Infectivity of human norovirus in live challenge trials: a systematic review and meta-analysis

W. Zane Billings1, Anne Marie Dye2, and Andreas Handel1

2021-12-13

Human norovirus is the most common cause of acute gastroenteritis and food-borne illness in the United States. A low infectious viral load, high environmental persistence, potential for aerosolization, and long-term induction of viral shedding work in tandem to make norovirus extremely transmissible. However, norovirus outbreaks are often self-limiting, and the wide variety of potential routes of exposure make estimates of transmissibility from outbreak data somewhat unreliable. In contrast, human challenge studies involve a controlled inoculation, after which subjects are closely monitored for the duration of inducted illness. Such challenge studies provide the opportunity to control factors like inoculum strain, inoculum dose, and participant histo-blood group. Recall bias is also less likely to influence the result of challenge studies. Thus, using challenge studies to explore norovirus infectivity is less prone to bias than challenge studies, even if the results are not always generalizable to natural outbreaks. We conducted a systematic review of the literature in order to find all studies which report using norovirus challenge data, and from these studies, we abstract data to obtain unique challenge cohort data. From unique cohorts, we conduct a meta-analysis of the proportion of individuals infected during each study, including subgroup analyses by study risk of bias, inoculum genogroup, and FUT2 participant genotype control. We also perform a meta regression on inoculum dose. We find that [EXPLAIN RESULTS].

1 Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia  
2 Department of Environmental Health, College of Public Health, University of Georgia

# Introduction

Human norovirus (NoV) is a small, round-structured, unenveloped, positive-sense single-strand RNA virus with a linear, unsegmented genome, belonging to the caliciviridae family. NoV was first isolated in .

# Methods

We conducted a systematic review of papers using data from norovirus challenge studies, and from the included reports, we identified unique studies. Then, we abstracted data on the proportion of infected individuals from each study, and conducted a meta-analysis of these data.

## Literature search

We searched two databases to find literature:

* PubMed [(https://www.ncbi.nlm.nih.gov/pubmed/)](https://www.ncbi.nlm.nih.gov/pubmed/) and
* Web of Science [(https://www.webofscience.com/)](https://www.webofscience.com/).

The search terms, included in Table 1, contained terms for norovirus, terms related to human studies, and terms related to the challenge (or volunteer) study design. The databases were searched on September ##, 2021, with no filter for date used during the search.

Table 1: Search strings for the two databases searched.

| **database** | **search strategy** |
| --- | --- |
| PubMed | ("norovirus" [MeSH Major Topic]) AND ("norovirus" OR "Norwalk virus" OR "snow mountain virus" OR  "Norwalk agent" OR "nonbacterial gastroenteritis" OR  "viral gastroenteritis" [Title/Abstract]) AND (human OR challenge OR experimental OR infect\* OR volunteer OR  vaccin\* OR adult OR clinical OR randomized OR  individual [Title/Abstract]) NOT ("mouse" or "murine" or "mice" [Title]) |
| WoS | (TS=("norovirus" OR "Norwalk virus" OR "snow mountain virus" OR  "Norwalk agent" OR "nonbacterial gastroenteritis" OR  "viral gastroenteritis")) AND (TS=(human OR man OR adult OR volunteer)) AND (TS=(volunteer OR challenge OR experimental OR infect\* OR vaccin\* OR  inoculum)) |

Reports were selected for inclusion if they referenced a human norovirus challenge study with controlled inoculation of participants. No other inclusion criteria (location, date, study design, etc.) were used. Reports were excluded from the review if they were not primary research articles (i.e. reviews or letters to the editor), not written in English, or not available to the authors. Two reviewers (WZB and AMD) independently reviewed titles and abstracts, with disagreements being resolved by a third reviewer (AH). The reviewers repeated the process for the full-text review.

