditional tests and the mid-p approach ought to be used more than they now are." ... Elsewhere in the paper: "We consider exact unconditional tests to be the gold standard for testing association in 2×2 tables."

In other words, in many cases, and especially in studies with small independent binomial samples, the conditional Fisher Exact Test is less powerful and less accurate or less exact for the determination of statistical significance when comparing two independent binomial samples. One concern with using the Mid-*p* value, however, has been that Mid-*p* statistically significant outcomes might produce more Type I errors compared to those occurring with the Fisher Exact Test (determinations of statistical significance when nonsignificance exists). Concerning this issue, researchers have been reporting that the problem is either minimal or that it does not exist (e.g. Hwang & Yang, 2001; Crans & Shuster, 2008; Parzen, 2009; Biddle & Morris, 2011). It is also noted that using Mid-*p* values reduce Type II errors produced by the Fisher Exact Test, which seems to have been the problem when using the Fisher Exact Test with the Shaffer and Renaud study data. Concerning this issue, Biddle and Morris (2011) state:

"In recent years, there has been growing support in the statistical literature regarding an adjustment to the FET [Fisher exact test], namely Lancaster's mid-P (LMP) test (Lancaster, 1961), which yields a better balance of Type I and Type II errors. The LMP test produces Type I error rates that are generally closer to the nominal significance level (typically set at .05) than the FET, which often shows Type I error rates substantially below the nominal level (Crans & Shuster, 2008). Consequently, the LMP test often has greater statistical power (lower Type II error rates) than the FET. The LMP test has been widely accepted in the medical and statistics fields..." (p. 956)

In the "Statistical Significance Tests" table, results from varied unconditional test and Mid-*p* analyses are given for both the Shaffer and Renaud studies. In the great majority of cases (and more so in one-sided tests), results generated from unconditional tests and for Mid-*p* values indicate statistical significance for the difference between the two independent binomial samples in both studies. On this basis, it is therefore concluded that homosexually oriented adolescents are likely to be at greater risk for suicide compared to their heterosexual counterparts. Parzen (2009) reports that Mid-*p* values also closely approximate what would be the result of a Bayesian analysis, this being what was observed given the Bayesian analysis of the two studies, as summarized below.

	Shaffer et al. (1995) n1 = 120, s1 = 3 n2 = 147, s2 = 0		Renaud et al. (2010) n1 = 55, s1 = 4 n2 = 55, s2 = 0	
	One-Tailed	Two-Tailed	One-Tailed	Two-Tailed
Fishers Exact [1,7] (conditional)	.090	.090	.059	.118
Fishers Exact [6] (unconditional)	.073	.145	.042*	.084
Fisher-Boschloo[7]				
No confidence interval method	.073	.085	.042*	.084
With confidence interval Method [8]	.054	.069	.034*	.067
Barnards Test – real *p*-Value[6,10]	.040*	.046*	.021*	.042*
Barnards Test - estimated *p*-Value[6]	.032*	.036*	.021*	.042*
Barnards Test [2]	.030*	.060	.022*	.045*
Barnards Test [9]	.032*	.056	.022*	.045*
z-Pooled statistics[6]	.032*	.056	.022*	.045*
z-Pooled statistics[7]				
No confidence interval method	.032*	.056	.022*	.045*
With confidence interval Method[8]	.033*	.053	.023*	.044*
z-Unpooled statistics[7]				
No confidence interval method	.052	.090	.022*	.045*
With confidence interval Method [8]	.041*	.091	.023*	.044*
Unconditional Arcsine statistic[6]	.032*	.057	.021*	.043*
Mid-p[3]	.045*	.045*	.030*	.059
Mid-p[4]	.045*	.045*	.030*	.089
Mid-p[6]	.050	.049*	.030*	.060
Overall Continuity Corrected[4]	.040*	.040*	.028*	.057
N-1 χ^2 Upton[4,5]	.054		.043*	

Log-Likelihood χ^2 [4]	.028*	.017*
Log-Likelihood χ^2 With Yates Correction[4]	.167	.106

[1]Calculated with function fisher.test in R, similar results with
http://www.quantitativeskills.com/sisa/statistics/fisher.htm
[2]Calculated in R with the function given here: http://www.r-statistics.com/wp-content/uploads/2010/02/Barnard.R.txt
[3]Calculated with http://www.quantitativeskills.com/sisa/statistics/fisher.htm
[4]Calculated with WinPepi 11.17
[5]Only recommended if cell frequencies not less than 1
[6]Calculated with SMP Program http://www.ugr.es/~bioest/software.htm#pt2x2m
[7]Calculated with http://www.stat.ncsu.edu/exact/
[8]Confidence Interval set to 0.999, as recommended.
[9]Calculated with the Barnard package in R, resolution factor for the nuisance parameter set to dp = 0.0001
[10]According to Martín Andrés (pers. comm.), there is no other software solution to run the original Barnard test.

