Reproducible report: Meta-analysis of relationships between white mold and soybean yield

Created: 2016-10-01 Last update: 2016-11-01

Table of Contents

# The article

This report describes all steps for reproducing the analysis of data from multiple field trial using meta-analytic approaches in a recent article published in Plant Pathology. The work was conducted in the [Epidemiology Lab](http://www.dfp.ufv.br/epidemiology) (Del Ponte Lab) in collaboration with the [EVADE](http://evade.pppmb.cals.cornell.edu/)- Epidemiology of vegetable diseases Lab (Pethybridge Lab):

Lehner, M. S., Pethybridge, S. J., Meyer, M. C., & Del Ponte, E. M. (2016). Meta-analytic modelling of the incidence-yield and incidence-sclerotial production relationships in soybean white mould epidemics. Plant Pathol. <doi:10.1111/ppa.12590>

In the [article](http://onlinelibrary.wiley.com/doi/10.1111/ppa.12590/abstract) ([pre-print here](Lehner-etal-white-mold-PP.pdf)), two relationships were studied: soybean white mold incidence (inc, %) and soybean yield (yld, kg/ha) and incidence and sclerotia weight (scl, g/ha). The data were obtained from a scientific [report](http://ainfo.cnptia.embrapa.br/digital/bitstream/item/101371/1/Ensaios-cooperativos-de-controle-quimico-de-mofo-branco-na-cultura-da-soja-safras-2009-a-2012.pdf) (PT language) on fungicide efficacy evaluated in 35 trials conducted across several locations and 4-year period in Brazil. The data was organized in tables reporting the mean values of the three variables of interest, one table for each trial.

The two relationships were summarized using [meta-analytic](http://apsjournals.apsnet.org/doi/abs/10.1094/PHYTO-03-10-0069) models using three effect-sizes: 1) Fisher's z (from transforming Pearson's r); 2) intercept and 3) slopes of a random-coefficients models fitted to the data. We followed the procedures detailed in previous studies in plant pathology (Madden and Paul, 2009; Dalla Lana et al., 2015). For summarizing the correlation coefficients, we calculated Fisher's z and fitted random-effects and mixed-effects model; the latter for testing the effect of moderator variables. For summarizing the slopes and intercepts, we fitted multi-level models (random-coefficient) as described in Madden and Paul (2009) who used SAS. Here, we demonstrate the analysis in R using the lmer package as described in this metafor [tutorial](http://www.metafor-project.org/doku.php/tips:two_stage_analysis). Similar to the tutorial, the estimates were similar between the two-stage and the multilevel approach, and so we decided to use the latter.

# This report

My goal is to demonstrate, using [R](https://www.r-project.org/) programming, each step of the analyses, from data preparation to presentation of results, which can be reproduced either by myself or other people interested in the same topic. Instead of giving away all the data and my (messy!) original code, I thought that investing time to prepare this report can contribute to popularize the use of meta-analytic approaches in plant pathology. The data, codes, pre-print version of the article and supplementary materials were also made available at this GitHub [repository](https://github.com/emdelponte/paper-white-mold-meta-analysis).

The report is the html output of an [R Markdown](http://rmarkdown.rstudio.com) file prepared with the [R Studio](https://www.rstudio.com/) IDE for R. As much as possible I used the pipe operator %>% of the magrittr [package](https://cran.r-project.org/web/packages/magrittr/vignettes/magrittr.html), which really makes the code easier to write and understand. The plots were prepared using both the base R graphics and [ggplot2](http://ggplot2.org/), whichever was more convenient. Most of them are simple versions for a quick visualization, not actually formatted for final publishing.

# Data import

First, we need to load the packages.

library(tidyverse); library(broom); library(tidyr);  
library(cowplot); library(tibble); library(knitr); library(nlme); library(lme4)

The raw data was organized in the long or [tidy format](http://garrettgman.github.io/tidying/) where each treatment (observation) in a fungicide trial (hereafter study) is placed in its row and each variable in its column. Let's now load the data and group by study for separated data frames, one for each relationship.

# inc-yld  
dat\_yld <- read\_csv("dat-white-mold-br.csv") %>%   
 group\_by(study)  
# inc-scl  
dat\_scl <- read\_csv("dat-white-mold-br.csv") %>%  
na.omit(dat\_scl)%>% # some trials did not have sclerotia data, and we omit them here  
 group\_by(study)

See the structure of the data set and all variables types (scroll the content to the left). The full data set in csv format can be downloaded here: <dat-white-mold-br.csv>

dat\_yld

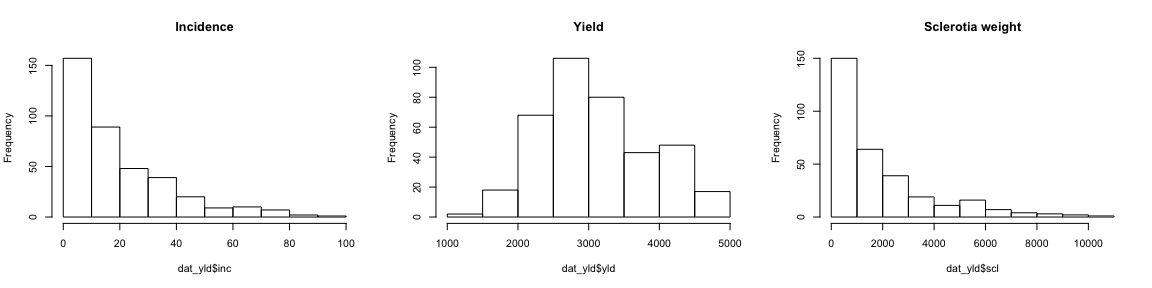
## Source: local data frame [382 x 17]  
## Groups: study [35]  
##   
## study treat season harvest\_year location state country elevation region elevation\_class inc\_check inc\_class yld\_check yld\_class inc scl yld  
## <int> <int> <chr> <int> <chr> <chr> <chr> <int> <chr> <chr> <dbl> <chr> <dbl> <chr> <dbl> <int> <dbl>  
## 1 9 1 2009/2010 2010 Agua Fria GO Brazil 891 Northern low 37.7 low 3729 high 37.7 5092 3729  
## 2 9 2 2009/2010 2010 Agua Fria GO Brazil 891 Northern low 37.7 low 3729 high 11.6 6154 3739  
## 3 9 3 2009/2010 2010 Agua Fria GO Brazil 891 Northern low 37.7 low 3729 high 33.5 200 3863  
## 4 9 4 2009/2010 2010 Agua Fria GO Brazil 891 Northern low 37.7 low 3729 high 1.0 180 3904  
## 5 9 5 2009/2010 2010 Agua Fria GO Brazil 891 Northern low 37.7 low 3729 high 5.6 1123 4471  
## 6 9 6 2009/2010 2010 Agua Fria GO Brazil 891 Northern low 37.7 low 3729 high 1.0 641 4313  
## 7 9 7 2009/2010 2010 Agua Fria GO Brazil 891 Northern low 37.7 low 3729 high 3.7 1203 4177  
## 8 9 8 2009/2010 2010 Agua Fria GO Brazil 891 Northern low 37.7 low 3729 high 0.0 521 4001  
## 9 9 9 2009/2010 2010 Agua Fria GO Brazil 891 Northern low 37.7 low 3729 high 1.1 20 4090  
## 10 9 10 2009/2010 2010 Agua Fria GO Brazil 891 Northern low 37.7 low 3729 high 0.0 0 4254  
## # ... with 372 more rows

