growthcleanr Factsheet:
A Cleaning Algorithm for Anthropometric Data

growthcleanr is a useful tool for cleaning height, weight, and head circumference data collected in electronic health records or other electronic health information. It runs an algorithm that processes data for people ages 0-65 years, identifies erroneous measurements, and yields ready-to- use datasets that can be used for subsequent analyses, public health surveillance, or other activities.[1]



Who Can Use *growthcleanr*?

Researchers in various settings such as academic institutions, government agencies, and non-profits.

Organizations conducting data-driven health initiatives, quality improvement, or multi-site research, requiring standardized growth data.

Implementers including technical staff, data engineers, and informatics teams working with clinical systems or analytics workflows.

Importance & Benefits of *growthcleanr*

- *growthcleanr* helps prepare height and weight data for secondary uses such as research, quality improvement, and public health surveillance.
- Height and weight measurements are essential for studying health conditions that depend on accurate understanding of growth patterns, such as obesity, diabetes, nutrition, child development, etc. However, these measurements are prone to data entry errors including swapped values, incorrect units, and misplaced decimal points.[2,3]
- *growthcleanr* identifies height and weight measurements that are likely implausible, using each person's current and prior measurements.

Improves Data Quality

growthcleanr determines whether values should be retained or excluded based on plausibility. This includes retaining values outside the normal range if they represent true atypical growth for a given individual.

Reduces Manual Cleaning Effort

growthcleanr flags measurements likely affected by data entry errors and provides a clear recommendation to keep or exclude each value. What previously required months of chart review can often be completed in minutes or hours through automation.

Supports Patient-Centered Outcomes Research

growthcleanr uses patient-specific longitudinal analysis to ensure growth patterns reflect realistic individual trajectories, supporting real-world outcomes research. Cleaning the data reduces bias and improves the validity of analyses.

Common Use Cases for growthcleanr

Upload data using the *growthcleanr* R package

Code-based workflow
Run code through R software or a command line interface

Point-and-click workflow
No R coding knowledge required

Upload data using the *growthcleanr* Shiny app

growthcleanr first prepares your data for analysis

- Accepts long, semi-wide, or wide data
- Maps user column names and coding schemes
- Calculates age from dates (optional)

The wrapper processing then uses age range to send formatted data to child or adult functions

Some steps in the child and adult algorithms address similar questions but use different methods aligning with needs of that age group. Other steps are unique to either the child or adult algorithm.

Child

Could premature birth affect how this measurement should be interpreted?

Accounts for premature birth and development

Is this a true repeat measurement or a copied value?

Evaluates whether repeated values are likely true values or values copied forward from earlier visits

Does this head circumference measurement fit the child's growth pattern?

Evaluates head circumference measurements to exclude implausible measurements

Overlapping Functions

Is this height and/or weight plausible regardless of the person's other measurements?

Excludes very extreme values before longitudinal evaluation

Which value should we use for this day?

When there are multiple values on the same day, selects the value that best fits the person's other values

Does this measurement fit with nearby measurement values?

Detects implausible patterns, accounting for expected differences in growth pattern with age

Are this person's measurements reliable?

Identifies people for whom a high proportion of values are excluded

Adult

What is most likely this person's true height at this time?

Provides a stable height value for analysis between periods of height gain or loss

The flowchart presents examples and highlights key functionality and is not a complete representation of every step in the algorithms.

growthcleanr's Functionalities

The *growthcleanr* package comprises separate **Child (0-20y)** and **Adult (18-65y)** algorithms to efficiently clean height and weight data while preserving realistic growth patterns across children and adults. While *growthcleanr* provides a standardized, evidence-based data cleaning framework, it also allows investigators to configure key parameters for the needs of a specific project. Researchers can adjust thresholds, error-handling strategies, reference standards, and inclusion criteria to reflect their study population, data quality, and analytic objectives, while maintaining a transparent and reproducible workflow.



Examples of *growthcleanr's* Functionalities

Subject ID	Age (Years)	Sex	Weight (kg)	*growthcleanr* Result
P1	5.8	M	25.84	Include
P1	6.6	M	28.61	Include
P1	7.4	M	28.61	Exclude

growthcleanr recommends excluding this repeated weight value that was likely **carried forward from the previous measurement**.

Subject ID	Age (Years)	Sex	Height (cm)	*growthcleanr* Result
P2	40.1	F	165	Include
P2	40.2	F	62	Exclude
P2	40.3	F	170	Include

growthcleanr recommends excluding this measurement, which **likely represents a mis-recorded weight.**

Subject ID	Age (Months)	Sex	Weight (kg)	*growthcleanr* Result
P3	Birth (0)	M	1.14	Include
P3	6	M	5.8	Include
P3	9	M	7.3	Include
P3	12	M	10.5	Exclude
P3	24	M	10.8	Include

Despite the weight falling outside typical age-based values, *growthcleanr* recommends including this measurement because it is a **biologically plausible growth pattern for a premature infant**.

Conversely, despite falling within typical age-based values, *growthcleanr* recommends excluding this measurement due to an **implausibly rapid increase in weight.**

Get Started with *growthcleanr*

growthcleanr is open source and reproducible. **It is available as: 1) an R package for use in the R interface or 2) a Shiny application for a point-and-click interface.**

- **R interface: Comprehensive R Archive Network (CRAN)**: install.packages("growthcleanr")
- **Shiny application:** Access application and *growthcleanr* documentation here: www.growthcleanr.org.

[1] Daymont C, Grundmeier R, Miller J, Campos D, et al. *growthcleanr*. https://growthcleanr.org/
[2] Growthcleanr –Public Health Informatics Institute. https://phii.org/module-6/growthcleanr-4/
[3] Daymont C, Ross ME, Russell, et al. Automated identification of implausible values in growth data from pediatric EHRs. JAMIA. 2017;24(6):1080-1087. doi:10.1093/jamia/ocx037