The Journal of Open Source Software

fgwqsr: An R package for Frequentist Grouped Weighted Quantile Sum Regression

Daniel Rud [1*] **and Juan Pablo Lewinger** [1*]

1 Department of Population and Public Health Sciences, University of Southern California, USA * These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Nick Golding

Reviewers:

- @ibarraespinosa
- @msalibian

Submitted: 01 April 2025
Published: unpublished

Summary

Environmental epidemiologists frequently study the effects of chemical and pollutant exposures on health outcomes. Beyond single-constituent models, recent epidemiological methods focus on modeling exposure mixtures jointly, accounting for the correlation between exposures arising from common sources (Carrico et al., 2015; Gibson et al., 2019; Hamra & Buckley, 2018; Joubert et al., 2022; Wheeler et al., 2021a, 2021b).

Among existing approaches, Weighted Quantile Sum Regression (WQSR) has gained traction for evaluating associations between exposure mixtures and health outcomes (Carrico et al., 2015; Wheeler et al., 2021a, 2021b). WQSR estimates both (1) group effects, which quantify the impact of a mixture group, and (2) sets of group weights, which represent the relative contributions of individual constituents within a mixture. In the binary outcome setting, the WQSR model is formulated as:

$$y_i \sim \text{Bernoulli}(\pi_i)$$
$$\text{logit}(\pi_i) = c_0 + \sum_{g=1}^{G} \gamma_g \left(\sum_{k=1}^{c_g} w_{g,k} \cdot q_{g,k,i} \right) + \sum_{r=1}^{R} \phi_r z_{r,i}$$

where, for subject i, y_i represents the observed disease outcome, π_i the probability of disease, $q_{g,k,i}$ the exposure to chemical k in mixture group g, and $z_{r,i}$ the r^{th} confounder adjustment. The weights for mixture group g satisfy $\sum_{k=1}^{c_g} w_{g,k} = 1$ and $w_{g,k} \geq 0$. The parameter γ_g represents the group effect for a given mixture group, capturing the impact of a one-quantile increase in all chemical constituents within the group. WQSR models are constrained such that all constituents from a particular mixture group have effects in the same direction, which functions as a form of regularization to stabilize the effect estimates of the highly correlated exposures.

The `fgwqsr` package implements the Frequentist Grouped Weighted Quantile Sum Regression (FGWQSR) model introduced in Rud et al. (2025). Its main function, `fgwqsr`, accommodates binary, continuous, and count outcome types. To fit a FGWQSR model, users must specify a special model formula using vertical bars (`|`) to separate mixture group elements and a forward slash (`/`) to separate mixture groups from unconstrained covariates. Categorical covariates must be prefixed with `i.`. For instance, in an analysis with outcome Y, mixture groups $\{A_1, A_2\}$ and $\{B_1, B_2\}$, and confounders $\{W_1, W_2, W_3\}$ (where W_1, W_2 are numeric and W_3 is categorical), the model formula is:

```
model_formula = Y ~ A1 + A2 | B1 + B2 / W_1 + W_2 + i.W_3
```

If no adjusting covariates are included, no forward slash is required:

```
model_formula = Y ~ A1 + A2 | B1 + B2
```

For a single mixture group, vertical bars are not necessary:

```
model_formula = Y ~ A1 + A2 / W_1 + W_2 + i.W_3
```

Given `model_formula`, the outcome family type `family` being one of (`"binomial"`, `"gaussian"`, `"poisson"`), the number of quantiles `q` desired for the quantization of the mixture constituents, the number of multivariate normal simulations `n_mvn_rep` performed for each hypothesis test, and the number of cores `cores` one is willing to parallelize over, an FGWQSR model can be fitted with the call:

```
fgwqsr_model = fgwqsr(formula = model_formula,
                      data = data,
                      quantiles = q,
                      family = family,
                      n_mvn_sims = n_mvn_rep,
                      verbose = T,
                      cores = cores)
```

Results can be examined using `summary(fgwqsr_model)`, which provides parameter estimates for group effects, group weights, and statistical tests for both group and single-constituent effects. Forest plots for both group effects and single constituent effects can be plotted using `plot(fgwqsr_model)`.

An optional tuning parameter, `zero_threshold_cutoff`, is used in the non-regular statistical testing procedure. This parameter determines how often near-boundary estimates are assigned a boundary cone in the constrained multivariate normal Monte Carlo inference procedure. A default value of 0.5 has been shown to perform well across various scenarios, though reasonable values range from [0.05, 0.5]. More details are provided in Rud et al. (2025).