# Data visualization

## Raw data

The histograms below summarize the distribution of the three variables used to calculate the three effect-sizes for the meta-analysis. We can see that the two disease-related variables (inc and scl) are skewed to the left and yield data are more normally distributed.

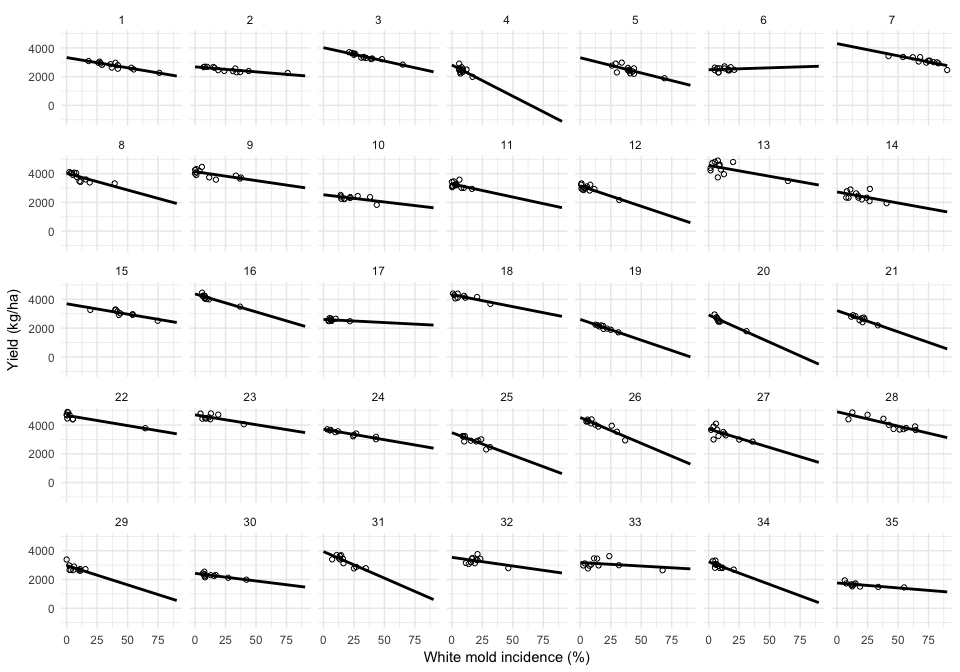
par(mfrow= c(1,3))  
hist(dat\_yld$inc, main = "Incidence")  
hist(dat\_yld$yld, main = "Yield")  
hist(dat\_yld$scl, main = "Sclerotia weight")



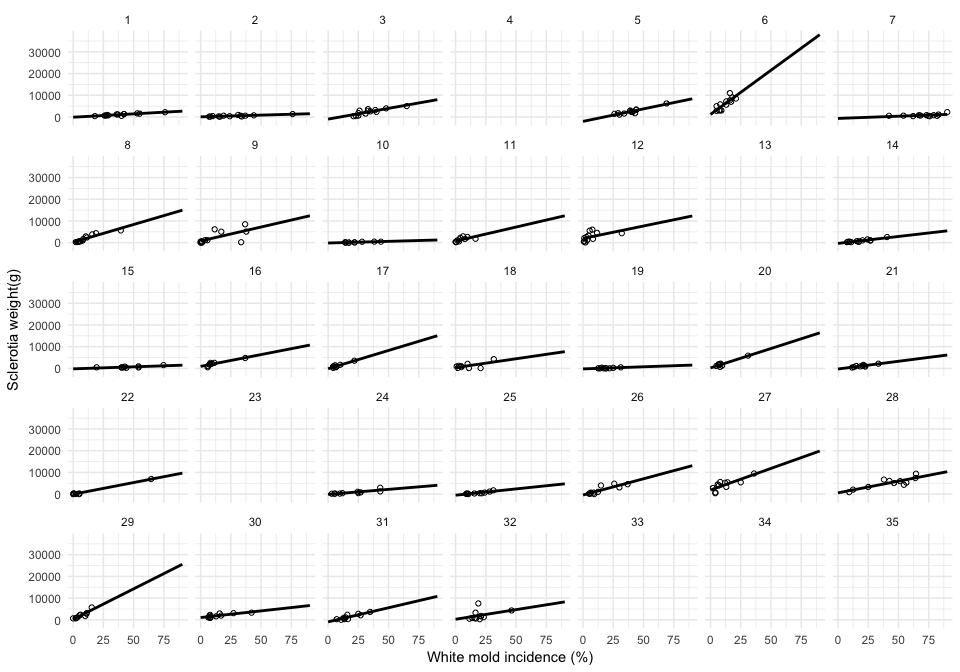
## Individual regressions

Let's now visualize the two relationships of interest, conditioned to study, and add a regression line (extended to the full range) for each study with variable number of pair of observation. Note that soybean yield decreases with the increase of white mold incidence. On the other hand, sclerotia weight increases with the increase of disease incidence. The variability in the incidence in a single study was due to differences in fungicide efficacy. Note that 35 and 29 studies were available for the inc-yld and inc-scl, respectively (6 plots are empty for inc-scl because data was not available).

