

SAMBA

(Statistical Apparatus for Micronutrient Biomarker Analysis)

User Manual

SAMBA development team

Introduction to the SAMBA R package (I)

- SAMBA:
 - **Statistical Apparatus**
for Micronutrient Biomarker Analysis
- SAMBA is an easy-to-use and **“all-in-one” R** package
 - R is a free and powerful software for statistical computing and graphics.
No additional costs on software purchase required;
- Co-developed by
 - UC Davis/BRINDA: Hanqi Luo;
 - GAIN: Ty Beal;
 - BRINDA: Yaw Addo, Madeleine Zeiler, Jiaxi Geng, E Rochelle Werner, Parminder Suchdev, Melissa Young;
 - MAPS: Tineka Blake
- With support from USAID Advancing Nutrition



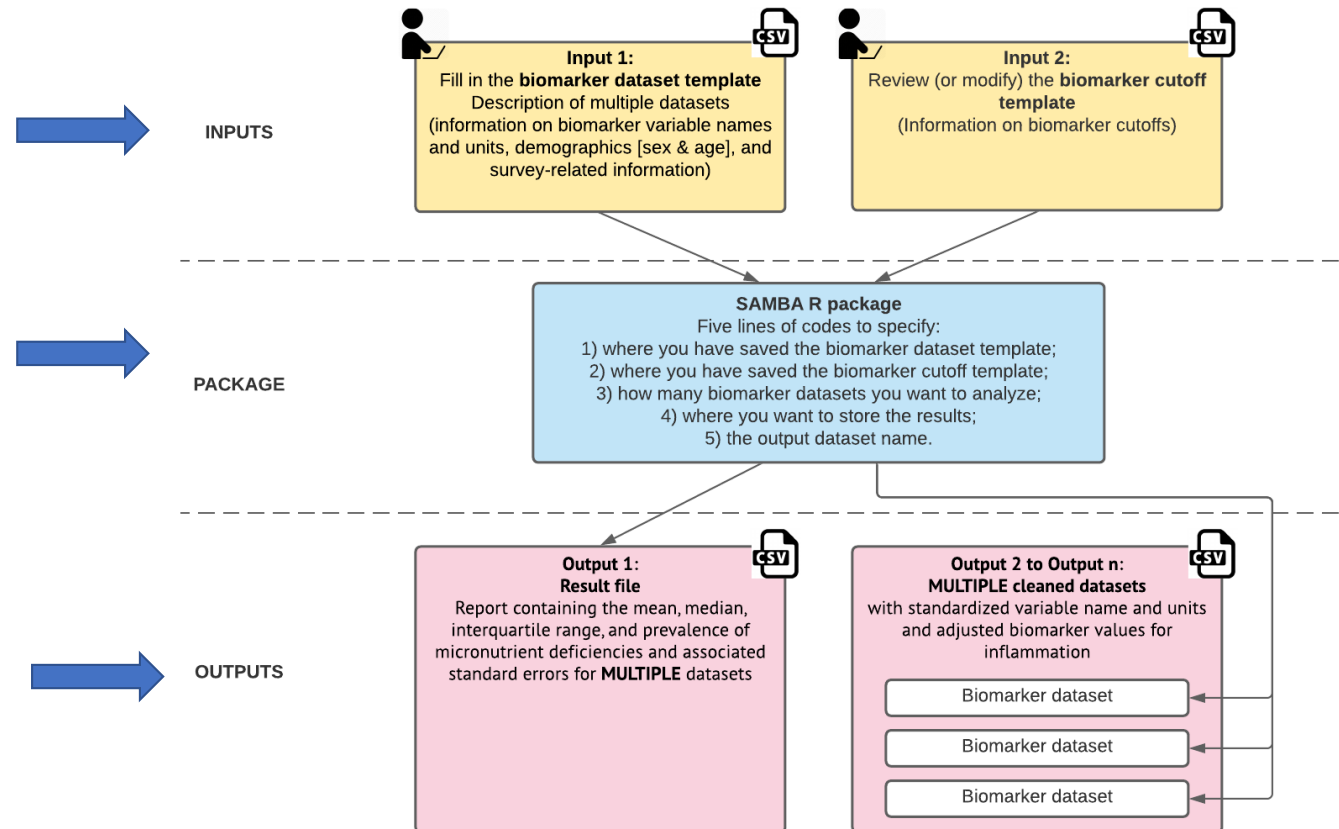
Introduction to the SAMBA R package (II)

- An “all-in-one” R package
 - Can estimate **prevalence** of any measured micronutrient deficiency and the associated standard errors/confidence intervals.
 - Can estimate **mean, geometric means, and percentiles** of multiple biomarkers and the associated standard errors/confidence intervals;
 - Can be used to analyze multiple datasets simultaneously;
 - Can provide both unadjusted and adjusted biomarker values for inflammation (BRINDA method);
 - Can be used to analyze both single-site studies and studies/surveys with complex survey designs;
 - Can provide a summary of results and a clean dataset with standardized variable names and units.

