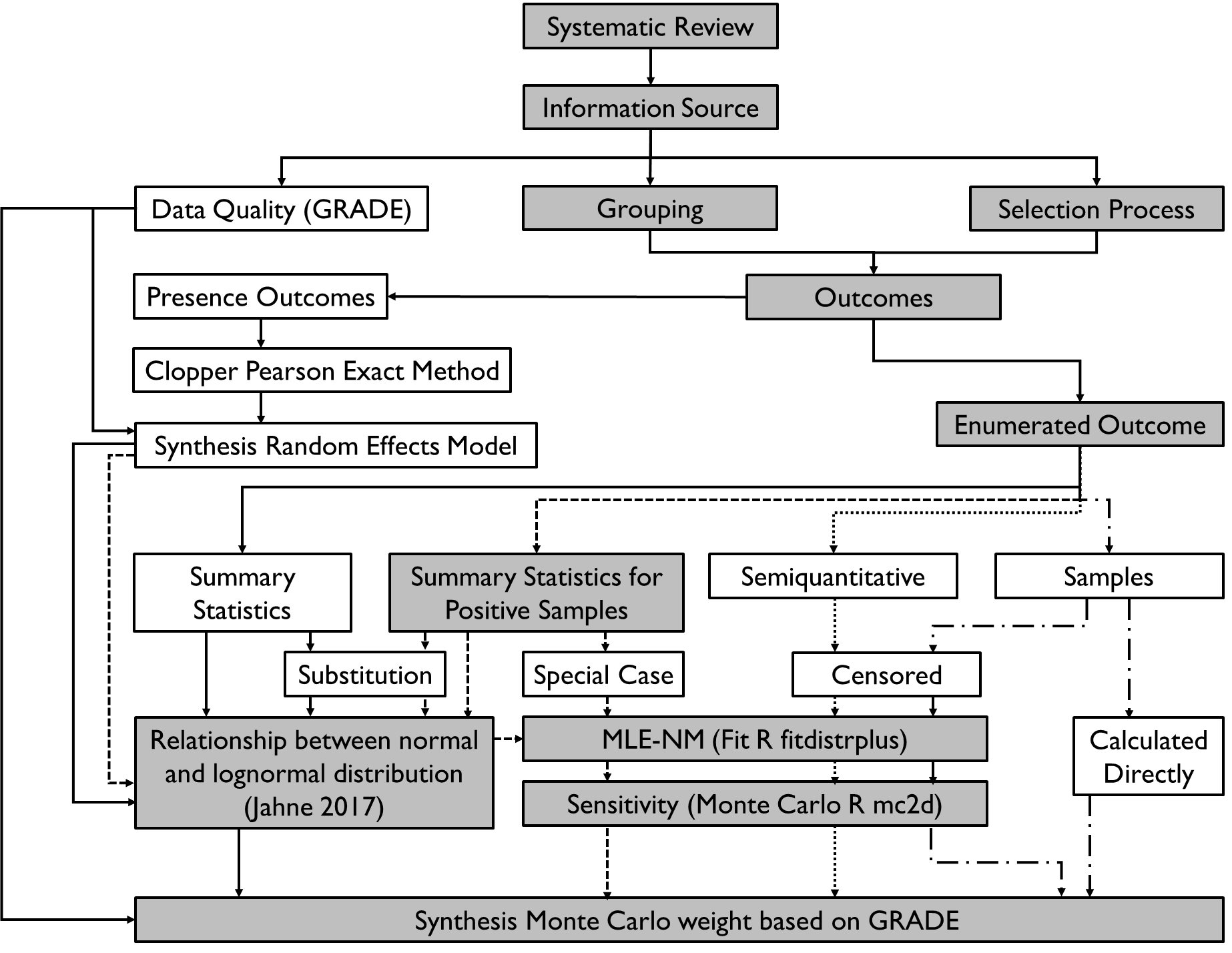
Example Data Processing an Meta-Analysis for Systematic review and statistical synthesis of rooftop animal fecal pathogens: for risk assessment

The following will give an example of how pathogen concentration outcomes were processed for one study.