library(ggplot2)  
library(ggthemes)  
ggplot(dat\_yld, aes(inc, yld))+  
 geom\_point(shape = 1)+  
 stat\_smooth(method = lm, fullrange=TRUE, se = F, col = "black")+  
 ylab("Yield (kg/ha)")+  
 xlab("White mold incidence (%)")+  
 theme\_minimal()+  
 facet\_wrap(~ study, ncol = 7, scales = "fixed")

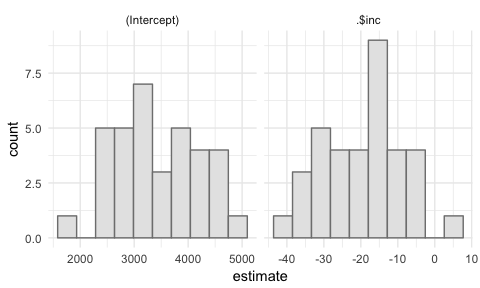


ggplot(dat\_yld, aes(inc, scl))+  
 geom\_point(shape = 1)+  
 stat\_smooth(method = lm, fullrange=TRUE, se = F, col = "black")+  
 ylab("Sclerotia weight(g)")+  
 xlab("White mold incidence (%)")+  
 theme\_minimal()+  
 facet\_wrap(~ study, ncol = 7, scales = "fixed")



#### Coefficients

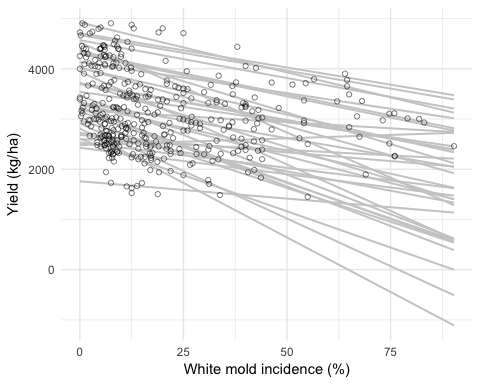
dat\_yld %>%  
 do(tidy(lm(.$yld ~ .$inc), conf.int=TRUE)) %>%   
 ggplot(aes(estimate))+  
 geom\_histogram(bins = 10, fill = "grey90", color="grey50")+  
 theme\_minimal()+  
 facet\_wrap(~term, scales = "free\_x")



#### Fitted lines

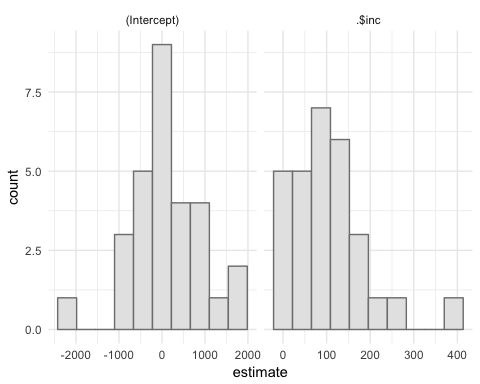
Let's see all the regression lines in the same plot.

ggplot(dat\_yld, aes(inc, yld))+  
 geom\_smooth(method="lm", fullrange=TRUE, se=F, size=0.7, color="grey80", aes(group = factor(study)))+  
 geom\_point(alpha = 0.5, shape = 1)+  
 ylab("Yield (kg/ha)")+  
 xlab("White mold incidence (%)")+  
 theme\_minimal()



### Coefficients

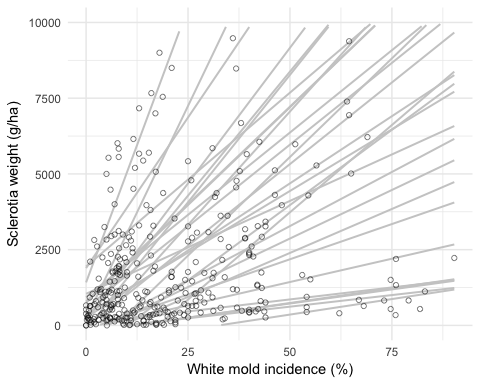
dat\_scl %>%   
 do(tidy(lm(.$scl ~ .$inc), conf.int=TRUE)) %>%  
 ggplot(aes(estimate))+  
 geom\_histogram(bins = 10, fill = "grey90", color="grey50")+  
 theme\_minimal()+  
 facet\_wrap(~term, scales = "free\_x")



#### Fitted lines

Regression plot.

ggplot(dat\_scl, aes(inc, scl))+  
 geom\_smooth(method="lm", fullrange=TRUE, se=F, size=0.7, color="grey80", aes(group = factor(study)))+  
 geom\_point(alpha = 0.5, shape = 1)+  
 ylab("Sclerotia weight (g/ha)")+  
 xlab("White mold incidence (%)")+  
 ylim(0,10000)+  
 theme\_minimal()



# Meta-analytic models

The procedures described below are exactly the same for the two studied relationships. For each one, we firstly summarize the Fisher's z as the effect size for the study of the strength of the association between white mold incidence and soybean yield. Secondly, we will fit a regression model to individual studies. Finally, we will fit random-coefficients model to estimate the coefficients.

