Applying data synthesis methods to turn data you have into data you need

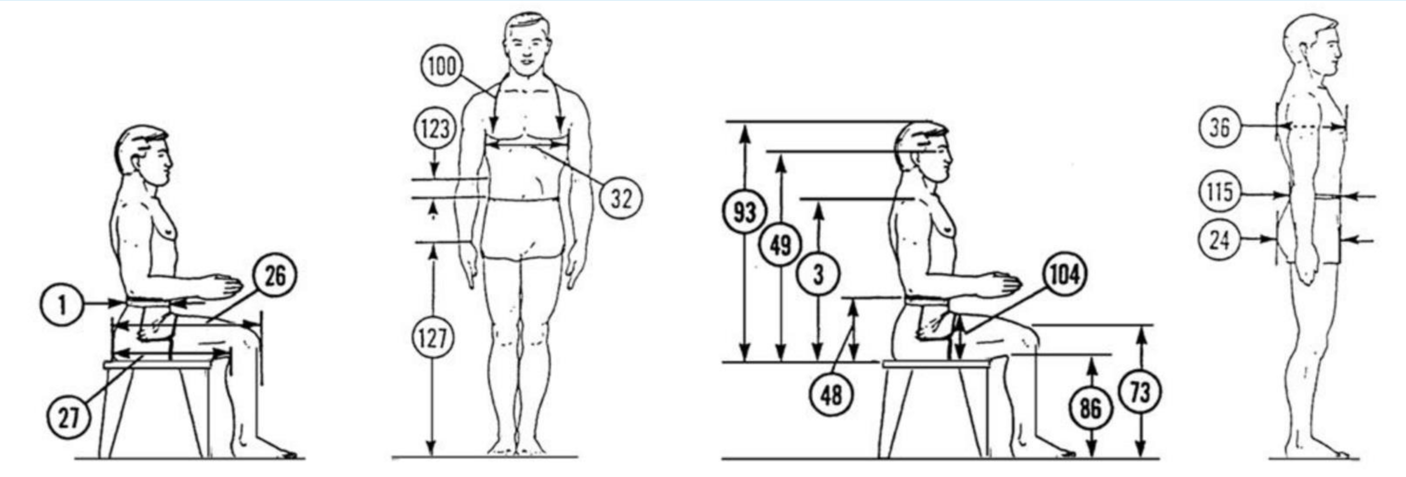
Zoe Marazita

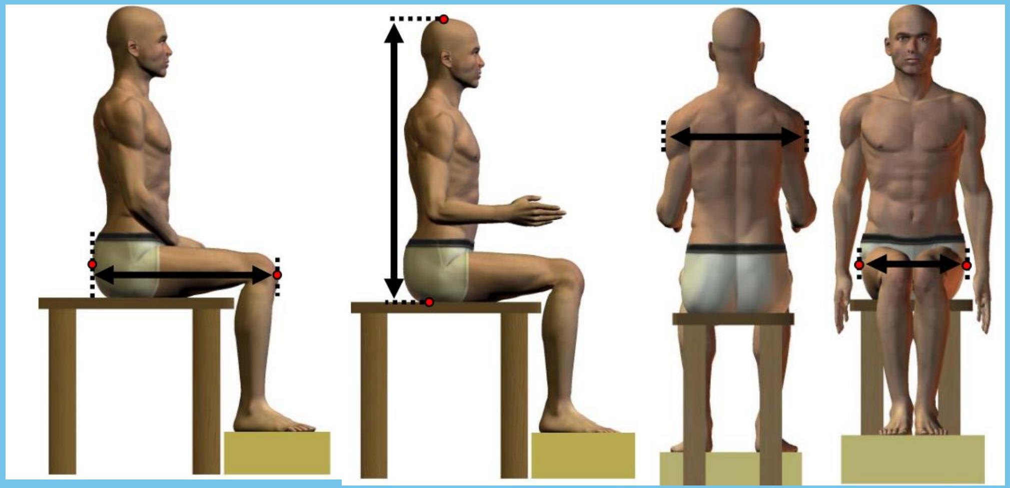
Matt Parkinson

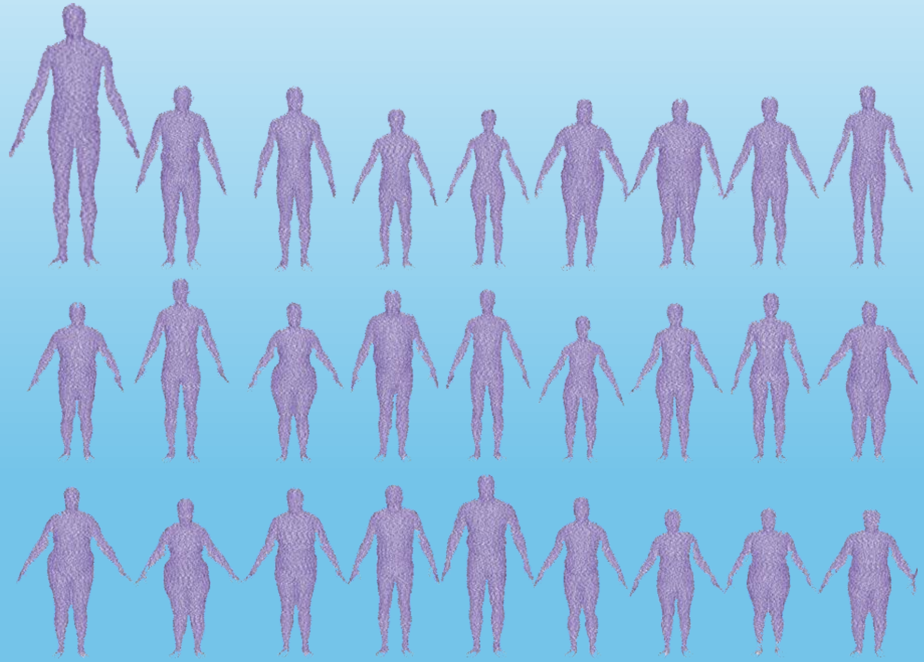
open design lab

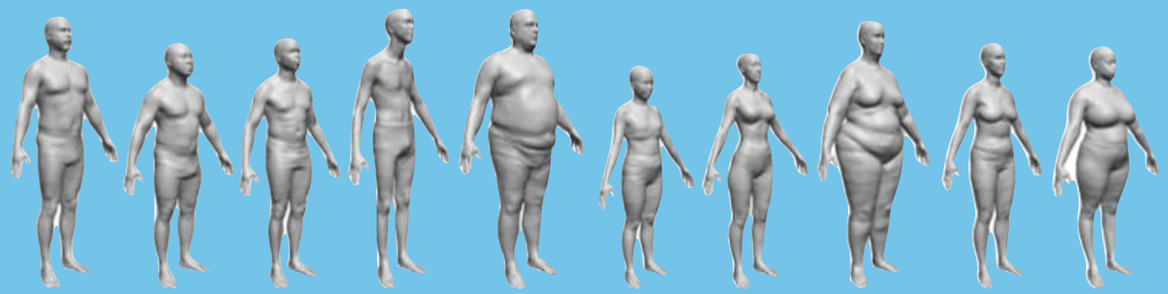
**We are submitting a journal paper on this topic shortly, and it will be sent along when published*

Anthropometric data that are both comprehensive and representative of a target population are rare, making data synthesis often necessary for design evaluations.

