Help Document: APC Visualization Platform

Junyu Long

2026-02-26

1. About this platform

1.1 Aim

The primary aim of this platform is to conveniently facilitate Hierarchical Age-Period-Cohort (HAPC) analysis for researchers. It provides an integrated workflow for data visualization, model estimation, and the generation of result tables and figures.

A key highlight is the interactive 3D slicing capability, which transforms static data into dynamic visualizations to help researchers uncover hidden patterns within the Age-Period-Cohort structure.

1.2 Theoretical basis

APC identification problem

In demographic and public health research, it is crucial to disentangle the effects of Age (biological aging), Period (historical events at the time of survey), and Cohort (generational experiences). However, estimating these three effects simultaneously in standard regression models is mathematically impossible due to perfect linear dependency: $Cohort = Period - Age$. This is known as the classic "APC identification problem", which leads to collinearity and model failure.

The HAPC-CCREM solution

The Hierarchical Age-Period-Cohort Cross-Classified Random Effects Model (HAPC-CCREM), pioneered by Yang and Land (2013), resolves this collinearity by treating the data hierarchically rather than as a single-level flat structure.

Because individuals are simultaneously nested within a specific survey year (Period) and a specific birth generation (Cohort)—and periods and cohorts overlap rather than being purely nested within each other—the model structure is cross-classified. The HAPC framework thus treats age as fixed effect, period and cohort as random effects.

The combined equation can be expressed as:

$$\text{logit}(P(Y_{ijk} = 1)) = \beta_0 + \beta_1 \text{Age}_{ijk} + \beta_2 \text{Age}_{ijk}^2 + \beta^T \mathbf{X}_{ijk} + u_j + v_k$$

Where:

- Y_{ijk} is the binary disease outcome for individual i, in period j, and cohort k.
- β_0 is the overall intercept.
- $\mathbf{X}_{ijk}$ represents the vector of other user-specified fixed-effect covariates.
- $u_j \sim N(0, \sigma_u^2)$ is the random effect for Period j.
- $v_k \sim N(0, \sigma_v^2)$ is the random effect for Cohort k.

1.3 Citation, credits and contributors

Methodological framework:

The statistical framework of this platform is primarily based on the Hierarchical Age-Period-Cohort (HAPC) models as detailed in:

Yang, Y., & Land, K. C. (2013). *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications.* CRC Press. (Chapters 7 & 8).

Software

This interactive platform was built using R version 4.2.0. Key R packages include:

Category	R Packages
Core Infrastructure	shiny, shinyWidgets, htmlwidgets, digest
Data Import & Wrangling	tidyr, dplyr, haven, readxl, tools
Statistical Modeling	glmmTMB, emmeans, bruceR
Visualization	plotly, ggplot2, ggsci
Results Reporting	DT, sjPlot

Contributors

Principal Developer: Junyu Long, School of Public Health, Peking University.

Technical Assistance: Supported by AI assistants for code optimization and UI refinement.

2. User guideline

2.1 Data input

To ensure a smooth and accurate analysis, please prepare your dataset according to the following specifications.

File requirement

Currently, this platform supports .csv, .dta, .sav, .xlsx, .RData, and .rds files. Your dataset must be individual-level data containing:

Column	Description	Requirement
disease	The independent variable	Must be numeric, with coding:0 = no disease;1 = having disease
age	Individuals' age at the year of survey	numeric
period	The survey year	numeric

Period range

Please select the start year and end year of your dataset here.

Age range

Select the minimum and maximum age of your dataset. Due to data reliability, maximum age exceeding 105 years will need a double check.

The platform automatically performs centering and scaling: $age\_c = (age - min)/10$. You do not need to pre-center your age variable. Also, you are recommended to provide dataset with age range more than 20 years.

Intervals

Intervals will be used to group the cohort years. You should provide an interval within 2-10 years long.

Fixed effects structure

Once the upload process is successfully done, the fixed effects structure will be automatically filled based on the column names of your dataset. You can:

- Select 1 var and click **"Add Main"** for main terms.
- Select 2 vars and click **"Add Interaction"** for terms like "age:sex".

Random effect structure

The platform allows to specify random slopes for both period and cohort effects. By selecting variables from dropdown lists, you can allow the estimated effect (slope) of one or several specific predictor to vary across different survey years or birth generations.

While this adds significant depth to the analysis, it is important to note that specifying random slopes increases the mathematical complexity of the estimation. Therefore, it should be used judiciously, as models with multiple random slopes may require larger datasets to achieve stable convergence and avoid singular fit errors.

2.2 3D figure

This module provides a dynamic way to explore the distribution of rates across Age, Period, and Cohort dimensions. Please note that this is a descriptive trend figure. You can check this 3D figure without running HAPC-CCREM model.

3D figure view

The 3D surface plot uses age and period as X and Y axes, with the rate mapped to the Z-axis.

- Rotate: Hold and drag the left mouse button.
- Zoom: Scroll the middle mouse button.

2D slice view

The intersection of a geometric plane and the 3D surface generates a 2D slice.

Control panel

The control panel allows you to manipulate the slice.

- Choose the slice dimension (age/period/cohort) and drag the slider to move it.
- Clicking the play button allows automatic slicing.
- Choosing the dimension of "null" will hide the slice and display the 3D surface plot alone.
- Clicking directly on the 3D surface will snap the 2D slice to that location.

2.3 Model results

This module calculates predicted probabilities:

$$Predicted\ Probability = 1/(1 + \exp(-\beta))$$

You can select needed figures through Visual Settings.

- Choose Age, Period, or Cohort as your X-axis.
- Stratified by: Only displays variables specified with an interaction term involving age OR variables assigned as a random slope.

2.4 Export results

This platform allows you to download all the important results of your HAPC model. You can select specific tables from the dropdown or click **"Download All"** to get a .zip file.

- Tables 1-3: Statistical results (HAPC Effect Table, etc.) for manuscript reporting.
- Tables 4-6: Predicted probability trend data for plotting.

3. Problems you might meet

Missing core columns

If your uploaded dataset is missing any of the mandatory variables (disease, age, or period), a warning message will pop up. Please double-check your dataset and re-upload.

Model fitting time

Even with parallel computing enabled, fitting the model takes time. For reference, a dataset with about 100,000 rows typically takes about 20 seconds to process.

You can track the status using the progress bar in the bottom right corner. Please be patient while the app calculates.

Convergence problem

If you receive a "Model convergence problem" warning, it usually means your model is too complex for the amount of data provided.

To fix this, you can either remove some variables from the random slopes to simplify your model, or use a larger dataset with more observations per group.

4. Functions to be updated

We are continuously working to expand the platform's capabilities. The following features are currently on our development roadmap.

4.1 Support of aggregate data

Currently, the HAPC-CCREM framwork requires individual-level microdata to compute cross-classified random effects. It does not support aggregate data (such as regional prevalence rates from the Global Burden of Disease database).

We plan to adapt the mathematical framework to handle aggregated count/rate data using appropriate theoretical models in a future update.

4.2 Hierarchical partitioning

To better interpret complex models, we plan to introduce Hierarchical Partitioning, a method widely used in ecological research for Generalized Linear Mixed Models. This feature will visualize the independent proportion of variance (R^2) explained by each fixed effect in the model. Considering the heavy computation burden of hierarchical partitioning, this function will be selective while running.

4.3 Interactive figure editing

We are developing an interactive figure editing module that will allow users to customize the appearance of their plots directly within the platform. This will include options for adjusting colors, fonts, axis labels, and other stylistic elements to enhance the visual appeal and clarity of the figures.

5. Contact us

Contact **longjunyuhooper@163.com** for assistance.