In addition to FGWQSR, the package includes an implementation of Bayesian Grouped Weighted Quantile Sum Regression (BGWQSR) for binary outcomes. Unlike the BayesGWQS package, our implementation leverages the `runjags` package for parallelized Markov Chain Monte Carlo (MCMC) sampling. BGWQSR models can be fitted using the `bgwqsr` function, with additional MCMC control parameters available. Visualization tools such as `plot_results`, `plot_betas`, and `plot_weights` provide graphical summaries of group effects, weights, and confounder estimates with posterior credible intervals.

For further guidance, see the package vignette here.

Statement of need

FGWQSR was developed to address several limitations of existing WQSR methods. Many existing approaches (Carrico et al., 2015; Wheeler et al., 2021a) rely on data splitting, requiring separate datasets to first estimate group weights and then assess group effect parameters. In contrast, FGWQSR jointly estimates group effects and group weights using a constrained optimization procedure (Rud et al., 2025), eliminating the need for data splitting.

Moreover, existing WQSR implementations struggle with large datasets. FGWQSR was designed to efficiently handle large datasets and was successfully applied to a dataset with 317,767 observations, which previous implementations struggled to accommodate. Additionally, FGWQSR extends the statistical framework by introducing statistical tests for both group and single-constituent effects, whereas previous WQSR models focused solely on group effects.

Thus, FGWQSR represents a significant advancement in WQSR methodology, providing a scalable, statistically rigorous approach that does not require data splitting, handles large datasets, and enables statistical inference for both group and individual constituent effects.

Installation

The most current version of `fgwqsr` package can be downloaded from github using the following instructions:

```
install.packages("remotes")
remotes::install_github("Daniel-Rud/fgwqsr")
```

References

Carrico, C., Gennings, C., Wheeler, D. C., & Factor-Litvak, P. (2015). Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. *Journal of Agricultural, Biological, and Environmental Statistics*, *20*(1), 100–120. https://doi.org/10.1007/s13253-014-0180-3

Gibson, E. A., Nunez, Y., Abuawad, A., Zota, A. R., Renzetti, S., Devick, K. L., Gennings, C., Goldsmith, J., Coull, B. A., & Kioumourtzoglou, M.-A. (2019). An overview of methods to address distinct research questions on environmental mixtures: An application to persistent organic pollutants and leukocyte telomere length. *Environmental Health*, *18*(76). https://doi.org/10.1186/s12940-019-0515-1

Hamra, G. B., & Buckley, J. P. (2018). Environmental exposure mixtures: Questions and methods to address them. *Current Epidemiology Reports*, *5*(2), 160–165. https://doi.org/10.1007/s40471-018-0145-0

Joubert, B. R., Kioumourtzoglou, M.-A., Chamberlain, T., Chen, H. Y., Gennings, C., Turyk, M. E., Miranda, M. L., Webster, T. F., Ensor, K. B., Dunson, D. B., & Coull, B. A. (2022). Powering research through innovative methods for mixtures in epidemiology (PRIME) program: Novel and expanded statistical methods. *International Journal of Environmental Research and Public Health*, *19*(3), 1378. https://doi.org/10.3390/ijerph19031378

Rud, D., Mostafijur Rahman, M., Xiang, A. H., McConnell, R., Lurmann, F., Kleeman, M. J., Schwartz, J., Chen, Z., Eckel, S., & Pablo Lewinger, J. (2025). Frequentist grouped weighted quantile sum regression for correlated chemical mixtures. *Statistics in Medicine*, *44*(7), e70078. https://doi.org/10.1002/sim.70078

Wheeler, D. C., Rustom, S., Carli, M., Whitehead, T. P., Ward, M. H., & Metayer, C. (2021a). Assessment of grouped weighted quantile sum regression for modeling chemical mixtures and cancer risk. *International Journal of Environmental Research and Public Health*, *18*(2). https://doi.org/10.3390/ijerph18020504

Wheeler, D. C., Rustom, S., Carli, M., Whitehead, T. P., Ward, M. H., & Metayer, C. (2021b). Bayesian group index regression for modeling chemical mixtures and cancer risk. *International Journal of Environmental Research and Public Health*, *18*(7). https://doi.org/10.3390/ijerph18073486