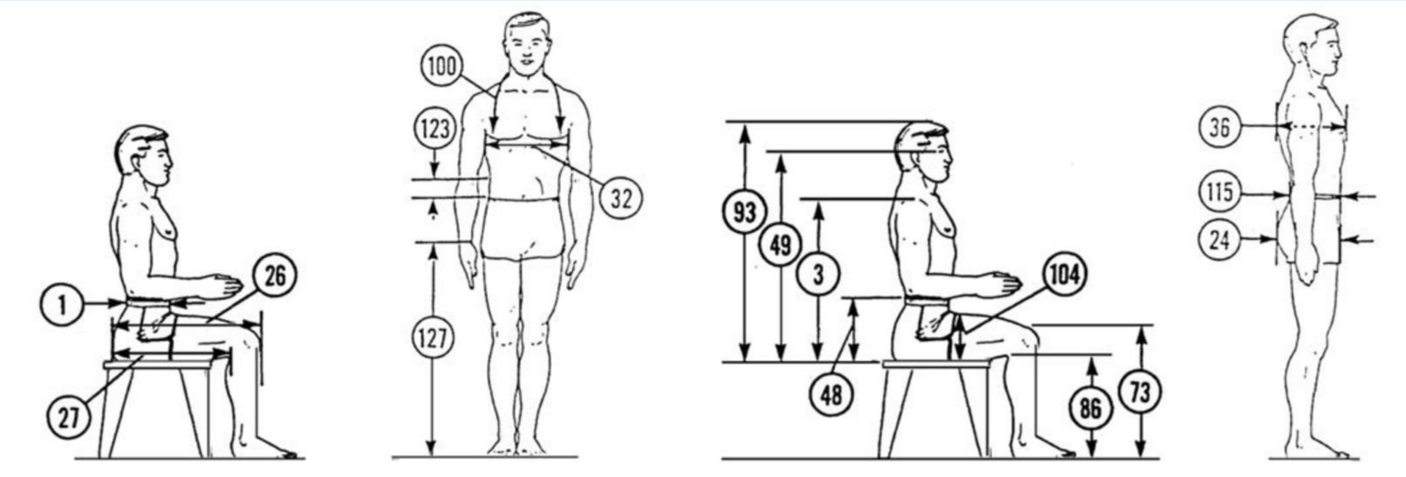


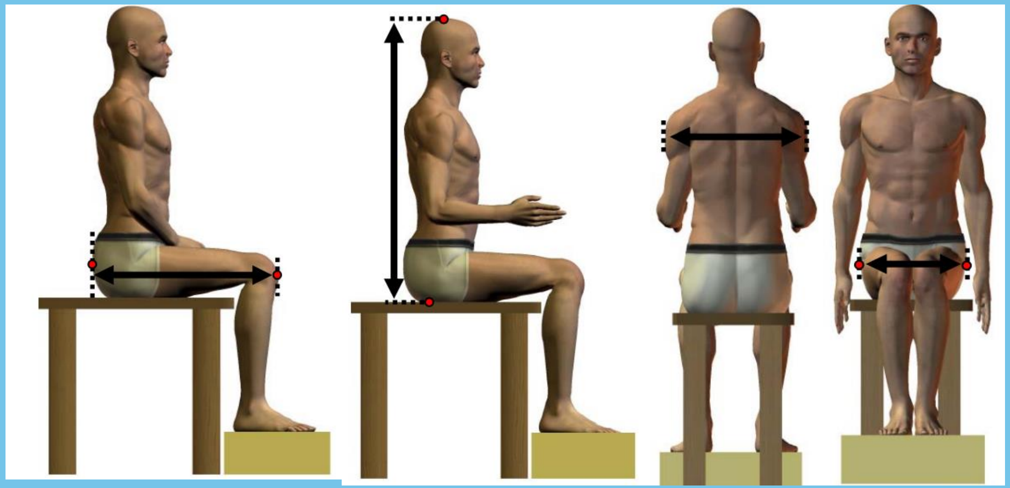


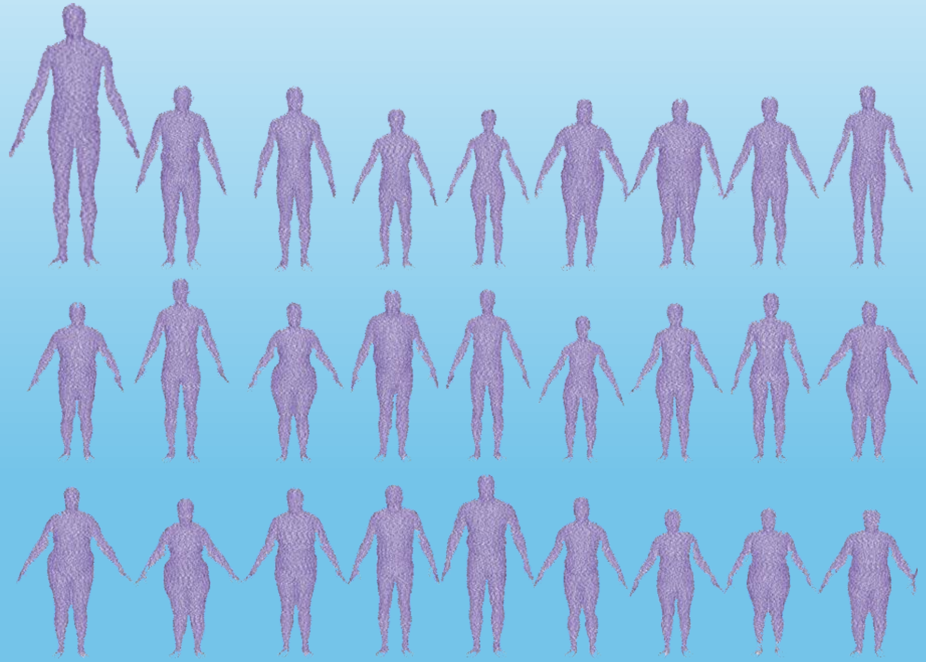


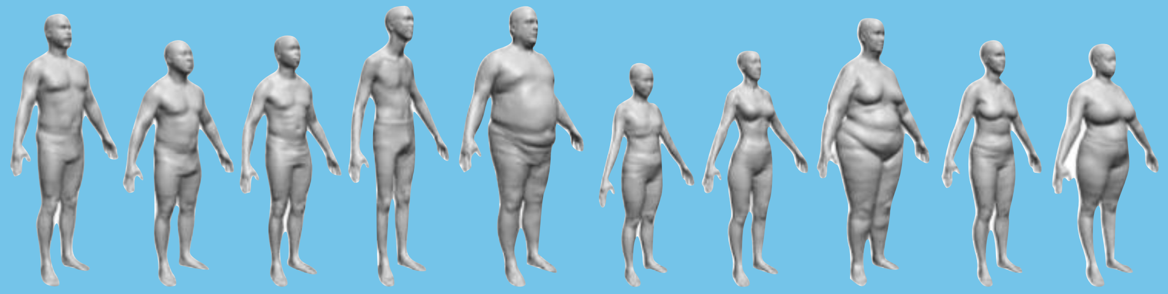


While several data synthesis methods have been developed, there are