## Correlation coefficient

### Incidence-yield

Here I use the dplyr::do and tidyr::broom functions to extract the correlation statistics from each of the studies and assign them to the cor\_yld\_inc data frame.

cor\_yld\_inc <- dat\_yld %>%   
 do(tidy(cor.test(.$inc, .$yld)))

Let's extract the first row of each study from the dat\_yld data frame and then combine with the new cor\_yld\_inc data frame that contains the correlation statistics. We will add a new column (n) for the number of data points per study using the mutate function.

dat\_yld2 <- filter(dat\_yld, row\_number() == 1)  
dat\_yld3 <- full\_join(cor\_yld\_inc, dat\_yld2, by = "study") %>%   
 mutate(n = parameter + 2)

The Fisher's z was used as effect-size because of its better statistical property than the Pearson's r. We obtain the Fisher's z and sampling variance of each study with the escalc function of the metafor package that calculates and adds them to the data frame. Note that the effect-size and sampling variance are indicated by yi and vi, the standard notations used in metafor when using the escalc function. Let's see how the data frame looks like.

library(metafor)  
dat\_yld3 <- escalc(measure = "ZCOR", ri = estimate, ni = n, data = dat\_yld3)  
  
head(dat\_yld3)

## study estimate statistic p.value parameter conf.low conf.high method alternative treat season harvest\_year location state country  
## 1 1 -0.8999278 -6.8450829 2.782235e-05 11 -0.9699609 -0.69213607 Pearson's product-moment correlation two.sided 1 2008/2009 2009 Montividiu GO Brazil  
## 2 2 -0.8149448 -4.6638241 6.894202e-04 11 -0.9426562 -0.47907499 Pearson's product-moment correlation two.sided 1 2008/2009 2009 Sao Miguel do Passa Quatro GO Brazil  
## 3 3 -0.9566920 -10.8999169 3.105025e-07 11 -0.9872663 -0.85795439 Pearson's product-moment correlation two.sided 1 2008/2009 2009 Silvania GO Brazil  
## 4 4 -0.6139396 -2.5795901 2.560901e-02 11 -0.8704696 -0.09513615 Pearson's product-moment correlation two.sided 1 2008/2009 2009 Ponta Grossa PR Brazil  
## 5 5 -0.7467931 -3.7242443 3.356899e-03 11 -0.9194503 -0.33270774 Pearson's product-moment correlation two.sided 1 2008/2009 2009 Maua da Serra PR Brazil  
## 6 6 0.1125625 0.3757154 7.142732e-01 11 -0.4674118 0.62479767 Pearson's product-moment correlation two.sided 1 2008/2009 2009 Nova Ponte MG Brazil  
## elevation region elevation\_class inc\_check inc\_class yld\_check yld\_class inc scl yld n yi vi  
## 1 921 Northern low 76 high 2265 low 76 2194 2265 13 -1.4718 0.1000  
## 2 1031 Northern high 76 high 2257 low 76 1331 2257 13 -1.1416 0.1000  
## 3 1050 Northern high 65 high 2839 high 65 5013 2839 13 -1.9053 0.1000  
## 4 1021 Southern high 17 low 1986 low 17 NA 1986 13 -0.7152 0.1000  
## 5 1029 Southern high 69 high 1893 low 69 6216 1893 13 -0.9657 0.1000  
## 6 999 Northern high 21 low 2476 low 21 8500 2476 13 0.1130 0.1000

#### Overall Fisher's z

A [random-effects](http://www.metafor-project.org/doku.php/tips:rma.uni_vs_rma.mv) meta-analytic model was fitted to these data using a maximum likelihood estimator for the amount of heterogeneity.

ma\_cor\_yld <- rma.uni(yi, vi, method = "ML", data = dat\_yld3)  
summary(ma\_cor\_yld)

##   
## Random-Effects Model (k = 35; tau^2 estimator: ML)  
##   
## logLik deviance AIC BIC AICc   
## -23.1897 53.8742 50.3794 53.4901 50.7544   
##   
## tau^2 (estimated amount of total heterogeneity): 0.0953 (SE = 0.0531)  
## tau (square root of estimated tau^2 value): 0.3087  
## I^2 (total heterogeneity / total variability): 42.97%  
## H^2 (total variability / sampling variability): 1.75  
##   
## Test for Heterogeneity:   
## Q(df = 34) = 61.1764, p-val = 0.0029  
##   
## Model Results:  
##   
## estimate se zval pval ci.lb ci.ub   
## -1.0111 0.0799 -12.6553 <.0001 -1.1677 -0.8545 \*\*\*   
##   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Back-transform z to obtain overall mean r.

pred\_r <- predict(ma\_cor\_yld, transf = transf.ztor)  
pred\_r

## pred ci.lb ci.ub cr.lb cr.ub  
## -0.7662 -0.8235 -0.6934 -0.9269 -0.3681

#### Effect of moderators

The random-effects model fitted previously assumes that the heterogeneity in the true correlation coefficients (Fisher's z) is purely random. However, there may be differences among the individual effects that are (at least in part) related to study-specific variables. These variable can be treated as "moderators" in the model. We considered here: year, region, elevation, or disease and yield levels in the untreated check, as indicative of the disease pressure.

The mixed-effect models test one moderator variable, each at a time, as a fixed effects. The goal was to examine the extent that the moderators included in the model influence the size of the average true effect. The heterogeneity among the true effect-sizes is evaluated based on significance of the Cochran Q test and the *I*2 index that measures the extent of heterogeneity of the true effect-sizes.

# season  
ma\_cor\_yld\_year <- rma(yi, vi, sei="", mods = ~season, method = "ML", data = dat\_yld3)  
  
## region  
ma\_cor\_yld\_region <- rma(yi, vi, mods = ~region, method = "ML",data = dat\_yld3)  
  
## elevation as continuous  
ma\_cor\_yld\_elevation <- rma(yi, vi, mods = ~elevation, method = "ML", data = dat\_yld3)  
  
## elevation as category  
ma\_cor\_yld\_elevation2 <- rma(yi, vi, mods = ~elevation\_class, method = "ML", data = dat\_yld3)  
  
## incidence in the check as continuous  
ma\_cor\_yld\_inc <- rma(yi, vi, mods = ~inc\_check, method = "ML", data = dat\_yld3)  
  
## incidence in the check as categorical  
ma\_cor\_yld\_inc2 <- rma(yi, vi, mods = ~inc\_class, method = "ML", data = dat\_yld3)  
  
## yield in the check as continuous  
ma\_cor\_yld\_yld <- rma(yi, vi, mods = ~yld\_check, method = "ML", data = dat\_yld3)  
  
## yield in the check as categorical  
ma\_cor\_yld\_yld2 <- rma(yi, vi, mods = ~yld\_class, method = "ML", data = dat\_yld3)

Let's check which variables significantly affected the heterogeneity and how much the variance was accounted for when including the moderator variable. It seems we have only one variable with a marginally significant P-value.