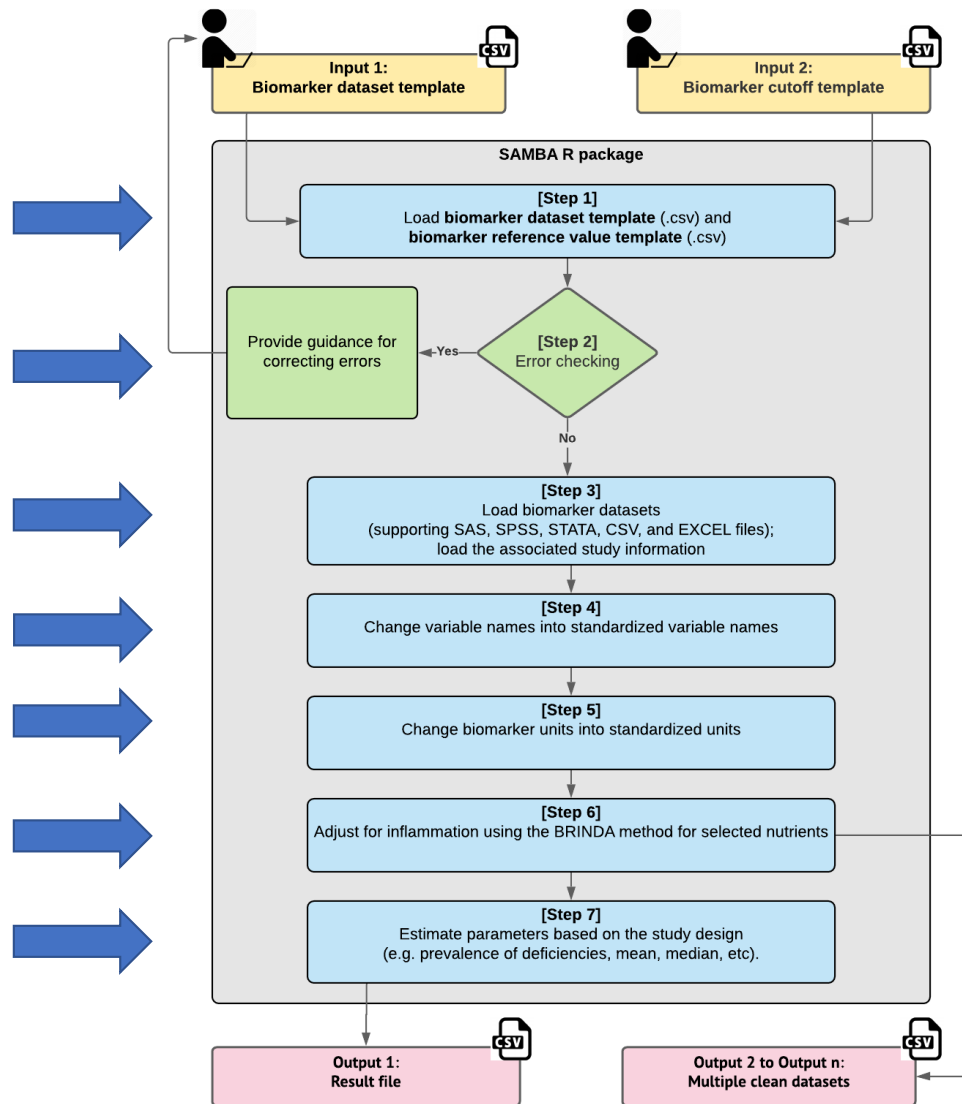
Introduction to the SAMBA R package (III)

- Flexibility (without modifying the R package)
 - Users can modify the cutoff values for micronutrient deficiencies;
 - Users can use the same package to analyze additional variables (such as HDL, cholesterol, & blood pressure).