limited guidelines to help designers select a method based on their design context.



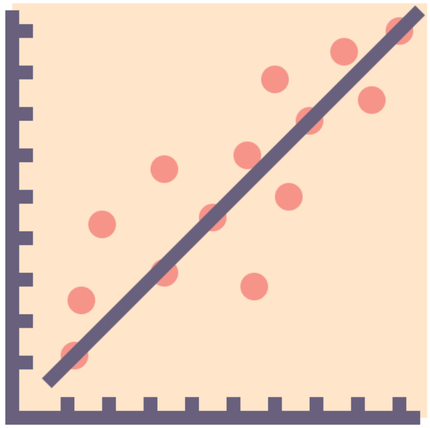






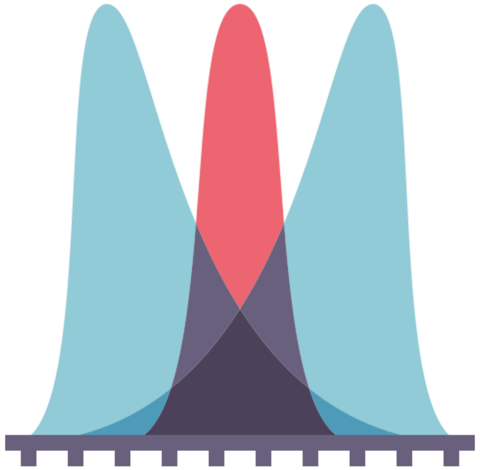
Data synthesis can be broadly categorized into three categories.