From the included reports, the reviewers identified unique studies by examining citations for data sources. We created a directed acyclic graph (DAG) based on citations in reference sections of papers–in this format, the end notes of the DAG are studies with original populations. Only reports which appeared to discuss original data collection and did not cite their data as coming from a previous report were considered to be unique studies.

## Data abstraction

From each study (only reports which included original data collection), we abstracted the following information:

1. Reference information including name of the first author and the year of publication;
2. Study design for the challenge portion of the study (either case series or randomized trial);
3. Demographic information reported for each study including age range, percentage of male participants, percentage of white participants, and location of the study site;
4. Whether the study controlled for participant FUT2 genotype;
5. Any other eligibility criteria for study enrollment;
6. Inoculum dose and genotype;
7. Criteria for reporting infection and illness; and
8. For each combination of FUT2 status, dose, and other study sub-samples (e.g. vaccine vs. placebo in vaccine trials), the number of participants challenged, number of participants with confirmed infection, and the number of participants with confirmed illness.

Reported inoculum strains were standardized to modern nomenclature where possible, criteria for infection and illness were standardized, and inoculum dose was converted to genome equivalent copies (1 RT-PCR unit = 400 genome equivalent copies) [CITE ATMAR PAPER HERE].

## Study quality assessment

## Statistical methods

We fit an overall meta-analysis model for the proportion of infected individuals (PI) by pooling together all subgroups reported within each study. We used a generalized linear mixed-effected modeling approach with logit-transformed proportions (log-odds) as the outcome.1,2 The method of maximum likelihood was used to estimate (between-study heterogeneity). Confidence intervals for pooled effects were calculated using the Knapp-Hartung adjustment,3,**sidik2002?** which is typically sensible2,4,**langan2019?**

The GLMM method with log-odds as the outcome has been previously recommended in the literature for the meta-analysis of proportions.5 When using a GLMM approach, no weights are estimated for each study, and (the estimate of between-study heterogeneity) can only be estimated through the method of maximum likehood, and a confidence interval for cannot be obtained.2

Influence of individual studies on the overall result was analyzed using three methods. First, we used a simple method which classifies studies as outliers if the estimate confidence interval for the individual study does not overlap with the confidence interval for the pooled estimate.2 Second, we used a leave-one-out approach and manually identified outliers using a Baujat plot6 and diagnostics.7 Third, we used the GOSH (Graphical display Of Study Heterogeneity) method, which fit 1,000,000 random subset models, and plots the estimated heterogeneity vs. the estimated effect size.8 From the GOSH results, we applied three unsupervised clustering algorithms: -means,**hartigan1979?** DBSCAN,**schubert2017?** and Gaussian mixture modeling.**fraley2002?** Study over-representation within clusters is used to determine which studies have an undue effect on heterogeneity.2

## Software

Reference management was conducted using both EndNote9 for deduplication and searching for missing reference fields, and Zotero10 for archival purposes. Data abstraction and review of reports was conducted using Microsoft Excel 365 (Microsoft Corporation, Santa Rosa, CA, USA) and Google Sheets (Google, Mountain View, CA, USA).

All analyses were conducted using R version 4.1.1.11 The packages meta (CITE), metafor (CITE THIS), and dmetar (CITE THIS) were used for meta-analysis. Figures were generated using the analysis packages along with ggplot2 (CITE THIS) and PRISMA2020.12–14 Tables were generated using the package flextable.15 This report was generated using R Markdown with the packages rmarkdown,16–18 knitr,19–21 and bookdown.22,23 Several additional packages were used for data cleaning and wrangling (CITE THESE PACKAGES). A complete printout of the R session information can be found in the Appendix.

# Results

## Identification of studies

## Study characteristics

## Overall meta-analysis

## Influence analyses

## Subgroup analyses

## Meta-regression

# Discussion

# Acknowledgements

The authors wish to acknowledge Dr. Mark Ebell (Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia) for guidance throughout the project.