Analysis procedure using SAMBA



Unpacked the Blackbox of SAMBA



Step 1: Fill in the biomarker dataset template

Study information	One dataset per column		
A	B	C	D
information	BioData1	BioData2	BioData3
country	United States	United States	Kenya
country_code	USA	USA	KEN
survey_year	2003-2016	2003-2016	2010
complex_survey_design_yes_no	YES	YES	NO
study_design	multi-stage cluster design	multi-stage cluster design	single-site
data	nhanes_with_diet.csv	nhanes_with_diet.csv	kenya_data.sas7bdat
directory	~/git/2billion/demo/bio_analysis/raw_data	~/git/2billion/demo/bio_ana	~/git/2billion/demo/bio
id	seqn	seqn	brinda_id
strata	sdmvstra	sdmvstra	
PSU	sdmvpsu	sdmvpsu	
PSU_ID_nested_in_strata ID_yes_no	yes	yes	
sex_var	riagendr	riagendr	sex
male_value	1	1	1
age	ridageyr	ridageyr	age_mo
age_unit	year	year	month
pregnancy_var	ridexprg	ridexprg	
pregnancy_value	1	1	
lactating_var	rhq200	rhq200	
lactating_value	1	1	
covariates	ridreth1 indfmin2 dmdeduc2 dmdhhsiz	ridreth1 indfmin2 dmdeduc2 dmdhhsiz	
subgroup	ridreth1	age_sex_sub2	age_group
retinol_binding_protein			rbp
retinol_binding_protein_unit			umol/l
retinol_binding_protein_survey_weight			
serum_retinol	lbdviasi	lbdviasi	
serum_retinol_unit	umol/l	umol/l	
serum_retinol_survey_weight	retinol_weight	retinol_weight	

If you need results by different subgroups, fill in information of the same dataset again except specifying different subgroup variables

Three pieces of information per biomarker:
 1) Variable name;
 2) Unit;
 3) Survey weight variables (leave blank for single-site studies)

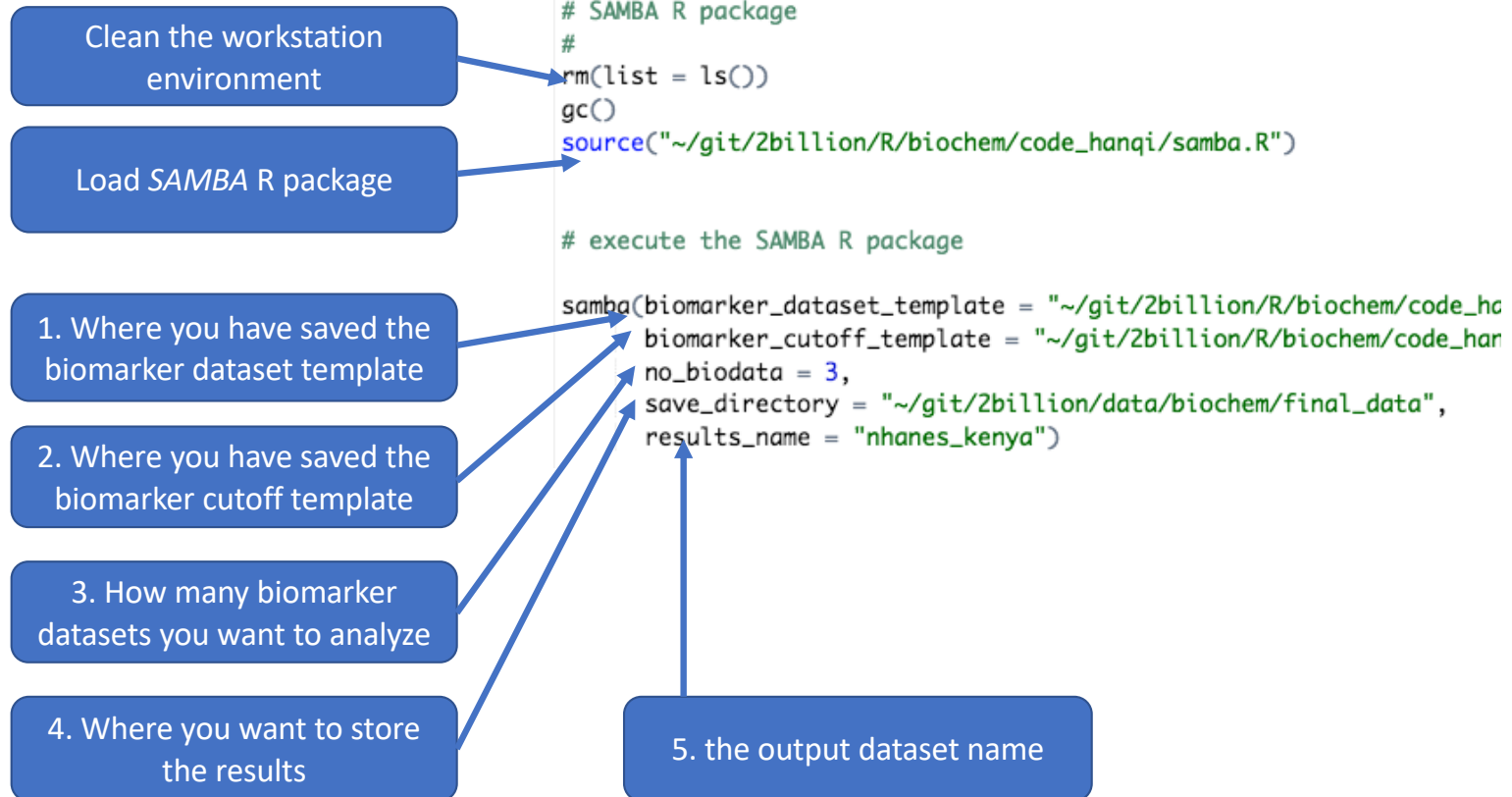
Step 2: review (or modify) the cutoff value template