**Fig. XXX.** This figure is Fig. XX from the review the boxes in grey represent the pathway that will be followed in the example below.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Step | Example | | | | | | | | |
| Systematic Review | Google Scholar search for: feces AND wildlife AND shedding AND pathogen | | | | | | | | |
| Information Source | Frick, C., J. Vierheilig, R. Linke, D. Savio, H. Zornig, R. Antensteiner, C. Baumgartner, C. Bucher, A. P. Blaschke, J. Derx, A. K. T. Kirschner, G. Ryzinska-Paier, R. Mayer, D. Seidl, T. Nadiotis-Tsaka, R. Sommer and A. H. Farnleitner (2018). "Poikilothermic Animals as a Previously Unrecognized Source of Fecal Indicator Bacteria in a Backwater Ecosystem of a Large River." Appl Environ Microbiol **84**(16): e00715-00718. | | | | | | | | |
| Selection Process | This article fit all of the eligibility criteria and was relevant since it had data on the concentration of pathogens in animal fecal material | | | | | | | | |
| Grouping | The following animals from this article were determined to have the ability to reach a roof top: gastropod, frog, and bird.  Frog was grouped under the amphibian class, while bird and gastropod were already classes.  The pathogens tested for were *Escherichia coli*, *Enterococci*, and *Clostridium perfringens.* Clostridium was included during data collection but was later excluded during the grouping step. *Escherichia coli*  and *Enterococci* were used as is because they were groups that had been decided on previously. | | | | | | | | |
| Outcomes | Pathogen | Source | Prev | n | Mean | Med. | 5% | 95% | Max |
| E. Coli | Bird | 15 | 73 | 5.0 | 4.8 | 2.3 | 8.5 | 9.2 |
| The mean, median, 5 and 95 percentile and max are in units of log10 CFU/g feces and all statistics were calculated excluding non-detects. The was also data for Enterococci, frogs, and gastropods but for the example we will only follow birds. The limit of detection was (LOD) was 1000 CFU/g | | | | | | | | |
| Enumerated Outcomes | |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Pathogen | Source | Mean | Median | 5% | 95% | Max | | E. Coli | Bird | 1.00×105 | 6.31×104 | 2.00×102 | 3.16×108 | 1.58×109 | | The mean, median, 5 and 95 percentile and max are in units of CFU/g feces and all statistics were calculated excluding non-detects. | | | | | | | | | | | | | | | |
| Summary Statistics for Positive samples | |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Pathogen | Source | Mean | Median | 5% | 95% | Max | | E. coli | Bird | 1.00×105 | 6.31×104 | 2.00×102 | 3.16×108 | 1.58×109 | | The mean, median, 5 and 95 percentile and max are in units of CFU/g feces and all statistics were calculated excluding non-detects. The limit of detection was (LOD) was 1000 CFU/g | | | | | | | | | | | | | | | |
| Relationship between normal and log normal distributiosn | For the rest of the rest of this example I will only follow E. coli in bird samples.  The following equations from the supplementary materials for (Jahne, Schoen et al. 2017) were used to fit a distribution to the | | | | | | | | |
| MLE-NM fit | the next step was to convert from a distribution for values with no non-to detects to include non detects. This is done by generating numbers above the Limit of Detection (15 positive samples) and values below the LOD (58). These values were generated using the rtrunc command from the mc2d package in R.  Values below LOD  rtrunc("rlnorm", n=LoDPrevBelow, mean = Impmean, sd = ImpSD, lsup = LOD)  rtrunc("rlnorm", n=58, mean = 13.6, sd = 6.93, lsup = 1000)  Values above LOD  rtrunc("rlnorm", n=LoDPrev, mean = Impmean, sd = ImpSD, linf = LOD)  rtrunc("rlnorm", n=15, mean = 13.6, sd = 6.93, linf = 1000)  The values above and below the LOD were combined to create a single data set. The fitdist function from the fitdistrplus R package was used to fit a distribution to the data set that included values below LOD. This would generate new µ and σ for the complete dataset. The fitdist function uses maximum likelihood estimation with the Nelder-Mead method to fit a distribution.  fitdist(e, "lnorm") | | | | | | | | |
| Sensitivity | To determine the sensitivity of this method 10000 of the complete datasets were generated and 10000 µ and σ.  for (i in 1:10000) {  c <- rtrunc("rlnorm", n=LoDPrevBelow, mean = Impmean, sd = ImpSD, lsup = LOD)  d <- rtrunc("rlnorm", n=LoDPrev, mean = Impmean, sd = ImpSD, linf = LOD)  e <- c(c,d)  f <- fitdist(e, "lnorm")  Ten thousand datasets were generated to account for the randomness in the way that data sets were generated. The average µ and σ for the 10000 datasets was considered the µ and σ for the whole dataset.  The distribution parameters for the example study were µ =5.55 nd σ = 4.06 | | | | | | | | |
| Synthesis | To synthesize data from multiple studies the studies were weighed based on there GRADE. Frick et al. 2018 had a GRADE of Moderate so had a weight of 100. There were 11 other datasets from 5 reports that had data on E. coli from birds.  Weight of Frick et al. 2018 =100  Weight of other datasets = 10 + 100 + 4 × 100 + 5 × 100 + 1000 = 2010  Total weight = 2110  A 10000 point data set was created by randomly generating a portion of the 10000 points from the distributions of each data set.  For Frick et al. 2018  EnumAniSubset$WValues <- EnumAniSubset$Weight/EnumSum\*10000  EnumAniSubset$WValues <- round(EnumAniSubset$WValues, digits=0)  To randomly generate the points the rlnorm function from base R was used.  rlnorm(EnumRandn, meanlog = EnumRandmean, sdlog = EnumRandsd)  rlnorm(474, 5.55 = EnumRandmean, sdlog = 4.06)  A distribution was then fit to the 10000 point synthesized dataset using the fitdist function from the fitdistrplus R package use the same method as in the MLE-NM fit section.  This process was repeated 10000 times and the average µ and σ from the synthesized datasets was calculated (µ = 9.19 and σ = 7.04). | | | | | | | | |

Jahne, M. A., M. E. Schoen, J. L. Garland and N. J. Ashbolt (2017). "Simulation of enteric pathogen concentrations in locally-collected greywater and wastewater for microbial risk assessments." Microb Risk Anal **5**: 44-52.