# References

1. Stijnen T, Hamza TH, Özdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine*. 2010;29(29):3046-3067. doi:[10.1002/sim.4040](https://doi.org/10.1002/sim.4040)

2. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. *Doing Meta-Analysis in R: A Hands-On Guide*. Chapmann & Hall/CRC Press; 2021. Accessed December 2, 2021. <https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/>

3. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*. 2003;22(17):2693-2710. doi:[10.1002/sim.1482](https://doi.org/10.1002/sim.1482)

4. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology*. 2014;14(1):25. doi:[10.1186/1471-2288-14-25](https://doi.org/10.1186/1471-2288-14-25)

5. Schwarzer G, Chemaitelly H, Abu-Raddad LJ, Rücker G. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Res Synth Methods*. 2019;10(3):476-483. doi:[10.1002/jrsm.1348](https://doi.org/10.1002/jrsm.1348)

6. Baujat B, Mahé C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. *Statistics in Medicine*. 2002;21(18):2641-2652. doi:[10.1002/sim.1221](https://doi.org/10.1002/sim.1221)

7. Viechtbauer W, Cheung MW-L. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*. 2010;1(2):112-125. doi:[10.1002/jrsm.11](https://doi.org/10.1002/jrsm.11)

8. Olkin I, Dahabreh IJ, Trikalinos TA. GOSH – a graphical display of study heterogeneity. *Research Synthesis Methods*. 2012;3(3):214-223. doi:[10.1002/jrsm.1053](https://doi.org/10.1002/jrsm.1053)

9. The EndNote Team. *EndNote*. Clarivate; 2013.

10. Center for History and New Media. *Zotero*. George Mason University; 2021.

11. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021. <https://www.R-project.org/>

12. Haddaway N, McGuinness L, Pritchard C. *Prisma2020: Make Interactive PRISMA Flow Diagrams*.; 2021. <https://CRAN.R-project.org/package=PRISMA2020>

13. Haddaway NR, Pritchard CC, McGuinness LA. *Prisma2020: R Package and ShinyApp for Producing PRISMA 2020 Compliant Flow Diagrams (Version 0.0.2)*.; 2021. doi:[10.5281/zenodo.5082518](https://doi.org/10.5281/zenodo.5082518)

14. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Published online July 15, 2021:2021.07.14.21260492. doi:[10.1101/2021.07.14.21260492](https://doi.org/10.1101/2021.07.14.21260492)

15. Gohel D. *Flextable: Functions for Tabular Reporting*.; 2021. <https://CRAN.R-project.org/package=flextable>

16. Allaire J, Xie Y, McPherson J, et al. *Rmarkdown: Dynamic Documents for r*.; 2021. <https://CRAN.R-project.org/package=rmarkdown>

17. Xie Y, Allaire JJ, Grolemund G. *R Markdown: The Definitive Guide*. Chapman; Hall/CRC; 2018. <https://bookdown.org/yihui/rmarkdown>

18. Xie Y, Dervieux C, Riederer E. *R Markdown Cookbook*. Chapman; Hall/CRC; 2020. <https://bookdown.org/yihui/rmarkdown-cookbook>

19. Xie Y. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*.; 2021. <https://yihui.org/knitr/>

20. Xie Y. Knitr: A comprehensive tool for reproducible research in R. In: Stodden V, Leisch F, Peng RD, eds. *Implementing Reproducible Computational Research*. Chapman; Hall/CRC; 2014. <http://www.crcpress.com/product/isbn/9781466561595>

21. Xie Y. *Dynamic Documents with R and Knitr*. 2nd ed. Chapman; Hall/CRC; 2015. <https://yihui.org/knitr/>

22. Xie Y. *Bookdown: Authoring Books and Technical Documents with r Markdown*.; 2021. <https://CRAN.R-project.org/package=bookdown>

23. Xie Y. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman; Hall/CRC; 2016. <https://bookdown.org/yihui/bookdown>

# Data and code availibility

All data sheets, along with cleaning and analysis code, are available on [GitHub](link), and the code at time of writing is archived on Zenodo: [doi url](link).