Users can define a micronutrient deficiency being smaller or greater than a user-specified cutoff

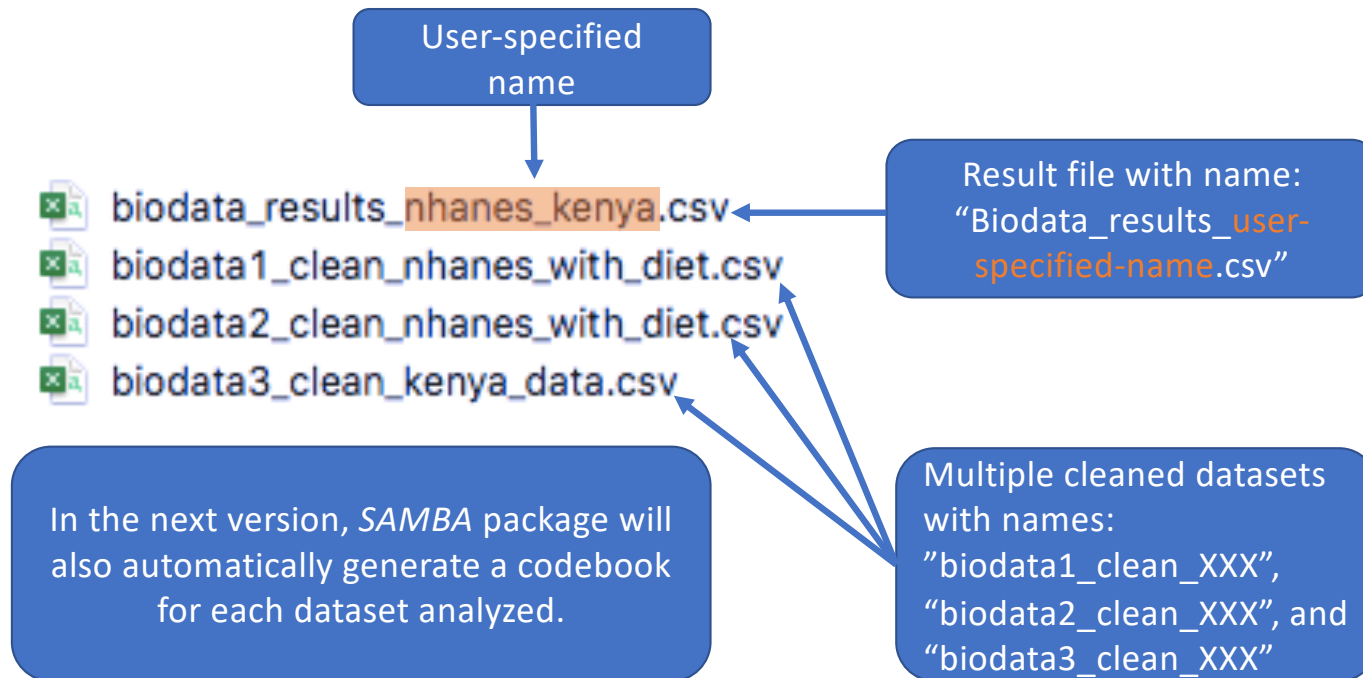
Cutoff values vary by population (male, female, and pregnant women) and age