References

Berger, Roger L (1996-2005). *Exact Unconditional Homogeneity/Independence Tests for 2X2 Tables*. http://www.stat.ncsu.edu/exact/ - "This performs exact, unconditional tests of homogeneity (binomial model) or independence (multinomial model) for 2X2 tables. These tests are usually **uniformly more powerful than Fisher's exact test**."
Berger, Roger L (1996). More powerful tests from confidence interval p values. *American Statistician*, 50: 314-318. [Download paper]
Berger, Roger L (1994). *Power comparison of exact unconditional tests for comparing two binomial proportions*. Institute of Statistics Mimeo Series No. 2266. [Download paper]
Berger RL, Boos DD (1994). P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association*, 89: 1012-1016. [Download paper]
"Three test statistics available [also from Berger (1996-2005): http://www.stat.ncsu.edu/exact/]: Fisher's exact as used in Boschloo (1970), Z-pooled or Z-unpooled from Suissa & Shuster (1985)". From: Berger RL (1996). FORTRAN program xun2x2.f Version 2.0 Computes exact unconditional p-values for 2X2 tables [Program used for large sample sizes]. https://www.jiscmail.ac.uk/cgi-bin/filearea.cgi?LMGT1=EXACT-STATS&a=get&f=/xun2x2v2for.txt
Boschloo R. D. (1970). Raised conditional level of significance for the 2X2 table when testing the equality of two probabilities. *Statistica Neerlandica*, 24: 1-35. PDF Download. Download Page.
Suissa S, Shuster JJ (1985). Exact unconditional sample sizes for the 2X2 binomial trial. *Journal of the Royal Statistical Society, Series A*, 148: 317-327. Abstract & Part of Introduction.
For a good explanation of why the *Barnard Exact Test* should be used instead of the *Fisher Exact Test* when comparing results in two small sample, see:
Mehta CR, Senchaudhuri P (2003). *Conditional versus Unconditional Exact Tests for Comparing Two Binomials*. PDF Download. Download Page.

For a Good Summary of Recommended Statistical Significance Tests for 2x2 Tables, see:
Lydersen S, Fagerland MW, Laake P (2009). Recommended tests for association in 2×2 tables. *Statistics in Medicine*, 28(7): 1159-1175. Abstract. PDF Download.

Abstract: The asymptotic Pearson's chi-squared test and Fisher's exact test have long been the most used for testing association in 2X2 tables. Unconditional tests preserve the significance level and generally are more powerful than Fisher's exact test for moderate to small samples, but previously were disadvantaged by being computationally demanding. This disadvantage is now moot, as software to facilitate unconditional tests has been available for years. Moreover, Fisher's exact test with mid-p adjustment gives about the same results as an unconditional test. Consequently, several better tests are available, and the choice of a test should depend only on its merits for the application involved. Unconditional tests and the mid-p approach ought to be used more than they now are. <u>The traditional Fisher's exact test should practically never be used</u> (Emphasis added).

Recommendations: Exact tests have the important property of always preserving test size. Our general recommendation is not to condition on any marginals not fixed by design. In practice, this means that an exact unconditional test is ideal. Pearson's chi-squared (z-pooled) statistic or Fisher–Boschloo's statistic works well with an exact unconditional test. Further, such a test can be approximated by an exact conditional mid-p test or, in large samples, see, for example, the traditional asymptotic Pearson's chi-squared test. However, when an exact test is chosen, an unconditional test is clearly recommended. <u>The traditional Fisher's exact test should practically never be used</u> (Emphasis added).

A very good discussion of unconditional vs. conditional tests that also covers the Fisher-Barnard controversy is from Martín Andrés (1991).
Martin Andres (1991). A review of classic non-asymptotic methods for comparing two proportions by means of independent samples. *Communication in Statistics – Simulation and Computation, 20,* 551-583.
Abtract: The 2 x 2 tables used to present the data in an experiment forcomparing two proportions by means of two observations of two independent binomial distributions may appear simple but are not. The debate about the best method to use is unending, and has divided statisticians into practically irreconcilable groups. In this article, all the available non-asymptotic tests are reviewed (except the Bayesian methodology). The author states which is the optimal (for each group), referring to the tables and programs that exist for them, and contrast the arguments used by supporters of each of the options. They also sort the tangle of solutions into "families", based on the methodology used and/or prior assumptions, and point out the most frequent methodological mistakes committed when comparing the different families.
Software for unconditional test is here: http://www.ugr.es/~bioest/software.htm
According to Martín Andrés, the his SMP software is the only one that realized the original Barnard Test.