# R session information

**R version 4.1.1 (2021-08-10)**

**Platform:** x86\_64-w64-mingw32/x64 (64-bit)

**locale:** *LC\_COLLATE=English\_United States.1252*, *LC\_CTYPE=English\_United States.1252*, *LC\_MONETARY=English\_United States.1252*, *LC\_NUMERIC=C* and *LC\_TIME=English\_United States.1252*

**attached base packages:** *stats*, *graphics*, *grDevices*, *utils*, *datasets*, *methods* and *base*

**other attached packages:** *dmetar(v.0.0.9000)*, *metafor(v.3.0-2)*, *Matrix(v.1.3-4)*, *meta(v.5.0-0)*, *PRISMA2020(v.0.0.3)*, *flextable(v.0.6.9)*, *knitr(v.1.36)*, *bookdown(v.0.24)* and *rmarkdown(v.2.11)*

**loaded via a namespace (and not attached):** *magic(v.1.5-9)*, *splines(v.4.1.1)*, *assertthat(v.0.2.1)*, *stats4(v.4.1.1)*, *pander(v.0.6.4)*, *robustbase(v.0.93-9)*, *yaml(v.2.2.1)*, *ggrepel(v.0.9.1)*, *gdtools(v.0.2.3)*, *pillar(v.1.6.4)*, *lattice(v.0.20-45)*, *glue(v.1.4.2)*, *uuid(v.0.1-4)*, *digest(v.0.6.28)*, *minqa(v.1.2.4)*, *colorspace(v.2.0-2)*, *MuMIn(v.1.43.17)*, *htmltools(v.0.5.2)*, *netmeta(v.2.0-1)*, *pkgconfig(v.2.0.3)*, *purrr(v.0.3.4)*, *scales(v.1.1.1)*, *officer(v.0.4.0)*, *lme4(v.1.1-27.1)*, *tibble(v.3.1.4)*, *generics(v.0.1.0)*, *ggplot2(v.3.3.5)*, *ellipsis(v.0.3.2)*, *nnet(v.7.3-16)*, *magrittr(v.2.0.1)*, *crayon(v.1.4.2)*, *mclust(v.5.4.7)*, *evaluate(v.0.14)*, *fansi(v.0.5.0)*, *nlme(v.3.1-153)*, *MASS(v.7.3-54)*, *xml2(v.1.3.2)*, *class(v.7.3-19)*, *tools(v.4.1.1)*, *data.table(v.1.14.2)*, *lifecycle(v.1.0.1)*, *stringr(v.1.4.0)*, *kernlab(v.0.9-29)*, *munsell(v.0.5.0)*, *cluster(v.2.1.2)*, *zip(v.2.2.0)*, *fpc(v.2.2-9)*, *compiler(v.4.1.1)*, *systemfonts(v.1.0.2)*, *rlang(v.0.4.11)*, *grid(v.4.1.1)*, *nloptr(v.1.2.2.2)*, *CompQuadForm(v.1.4.3)*, *base64enc(v.0.1-3)*, *boot(v.1.3-28)*, *gtable(v.0.3.0)*, *abind(v.1.4-5)*, *flexmix(v.2.3-17)*, *DBI(v.1.1.1)*, *R6(v.2.5.1)*, *gridExtra(v.2.3)*, *prabclus(v.2.3-2)*, *dplyr(v.1.0.7)*, *fastmap(v.1.1.0)*, *utf8(v.1.2.2)*, *mathjaxr(v.1.4-0)*, *poibin(v.1.5)*, *modeltools(v.0.2-23)*, *stringi(v.1.7.4)*, *parallel(v.4.1.1)*, *Rcpp(v.1.0.7)*, *vctrs(v.0.3.8)*, *DEoptimR(v.1.0-9)*, *tidyselect(v.1.1.1)*, *xfun(v.0.26)* and *diptest(v.0.76-0)*