

GoKinD Baseline Data Set

Format: The data set is stored in a SAS library under the name “Baseline”.

Structure: One record per patient (2866 in total, with 1894 probands and 972 parents) containing baseline covariates from a number of different evaluations and reporting forms. The conventions used to construct variable names are explained on the following pages.

Contents: A complete SAS contents listing is attached as Appendix A.

Summary of distributional statistics: Appendix B is organized in two parts, one for probands, and another one for their participating parents. In this summary, it provides the frequency distributions of all categorical variables, including treatment group and baseline retinopathy stratum, and descriptive statistics summarizing the distributions of all quantitative variables.

GoKinD Data Sets: Naming Conventions and Anomalies

1. **Variable Naming Conventions:** Raw data forms received at the Coordinating Center (CoC) are converted directly to SAS data sets. All the baseline variables are abbreviations, eg. CREATN is serum creatinine.
2. **Data Range and Missing Data:** Data is generally available for all measures for all patients. The statistical summary provided in Appendix B shows some instances where this is not the case. For example, waist circumference was not measured for Joslin clinic. Every effort has been made to obtain complete data on all subjects. The statistical summary also gives the range of values for each variable.
3. **ACR(Urine Albumin/Creatinine Ratio)/AER(albumin excretion rate):** If the patient has end stage renal disease (dialysis, kidney transplant), ACR will not be measured. If the ACR values of the first two tests are not both greater than 300 ug albumin/mg of urine creatinine, or less than 20 ug albumin/mg of urine creatinine, the third test is needed. Albumin excretion rate instead of ACR were measured for 27 patients.
4. **Calculation of GFR:** The calculations of average GFR are based on two methods. One is the MDRD GFR and another is the Cockcroft-Gault GFR.

MDRD Equation: $Race \cdot Sex \cdot 186.3 \cdot CREATN^{-1.154} \cdot AGE^{-0.203}$

If female, sex=0.742; if male, sex=1; If black, then race=1.21, else race=1;

Cockcroft-Gault Equation: $\frac{(140 - AGE) \cdot WEIGHT(kg)}{72 \cdot CREATININE(mg / dl)}$

If female, sex=0.85, if male, sex=1.

5. **BMI:** Body mass index is calculated as $\frac{\text{Weight in Kilograms}}{(\text{Height in Meters}) \times (\text{Height in Meters})}$
6. **Abnormal Creatinine:** Abnormal Creatinine is defined as serum creatinine greater or equal to 1.3 mg/dl for female and greater or equal to 1.5 mg/dl for male.
7. **Relative Code:** Parents are related to probands using relative code. The same first seven digits indicate that these patients are from a same family. The last two digits are composed as 0X. If the last digit reveals the family status of the patients: 1:Father, 2:Mother, 3:Probands, 4:Sibling.
8. **Protocol and Manual of Operations:** These two documents, describing the GoKinD in detail are available.

Privacy Considerations

In order to preserve the confidentiality of data all information that could lead to the direct identification of individual patients has been deleted. These variables include the original GoKinD study identification numbers, the GoKinD clinic number, patient initials and date of birth.

Individual study patients are identified in data set by a masked ID number (Mask_ID). The mask ID contains the information about eligibility status, trio status and family status, but conceals the original ID, which reveals the site identification. The study-coordinating center holds the key to this masked ID.