Table 1: Dataset Composition and Coverage

Component	N (%)	Data Type	Coverage
Total Participants	18,205 (100.0%)	Combined dataset	Johannesburg, SA
Clinical Trial Participants	9,103 (50.0%)	Biomarkers + Demographics	Facility-based
Survey Participants	9,102 (49.97%)	Socioeconomic + Demographics	Community-based
Biomarker Data Availability (Clinical Cohort)			
CD4 cell count	1,283 (14.1%)	cells/µL	HIV studies
Fasting glucose	2,731 (30.0%)	mg/dL	Metabolic studies
Blood pressure	4,957 (54.5%)	mmHg (systolic/diastolic)	All clinical trials
Cholesterol panel	2,502 (27.5%)	mg/dL (LDL/HDL/Total)	Metabolic studies
Hemoglobin	1,282 (14.1%)	g/dL	HIV/TB studies
Creatinine	1,251 (13.7%)	mg/dL	Renal function
Any biomarker data	6,244 (68.6%)	Multiple measures	Clinical subset
Spatial Coverage			
Participants with coordinates	18,205 (100.0%)	Latitude/longitude	GPS precision
Participants with ward info	9,102 (50.0%)	Administrative boundaries	Survey cohort only
Unique wards covered	55 wards	City of Johannesburg	Urban coverage
Temporal Coverage			
Study period	2002-2021	19 years	All seasons
Climate lag windows	0-21 days	Daily temperature	ECMWF ERA5
Study Type Distribution			
HIV treatment trials	5,847 (32.1%)	Randomized controlled	Facility-based
Metabolic studies	3,256 (17.9%)	Cross-sectional	Clinic recruitment
Population surveys	9,102 (50.0%)	Household surveys	Community-based

Notes:
• Clinical trial data includes biomarkers but limited socioeconomic variables
• Survey data includes comprehensive socioeconomic data but no biomarkers
• All participants have climate exposure data matched by coordinates and date
• Temperature data from ECMWF ERA5 reanalysis with 0.25° spatial resolution