table\_yld <- frame\_data(  
~"Moderator", ~"Test of moderator", ~"R2",  
"Season", ma\_cor\_yld\_year$QMp, ma\_cor\_yld\_year$R2,  
"Region", ma\_cor\_yld\_region$QMp, ma\_cor\_yld\_region$R2,  
"Elevation categorical", ma\_cor\_yld\_elevation$QMp, ma\_cor\_yld\_elevation$R2,  
"Elevation continuous", ma\_cor\_yld\_elevation2$QMp, ma\_cor\_yld\_elevation2$R2,  
"Incidence categorical", ma\_cor\_yld\_inc$QMp, ma\_cor\_yld\_inc$R,  
"Yield continous", ma\_cor\_yld\_yld$QMp, ma\_cor\_yld\_yld$R2,  
"Yield categorical", ma\_cor\_yld\_yld2$QMp, ma\_cor\_yld\_yld2$R2  
)  
  
kable(table\_yld, format = "pandoc", caption = "P-value for the test for the effect of moderator and the amount of variance (heterogeneity) accounted for by the moderator (R2 statistics)")

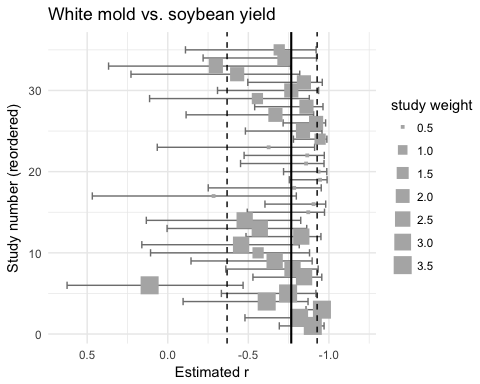
P-value for the test for the effect of moderator and the amount of variance (heterogeneity) accounted for by the moderator (R2 statistics)

|  |  |  |
| --- | --- | --- |
| Moderator | Test of moderator | R2 |
| Season | 0.4612331 | 15.2311624 |
| Region | 0.3253424 | 6.4018736 |
| Elevation categorical | 0.2620535 | 7.8577561 |
| Elevation continuous | 0.8234688 | 0.2695911 |
| Incidence categorical | 0.0604829 | 24.1929491 |
| Yield continous | 0.9372820 | 0.0985151 |
| Yield categorical | 0.8651536 | 0.3714379 |

#### Forest plot

Let's visualize the Pearson's r by study and the mean r (solid line) and confidence intervals (dashed lines) from back-transforming Z estimated by the random-effects model. This plot was not included in the paper, but is interesting to check the heterogeneity of the correlation coefficients and the estimated mean r from back-transforming Z (solid line).

wi <- 1/sqrt(dat\_yld3$vi)  
size <- 0.5 + 3.0 \* (wi - min(wi))/(max(wi) - min(wi))  
library(ggplot2)  
dat\_yld3 %>%   
 ggplot(aes(x = study, y = estimate)) +   
 geom\_errorbar(aes(ymin = conf.low, ymax = conf.high), color="grey50") +   
 geom\_point(aes(size = size), shape = 15, color="grey70") +  
 geom\_hline(yintercept = pred\_r$pred, size=0.75)+  
 geom\_hline(yintercept = c(pred\_r$cr.lb, pred\_r$cr.ub), linetype="dashed")+   
 coord\_flip()+  
 scale\_y\_reverse(limits = c(0.65,-1.2))+  
 labs(x = "Slope") +   
 theme\_minimal() +   
 labs(size = "study weight",   
 title = "White mold vs. soybean yield",   
 y = "Estimated r", x = "Study number (reordered)")



Follow the same procedures as described above.

cor\_scl\_inc <- dat\_yld %>%   
 na.omit(dat\_yld)%>%  
 do(tidy(cor.test(.$inc, .$scl)))

Join the two data frames

dat\_scl2 <- filter(dat\_yld, row\_number() == 1) %>%   
 na.omit(dat\_scl)  
dat\_scl3 <- full\_join(cor\_scl\_inc, dat\_scl2,   
 by = "study") %>%   
 mutate(n = parameter + 2)   
# create variable with the number of data points in the correlation

Obtain Fisher's z.

library(metafor)  
dat\_scl3 <- escalc(measure = "ZCOR", ri = estimate, ni = n, data = dat\_scl3)

### Incidence-sclerotia

#### Overall Fisher's z

Fit the random-coefficients model.

ma\_cor\_scl <- rma.uni(yi, vi, method = "ML", data = dat\_scl3)  
summary(ma\_cor\_scl)

##   
## Random-Effects Model (k = 29; tau^2 estimator: ML)  
##   
## logLik deviance AIC BIC AICc   
## -19.4014 44.9285 42.8028 45.5374 43.2643   
##   
## tau^2 (estimated amount of total heterogeneity): 0.0822 (SE = 0.0549)  
## tau (square root of estimated tau^2 value): 0.2866  
## I^2 (total heterogeneity / total variability): 39.32%  
## H^2 (total variability / sampling variability): 1.65  
##   
## Test for Heterogeneity:   
## Q(df = 28) = 48.7514, p-val = 0.0089  
##   
## Model Results:  
##   
## estimate se zval pval ci.lb ci.ub   
## 1.2564 0.0852 14.7462 <.0001 1.0894 1.4234 \*\*\*   
##   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Back-transform Z to r.

pred\_r\_scl <- predict(ma\_cor\_scl, transf = transf.ztor)  
pred\_r\_scl

## pred ci.lb ci.ub cr.lb cr.ub  
## 0.8501 0.7967 0.8903 0.5851 0.9510

#### Moderators effect

Same as previously described.