1	nutrients	biomarker	relationship	cutoff	unit	population	lower_age	upper_age
22	iron	serum_ferritin	<	12	µg/l	male	0	4
23	iron	serum_ferritin	<	12	µg/l	female	0	4
24	iron	serum_ferritin	<	15	µg/l	male	5	120
25	iron	serum_ferritin	<	15	µg/l	female	5	120
26	iron	serum_ferritin	<	15	µg/l	pregnant women	13	49
27	iron	transferrin_receptor	>	8.3	mg/L	male	0	4
28	iron	transferrin_receptor	>	8.3	mg/L	female	0	4
29	iron	transferrin_receptor	>	8.3	mg/L	male	5	120
30	iron	transferrin_receptor	>	8.3	mg/L	female	5	120
31	iron	transferrin_receptor	>	8.3	mg/L	pregnant women	13	49

Step 3: Execute the SAMBA R package



Step 4: Check result datasets



Output I: result file

country_code	country	survey_year	study_design	nutrient	unit	subgroup_name	subgroup	n	deficiency	deficiency_se	mean	mean_se
KEN	Kenya	2010	single-site study	retinol_binding_protein	umol/l	overall	overall	843	29.54	1.57	0.86	0.01
KEN	Kenya	2010	single-site study	retinol_binding_protein	umol/l	age_group	1	457	29.54	2.14	0.86	0.01
KEN	Kenya	2010	single-site study	retinol_binding_protein	umol/l	age_group	2	386	29.53	2.32	0.86	0.01
KEN	Kenya	2010	single-site study	retinol_binding_protein_adj	umol/l	overall	overall	843	9.02	0.99	1.1	0.01
KEN	Kenya	2010	single-site study	retinol_binding_protein_adj	umol/l	age_group	1	457	7.66	1.24	1.1	0.01
KEN	Kenya	2010	single-site study	retinol_binding_protein_adj	umol/l	age_group	2	386	10.62	1.57	1.11	0.02
KEN	Kenya	2010	single-site study	serum_ferritin	ug/l	overall	overall	843	19.22	1.36	57.77	2.22
KEN	Kenya	2010	single-site study	serum_ferritin	ug/l	age_group	1	457	23.19	1.98	52.91	2.89
KEN	Kenya	2010	single-site study	serum_ferritin	ug/l	age_group	2	386	14.51	1.79	63.51	3.41
KEN	Kenya	2010	single-site study	serum_ferritin_adj	ug/l	overall	overall	843	50.42	1.72	15.93	0.49
KEN	Kenya	2010	single-site study	serum_ferritin_adj	ug/l	age_group	1	457	54.92	2.33	14.93	0.66
KEN	Kenya	2010	single-site study	serum_ferritin_adj	ug/l	age_group	2	386	45.08	2.53	17.11	0.74


The result file also provides

- Estimated mean, prevalence of deficiencies, 10th, 25th, 50th, 75th, and 90th percentiles and the standard errors and confidence intervals;
- Estimated geometric means and the confidence intervals.

Due to limited space, results from NHANES analysis and the estimates of percentiles are not shown here.

Output 2 to Output n: cleaned datasets

Cleaned datasets with standardized variable names and units and adjusted biomarker values for inflammation for each study subject



retinol_binding_protein	serum_ferritin	transferrin_receptor	crp	agp	retinol_bindi	serum_ferritin	transferrin_r	crp_unit	agp_unit	male	np_female
0.74	23.2	32.68	3.89	1.4	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.61	118.4	14.42	22.49	2.02	umol/l	ug/l	mg/l	mg/l	g/l	1	0
1	7.01	15.91	3.09	1.05	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.89	38.12	7.22	1.12	1.58	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.78	21.39	25.2	0.36	0.69	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.79	6.59	20.03	1.05	1.31	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.59	172.35	14.95	40	1.67	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.92	9.4	14.65	0.58	1.14	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.65	1.71	21.45	0.2	0.51	umol/l	ug/l	mg/l	mg/l	g/l	1	0
1.14	56.32	11.94	0.79	1.16	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.89	149.95	18.25	22.8	1.55	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.93	7.4	12.92	0.05	0.71	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.81	38.29	7.68	1.25	1.06	umol/l	ug/l	mg/l	mg/l	g/l	1	0
1	7.79	17.35	2.83	1.2	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.85	120.34	6.52	10.64	1.01	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.82	26.29	10.73	4.05	1.04	umol/l	ug/l	mg/l	mg/l	g/l	1	0
1.02	19.75	8.74	0.2	0.63	umol/l	ug/l	mg/l	mg/l	g/l	1	0
0.4	16.56	12.39	27.21	1.34	umol/l	ug/l	mg/l	mg/l	g/l	1	0