Regression-based methods



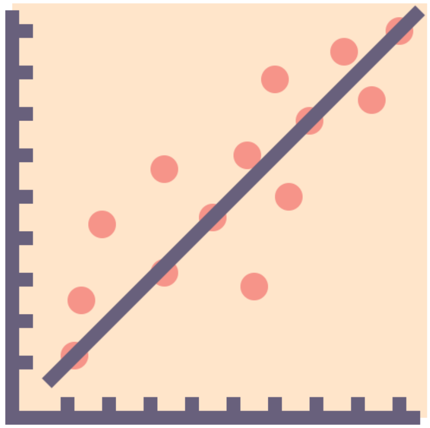
Population reweighting methods

Distribution-based methods



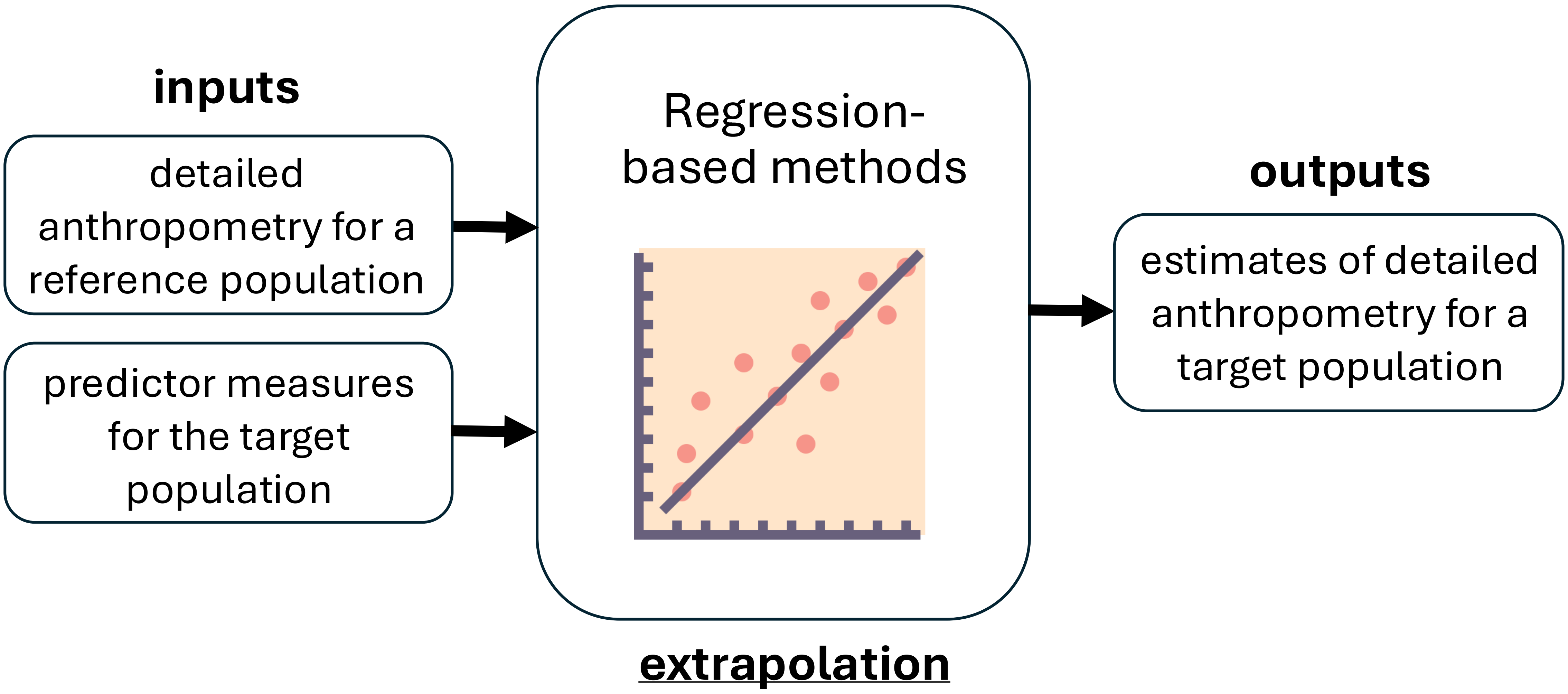
Data synthesis can be broadly categorized into three categories.