# season  
ma\_cor\_scl\_year <- rma(yi, vi, mods = ~season, method = "ML", data = dat\_scl3)  
  
## region  
ma\_cor\_scl\_region <- rma(yi, vi, mods = ~region, method = "ML",data = dat\_scl3)  
  
## elevation as continuous  
ma\_cor\_scl\_elevation <- rma(yi, vi, mods = ~elevation, method = "ML", data = dat\_scl3)  
  
## elevation as category  
ma\_cor\_scl\_elevation2 <- rma(yi, vi, mods = ~elevation\_class, method = "ML", data = dat\_scl3)  
  
## incidence in the check as continuous  
ma\_cor\_scl\_inc <- rma(yi, vi, mods = ~inc\_check, method = "ML", data = dat\_scl3)  
  
## incidence in the check as categorical  
ma\_cor\_scl\_inc2 <- rma(yi, vi, mods = ~inc\_class, method = "ML", data = dat\_scl3)  
  
## yield in the check as continuous  
ma\_cor\_scl\_yld <- rma(yi, vi, mods = ~yld\_check, method = "ML", data = dat\_scl3)  
  
## yield in the check as categorical  
ma\_cor\_scl\_yld2 <- rma(yi, vi, mods = ~yld\_class, method = "ML", data = dat\_scl3)

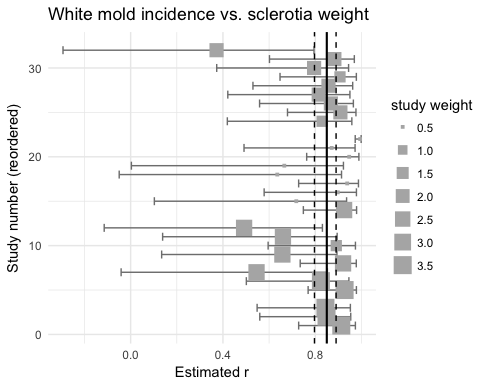
table\_scl <- frame\_data(  
~"Moderator", ~"Test of moderator", ~"R2",  
"Season", ma\_cor\_scl\_year$QMp, ma\_cor\_scl\_year$R2,  
"Region", ma\_cor\_scl\_region$QMp, ma\_cor\_scl\_region$R2,  
"Elevation categorical", ma\_cor\_scl\_elevation$QMp, ma\_cor\_scl\_elevation$R2,  
"Elevation continuous", ma\_cor\_scl\_elevation2$QMp, ma\_cor\_scl\_elevation2$R2,  
"Incidence categorical", ma\_cor\_scl\_inc$QMp, ma\_cor\_scl\_inc$R,  
"Yield continous", ma\_cor\_scl\_yld$QMp, ma\_cor\_scl\_yld$R2,  
"Yield categorical", ma\_cor\_scl\_yld2$QMp, ma\_cor\_scl\_yld2$R2  
)  
  
kable(table\_yld, format = "markdown", caption = "P-value of the test for the effect of the moderator and the amount of variance (heterogeneity) accounted for by the moderator (R2 statistics)")

|  |  |  |
| --- | --- | --- |
| Moderator | Test of moderator | R2 |
| Season | 0.4612331 | 15.2311624 |
| Region | 0.3253424 | 6.4018736 |
| Elevation categorical | 0.2620535 | 7.8577561 |
| Elevation continuous | 0.8234688 | 0.2695911 |
| Incidence categorical | 0.0604829 | 24.1929491 |
| Yield continous | 0.9372820 | 0.0985151 |
| Yield categorical | 0.8651536 | 0.3714379 |

#### Forest plot

Visualize the correlation coefficients by study and the mean r (solid line) and confidence intervals (dashed lines) from back-transforming Z estimated by the random-coefficients model. The size of the square is inversely proportion to the study's weight in the analysis.

wi <- 1/sqrt(dat\_scl3$vi)  
size <- 0.5 + 3.0 \* (wi - min(wi))/(max(wi) - min(wi))  
dat\_scl3 %>%   
 ggplot(aes(x = study, y = estimate)) +   
 geom\_errorbar(aes(ymin = conf.low, ymax = conf.high), color="grey50") +   
 geom\_point(aes(size = size), shape = 15, color="grey70") +  
 geom\_hline(yintercept = pred\_r\_scl$pred, size=0.75)+  
 geom\_hline(yintercept = c(pred\_r\_scl$ci.lb, pred\_r\_scl$ci.ub), linetype="dashed")+   
 coord\_flip()+  
 labs(x = "Slope")+   
 theme\_minimal()+   
 labs(size = "study weight",   
 title = "White mold incidence vs. sclerotia weight",   
 y = "Estimated r", x = "Study number (reordered)")



## Mixed effects model

### Incidence yield

We will use the lmer function of the lme4 to fit three different mixed models: random intercepts and slopes, random intercepts only and random slopes. We followed the procedures described in this [tutorial](http://www.metafor-project.org/doku.php/tips:two_stage_analysis) that compares the mixed model with the two-stage modeling approach.

#### Compare models

# null model  
mix\_yld <- lmer(yld ~ 1 + ( 1 |study), data=dat\_yld, REML=F)  
  
  
# random intercept and slopes  
mix\_yld1 <- lmer(yld ~ inc + (inc |study), data=dat\_yld, REML=F)  
  
# random slopes  
mix\_yld2 <- lmer(yld ~ inc + (1 | inc), data=dat\_yld, REML=F)  
  
# random intercepts  
mix\_yld3 <- lmer(yld ~ inc + (1 |study), data=dat\_yld, REML=F)

The AIC of the random intercepts and slopes model is the lowest.

AIC(mix\_yld1, mix\_yld2, mix\_yld3)

## df AIC  
## mix\_yld1 6 5319.291  
## mix\_yld2 4 6143.307  
## mix\_yld3 4 5331.700

anova(mix\_yld2, mix\_yld1)

## Data: dat\_yld  
## Models:  
## mix\_yld2: yld ~ inc + (1 | inc)  
## mix\_yld1: yld ~ inc + (inc | study)  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## mix\_yld2 4 6143.3 6159.1 -3067.7 6135.3   
## mix\_yld1 6 5319.3 5343.0 -2653.7 5307.3 828.02 2 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(mix\_yld3, mix\_yld1)

## Data: dat\_yld  
## Models:  
## mix\_yld3: yld ~ inc + (1 | study)  
## mix\_yld1: yld ~ inc + (inc | study)  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## mix\_yld3 4 5331.7 5347.5 -2661.8 5323.7   
## mix\_yld1 6 5319.3 5343.0 -2653.7 5307.3 16.41 2 0.0002733 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### Summary

summary (mix\_yld1)