Papers related to using Mid-P values
Hirji KF, Tau S-J, Elashoff RM (1991). A Quasi-Exact Test for Comparing Two Binomial Proportions. *Statistics in Medicine*, 10: 1137-1153. - Read Online & Download Page. PDF Download. Abstract: The use of the Fisher exact test for comparing two independent binomial proportions has spawned an extensive controversy in the statistical literature. Many critics have faulted this test for being highly conservative. Partly in response to such criticism, some statisticians have suggested the use of a modified, non-randomized version of this test, namely the mid-P-value test. This paper examines the actual type I error rates of this test. For both one-sided and two-sided tests, and for a wide range of sample sizes, we show that the actual levels of significance of the mid-P-test tend to be closer to the nominal level as compared with various classical tests. The computational effort required for the mid-P-test is no more than that needed for the Fisher exact test. Further, the basis for its modification is a natural adjustment for discreteness; thus the test easily generalizes to r x c contingency tables and other discrete data problems.
Berry G, Armitage P (1995). Mid-P confidence intervals : a brief review. *The Statistician*, 44(4): 417-423. PDF Download. PDF Download. Abstract: Significance tests that are based on discrete probabilities are conservative in that the average value of the significance level,

when the null hypothesis is true, always exceeds 0.5. An approach suggested by H. O. Lancaster over 40 years ago overcomes this problem. This is to calculate the mid-P value, where only half of the probability of the observed sample is included in the tail. The average value of the mid-P value is 0.5 and the variance is slightly less than that of a random variable uniformly distributed between 0 and 1. The mid-P concept has usually been advocated in the context of significance testing but it can be extended to the calculation of confidence intervals in an estimation approach by defining, for example, the 95% mid-P confidence limits as the values that have a one-sided mid-P value of 0.025. In this paper we review recent work supporting this approach.

Hwang JTG, Yang M-C (2001). An optimality theory for mid P-values in 2x2 contingency tables. *Statistica Sinica*, 11: 807-826. PDF Download. Abstract: The contingency table arises in nearly every application of statistics. However, even the basic problem of testing independence is not totally resolved. More than thirty-five years ago, Lancaster (1961) proposed using the mid *p*-value for testing independence in a contingency table. The mid *p*-value is defined as half the conditional probability of the observed statistic plus the conditional probability of more extreme values, given the marginal totals. Recently there seems to be recognition that the mid *p*-value is quite an attractive procedure. It tends to be less conservative than the *p*-value derived from Fisher's exact test. However, the procedure is considered to be somewhat ad-hoc. In this paper we provide theory to justify mid *p*-values. We apply the Neyman-Pearson fundamental lemma and the *estimated truth approach*, to derive optimal procedures, named *expected p-values*. The estimated truth approach views *p*-values as estimators of the truth function which is one or zero depending on whether the null hypothesis holds or not. A decision theory approach is taken to compare the *p*-values using risk functions. In the one-sided case, the expected *p*-value is exactly the mid *p*-value. For the two-sided case, the expected *p*-value is a new procedure that can be constructed numerically. In a contingency table of two independent binomial samplings with balanced sample sizes, the expected *p*-value reduces to a two-sided mid *p*-value. Further, numerical evidence shows that the expected *p*-values lead to tests which have type one error very close to the nomial level. Our theory provides strong support for mid *p*-values.

Agresti A, Gottard A (2007). Nonconservative exact small-sample inference for discrete data. *Computational Statistics & Data Analysis*, 51(12): 6447-6458. Abstract: Exact small-sample methods for discrete data use probability distributions that do not depend on unknown parameters. However, they are conservative inferentially: the actual error probabilities for tests and confidence intervals are bounded above by the nominal level. This article surveys ways of reducing or even eliminating the conservatism. Fuzzy inference is a recent innovation that enables one to achieve the error probability exactly. We present a simple way of conducting fuzzy inference for discrete one-parameter exponential family distributions. In practice, most scientists would find this approach unsuitable yet might be disappointed by the conservatism of ordinary exact methods. Thus, we recommend using exact small-sample distributions but with inferences based on the mid-P value. This approach can be motivated by fuzzy inference, it is less conservative than standard exact methods, yet usually it does well in terms of achieving desired error probabilities. We illustrate for inferences about the binomial parameter.

