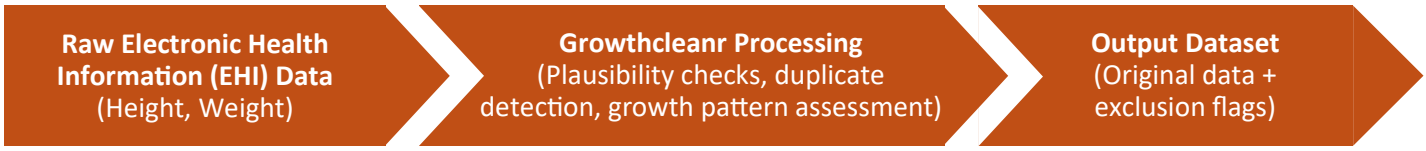
Researcher Guide for Interpreting and Using growthcleanr Output

growthcleanr is designed to identify potentially erroneous height, weight, and head circumference measurements while preserving measurements that reflect unusual but plausible growth patterns. Height and weight measurements are essential for studying health conditions that depend on accurate understanding of growth patterns (e.g., obesity, diabetes, nutrition, child development). Because these measurements are prone to data entry errors, including swapped values, incorrect units, and misplaced decimal points, methods for identifying such errors are critical.

The *growthcleanr* package can be used through the R programming environment or through a user-friendly point-and-click Shiny application. Users can access step-by-step instructions for both approaches in the *growthcleanr* How-To Guides. The *growthcleanr* package returns the user's original data alongside recommendations for values to be excluded, after optional review. The package is open source and produces reproducible results, allowing researchers to apply a standardized, transparent cleaning process to efficiently create analysis-ready datasets. This brief provides guidance for interpreting *growthcleanr* outputs and incorporating them into research workflows.

Understanding *growthcleanr* Output

growthcleanr does not delete measurements. Instead, it evaluates each observation and assigns flags that can be used to support decisions. Exclusion recommendations can be used as-is or customized. Researchers retain full control over which observations are ultimately included in analysis.



Output Elements

Output	Interpretation
Include	Measurement appears plausible and is generally appropriate for inclusion.
Exclude	Measurement may be affected by a data quality issue and should be reviewed.
Exclusion Flags	Information about why a measurement was flagged.
Unit Error Range Indicator	Measurement is in the range of a possible unit error, but may be a real extreme value.
Optional Diagnostics	Additional information to help researchers evaluate algorithm decisions and data quality.

Using *growthcleanr* Output in Research

The output dataset retains all original records and adds information to help researchers identify values that should be excluded. For most studies, the recommended approach is to create a dataset using measurements classified as "Include" while retaining the original dataset to support reproducibility and sensitivity analyses.

Recommended Analytic Practices

- Review exclusion frequencies across study sites, age groups, and populations.
- Examine unexpected patterns before making final exclusion decisions.
- Document any non-default algorithm settings.
- Retain both original and cleaned versions of the dataset.
- Consider performing sensitivity analyses when exclusions substantially affect sample size or study outcomes.

Key Considerations for Using *growthcleanr* Output

Understanding Growth Patterns in Context

A measurement may appear reasonable when viewed in isolation but become implausible when examined within an individual's growth trajectory. Conversely, a measurement outside typical population ranges may be retained when it is consistent with the individual's observed growth pattern.

Consideration: *growthcleanr* evaluates measurements within the context of individual growth trajectories, which may lead to different conclusions than approaches based solely on population-based cutoffs. This approach can be particularly valuable when studying populations with atypical growth patterns or clinical conditions that may not be well represented by population norms.

Effects of Customization Settings

growthcleanr includes configurable settings that can influence how measurements are evaluated, including carried-forward handling and other algorithm options. Default settings are recommended for most users. They have been designed to work for a wide variety of datasets and study needs.

Consideration: More restrictive settings (i.e., flagging more values) may be appropriate for studies of relatively healthy patients. Less restrictive settings may be appropriate for studies of patients with chronic conditions or in whom significant weight change is expected (e.g., patients using weight loss medications).

Interpreting Carried-Forward or Repeated Values

Carried-forward values are measurements that are identical to the previous value and may reflect copied data rather than a new measurement. The child *growthcleanr* algorithm attempts to distinguish true repeated values from copied values and may either exclude or reinstate these observations depending on the available evidence. The adult algorithm can also be configured to give less weight to repeated values when evaluating a patient's weight or height trajectory.

Consideration: Analyses of datasets with many repeated identical values may be particularly sensitive to decisions regarding carried-forward values. For analyses of such datasets, researchers may wish to consider customizing carried-forward and repeated value settings and/or to perform sensitivity analyses under different settings.

Reporting Recommendations

To support transparency and reproducibility, researchers should document the version of *growthcleanr* used, any non-default settings applied, and how flagged observations were handled during data preparation. When data cleaning decisions meaningfully affect sample size or study findings, reporting sensitivity analyses can help demonstrate the robustness of results. Retaining both original and *growthcleanr*-produced datasets can also facilitate reproducibility.

growthcleanr is open source and produces reproducible results.
View and download *growthcleanr* documentation and code here: www.growthcleanr.org.
Resources are also accessible via Comprehensive R Archive Network (CRAN):
https://cran.r-project.org/web/packages/growthcleanr/index.html

This work was funded by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF).
For more information about the OS-PCORTF projects described in this resource, please reach out to OSPCORTF@hhs.gov
Explore the OS-PCORTF project portfolio at aspe.hhs.gov/collaborations-committees-advisory-groups/os-pcortf/explore-portfolio