## Linear mixed model fit by maximum likelihood ['lmerMod']  
## Formula: yld ~ inc + (inc | study)  
## Data: dat\_yld  
##   
## AIC BIC logLik deviance df.resid   
## 5319.3 5343.0 -2653.6 5307.3 376   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.7271 -0.6010 -0.0295 0.5099 3.2501   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## study (Intercept) 602184.6 776.006   
## inc 37.6 6.132 -0.31  
## Residual 36904.6 192.106   
## Number of obs: 382, groups: study, 35  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 3455.710 132.936 26.00  
## inc -17.243 1.457 -11.84  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## inc -0.308

confint(mix\_yld1)

## Computing profile confidence intervals ...

## 2.5 % 97.5 %  
## .sig01 620.3827291 1005.3038234  
## .sig02 -0.6804408 0.1393279  
## .sig03 3.4498863 9.5276481  
## .sigma 177.8775077 208.3304081  
## (Intercept) 3187.6241104 3723.8831605  
## inc -20.5599336 -14.2780331

#### Goodness of fit

Let's get the p-value for the significance of the slope and also the pseudo R2 (conditional) of the model.

# Get the p-values of the significance of the slope  
library(car)  
Anova(mix\_yld1)

## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: yld  
## Chisq Df Pr(>Chisq)   
## inc 140.14 1 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

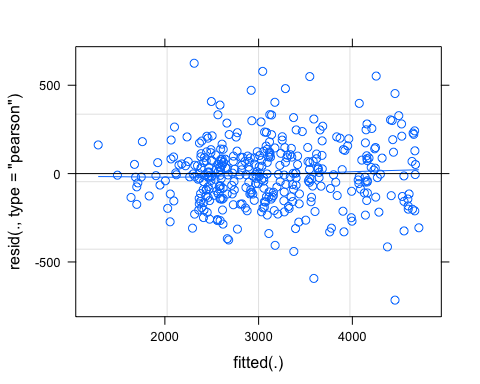
library(lmerTest)  
anova(mix\_yld1, ddf = "Kenward-Roger")

## Analysis of Variance Table  
## Df Sum Sq Mean Sq F value  
## inc 1 5171821 5171821 140.14

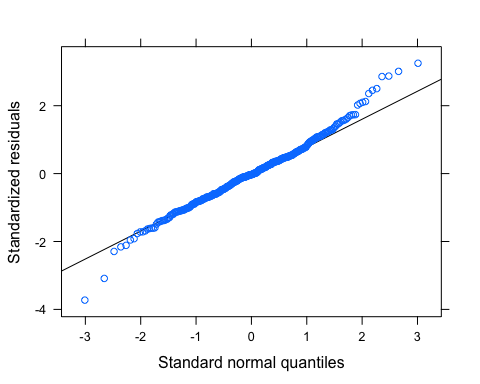
library(piecewiseSEM)  
sem.model.fits(mix\_yld1)

## Class Family Link N Marginal Conditional  
## 1 lmerMod gaussian identity 382 0.1384505 0.947683

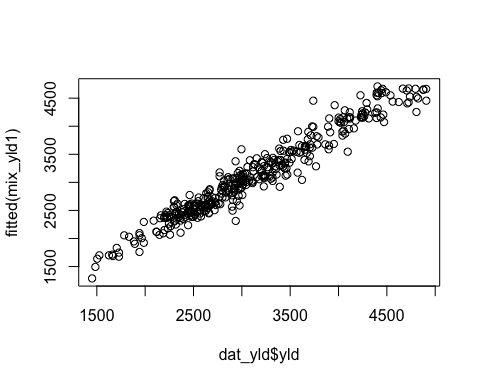
# plot the pearson's model residuals  
plot(mix\_yld1,type=c("p","smooth"))



# plot the qq-plot and qq-line  
library(lattice)  
qqmath(mix\_yld1)



# plot of the fitted and observations  
plot(fitted(mix\_yld1)~dat\_yld$yld)



#### Variance-covariance

We then proceed with the full model and calculate the variance and co-variance.

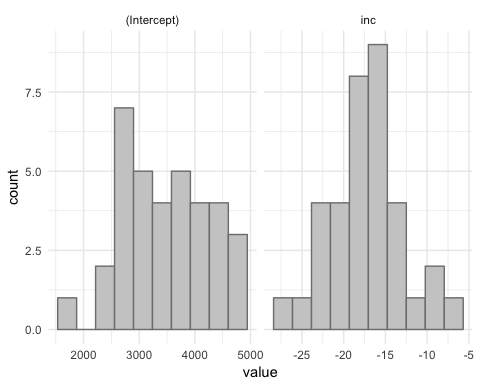
vc <- VarCorr(mix\_yld1)  
## variance only  
as.data.frame(vc, order="lower.tri")

## grp var1 var2 vcov sdcor  
## 1 study (Intercept) <NA> 602184.6028 776.0055430  
## 2 study (Intercept) inc -1477.4735 -0.3104987  
## 3 study inc <NA> 37.6002 6.1319002  
## 4 Residual <NA> <NA> 36904.6427 192.1058113

#### Predictions

We now extract the BLUEs and calculate the interdecile range.

# extract the blups  
cc2 <- coef(mix\_yld1)$study  
cc2 %>%   
 gather("coef", "value") %>%   
 ggplot(aes(x=value))+  
 theme\_minimal()+  
 geom\_histogram(bins = 10, fill = "grey80", color = "grey50")+  
 facet\_wrap(~coef, scales = "free\_x")



Calculate the interdecile range for the BLUEs of the slopes and intercepts

# Intercept  
dec90 <- quantile(cc2$`(Intercept)`,probs=c(.9))  
dec10 <- quantile(cc2$`(Intercept)`,probs=c(.1))  
dec90-dec10

## 90%   
## 1900.915

# Slopes  
dec90 <- quantile(cc2$inc,probs=c(.9))  
dec10 <- quantile(cc2$inc,probs=c(.1))  
dec90-dec10

## 90%   
## 10.41355

#### Effect of moderators

Forthcoming!

### Incidence-sclerotia

We repeat the same procedure as described above.