Parzen, Emanuel (2009). *United Applicable Statistics: Mid-Distribution, Mid-Quantile, Mid P Confidence Intervals Proportion p*. Keynote Paper. Advances in Statistics and Applied Probability: Unified Approaches. A Symposium in Honor of Benjamin N. Kedem. PDF Download. Abstract: We review the seminal influence of Ben Kedem on statistical time series analysis. We motivate our research on United Applicable Statistics ("analogies between analogies") approach to a learning framework for almost all of the Science of Statistics, which we distinguish from the Statistics of Science. We describe the exciting probability and statistical theory of mid-distributions, mid-quantiles, new way to calculate (for data with ties)

sample quantiles and median (mid), asymptotic normality of mid-distributions of Binomial, Poisson, hypergeometric distributions. We advocate statistical inference by mid-PVALUE function of a parameter whose inverse (under a stochastic order condition) is defined to be confidence quantile (of a confidence distribution). We show mid-P frequentist confidence intervals for discrete data have endpoint function equal to confidence quantile, which is algorithmically analogous to Bayesian posterior quantile. One computes frequentist (without assuming prior) but interprets Bayesian. We conclude with 0-1 data and quasi-exact (Beta distribution based) confidence quantiles of parameters p and log-odds(p). We claim quasi-identity of frequentist mid-P confidence intervals and Bayesian posterior credible intervals with uninformative Jeffreys prior. For parameters of standard probability models, calculating confidence quantiles yields Bayesian posterior quantiles for non-informative conjugate priors. "A problem often encountered in the practice of statistics is not that we don't have an answer to your question, but that we have too many answers and don't know which ones to choose as our "final" answer. A problem with an extensive literature, and many competing answers, is inference for parameters of discrete data, such as the true population proportion p when one observes K successes in n trials. What may be novel is our claim that for a proportion p of 0 – 1 data the mid-P frequentist confidence interval is approximately identical with the Bayesian Jeffrey's prior credible interval. An important inference method that is unknown (and perhaps difficult to accept) to many statisticians is the "mid-P" approach (usually credited to Lancaster, 1961). This paper presents theory to justify this frequentist approach and argues that it can be recommended as the "final" (benchmark) answer because it is identical with the Bayesian answer for a Jeffrey's Beta(.5,.5) prior for p. While for large samples other popular answers are approximately numerically equivalent, introductory courses will be happier if we teach only one way, the "right" way, the way that is accurate for small samples and zero successes. It is easy to compute from software for the quantile function of the Beta distribution."

Biddle DA, Morris SB (2011). Using Lancaster's mid-P correction to the Fisher's exact test for adverse impact analyses. *Journal of Applied Psychology*, 96(5): 956-965. Abstract: Adverse impact is often assessed by evaluating whether the success rates for 2 groups on a selection procedure are significantly different. Although various statistical methods have been used to analyze adverse impact data, Fisher's exact test (FET) has been widely adopted, especially when sample sizes are small. In recent years, however, the statistical field has expressed concern regarding the default use of the FET and has proposed several alternative tests. This article reviews Lancaster's mid-P (LMP) test (Lancaster, 1961), an adjustment to the FET that tends to have increased power while maintaining a Type I error rate close to the nominal level. On the basis of Monte Carlo simulation results, the LMP test was found to outperform the FET across a wide range of conditions typical of adverse impact analyses. The LMP test was also found to provide better control over Type I errors than the large-sample Z-test when sample size was very small, but it tended to have slightly lower power than the Z-test under some conditions.

Crans GG, Shuster JJ (2008). How conservative is Fisher's exact test? A quantitative evaluation of the two-sample comparative binomial trial. *Statistics in Medicine*, 27(18): 3598-611. Abstract. See also: Andrés MA, Tejedor HI (2008). Comments on 'How conservative is Fisher's exact test? A quantitative evaluation of the two-sample comparative binomial trial' by G. G. Crans and J. J. Shuster, Statistics in Medicine 2008; 27:3598-3611. *Statistics in Medicine*, 27(18):3598-3611. Reference.

Lancaster HO (1961). Significance tests in discrete distributions. *Journal of the American Statistical Association*, 56: 223–234. Abstract. PDF Download.

Lancaster HO (1949). The combination of probabilities arising from data in discrete distributions. *Biometrika*, 36: 370-382. Reference/Introduction.