#### Compare models

library(lme4)  
# random intercept and slopes  
mix\_scl1 <- lmer(scl ~ inc + (inc |study), data=dat\_scl, REML=F)  
mix\_scl2 <- lmer(scl ~ inc + (1 | inc), data=dat\_scl, REML=F)  
mix\_scl3 <- lmer(scl ~ inc + (1 |study), data=dat\_scl, REML=F)

The AIC shows the random intercepts and slopes model with the lowest value.

AIC(mix\_scl1, mix\_scl2, mix\_scl3)

## df AIC  
## mix\_scl1 6 5379.063  
## mix\_scl2 4 5694.996  
## mix\_scl3 4 5471.522

#### Summary

summary (mix\_scl1)

## Linear mixed model fit by maximum likelihood   
## t-tests use Satterthwaite approximations to degrees of freedom ['lmerMod']  
## Formula: scl ~ inc + (inc | study)  
## Data: dat\_scl  
##   
## AIC BIC logLik deviance df.resid   
## 5379.1 5401.6 -2683.5 5367.1 310   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.9693 -0.4602 -0.0555 0.2682 5.6779   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr  
## study (Intercept) 431439 656.84   
## inc 4201 64.81 0.94  
## Residual 981744 990.83   
## Number of obs: 316, groups: study, 29  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 280.50 163.00 27.49 1.721 0.0965 .   
## inc 98.59 13.30 22.51 7.414 1.77e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## inc 0.406

#### Goodness of fit

Let's get the p-value for the significance of the slope and also the pseudo R2 (conditional) of the model.

# Get the p-values of the significance of the slope  
library(car)  
Anova(mix\_scl1)

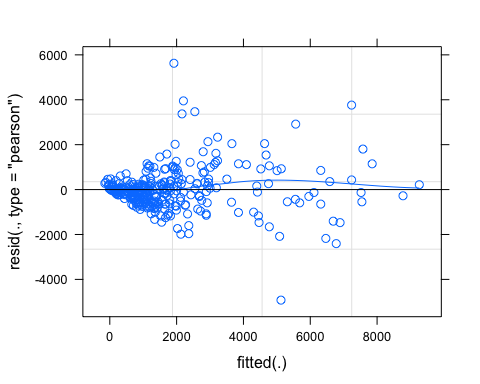
## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: scl  
## Chisq Df Pr(>Chisq)   
## inc 54.968 1 1.225e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

library(piecewiseSEM)  
sem.model.fits(mix\_scl1)

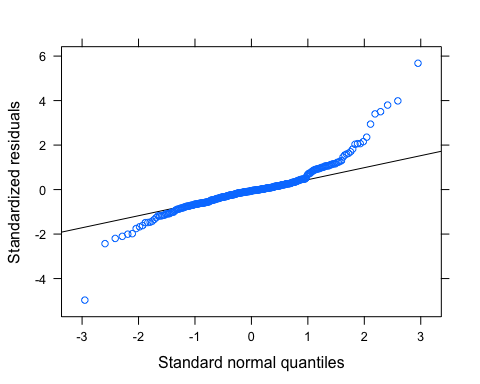
## Class Family Link N Marginal Conditional  
## 1 merModLmerTest gaussian identity 316 0.3454005 0.9004537

Check the goodness of fit of the model.

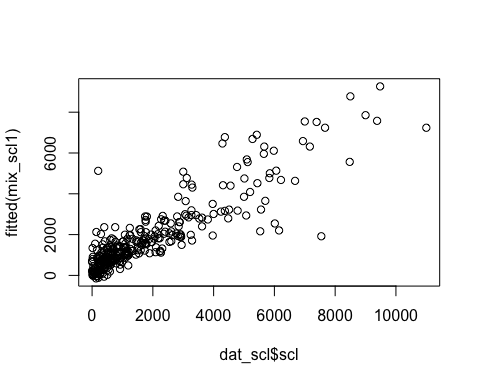
plot(mix\_scl1,type=c("p","smooth"))



library(lattice)  
qqmath(mix\_scl1)



plot(fitted(mix\_scl1)~dat\_scl$scl)



#### Variance-covariance

We then proceed with the full model and calculate the variance and co-variance.

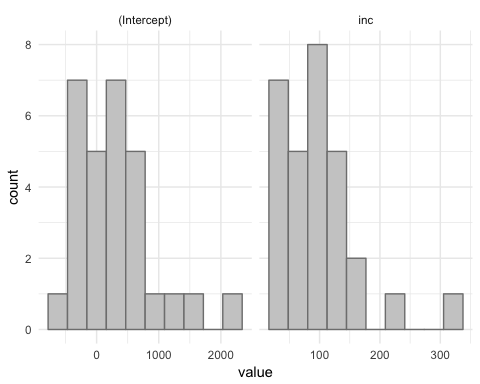
vc\_scl <- VarCorr(mix\_scl1)  
## variance only  
as.data.frame(vc\_scl, order="lower.tri")

## grp var1 var2 vcov sdcor  
## 1 study (Intercept) <NA> 431438.90 656.8400833  
## 2 study (Intercept) inc 40115.97 0.9422886  
## 3 study inc <NA> 4200.95 64.8147362  
## 4 Residual <NA> <NA> 981743.76 990.8298340

#### Predictions

Obtain the BLUEs and calculate the interdecile range.

# extract the blups  
cc2\_scl <- coef(mix\_scl1)$study  
cc2\_scl %>%   
 gather("coef", "value") %>%   
 ggplot(aes(x=value))+  
 theme\_minimal()+  
 geom\_histogram(bins = 10, fill = "grey80", color = "grey50")+  
 facet\_wrap(~coef, scales = "free\_x")



#### Effect of moderators

Forthcoming!

# References

Dalla Lana, F., Ziegelmann, P. K., Maia, A. de H. N., Godoy, C. V., & Del Ponte, E. M. (2015). Meta-Analysis of the Relationship Between Crop Yield and Soybean Rust Severity. Phytopathology, 105(3), 307–315. <doi:10.1094/phyto-06-14-0157-r>

Madden, L. V., & Paul, P. A. (2009). Assessing Heterogeneity in the Relationship Between Wheat Yield and Fusarium Head Blight Intensity Using Random-Coefficient Mixed Models. Phytopathology, 99(7), 850–860. <doi:10.1094/phyto-99-7-0850>