Regression-based methods

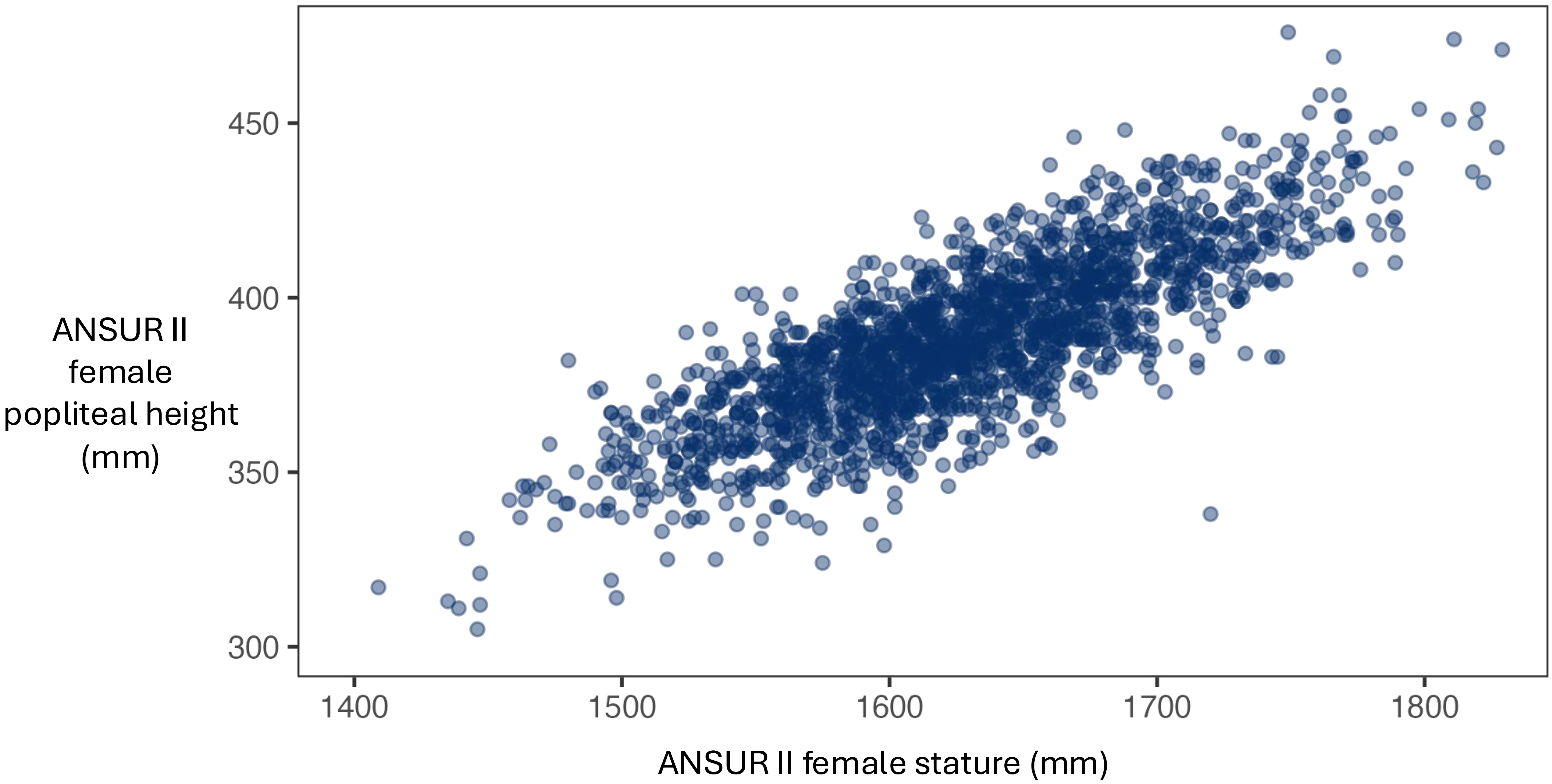


Population reweighting methods

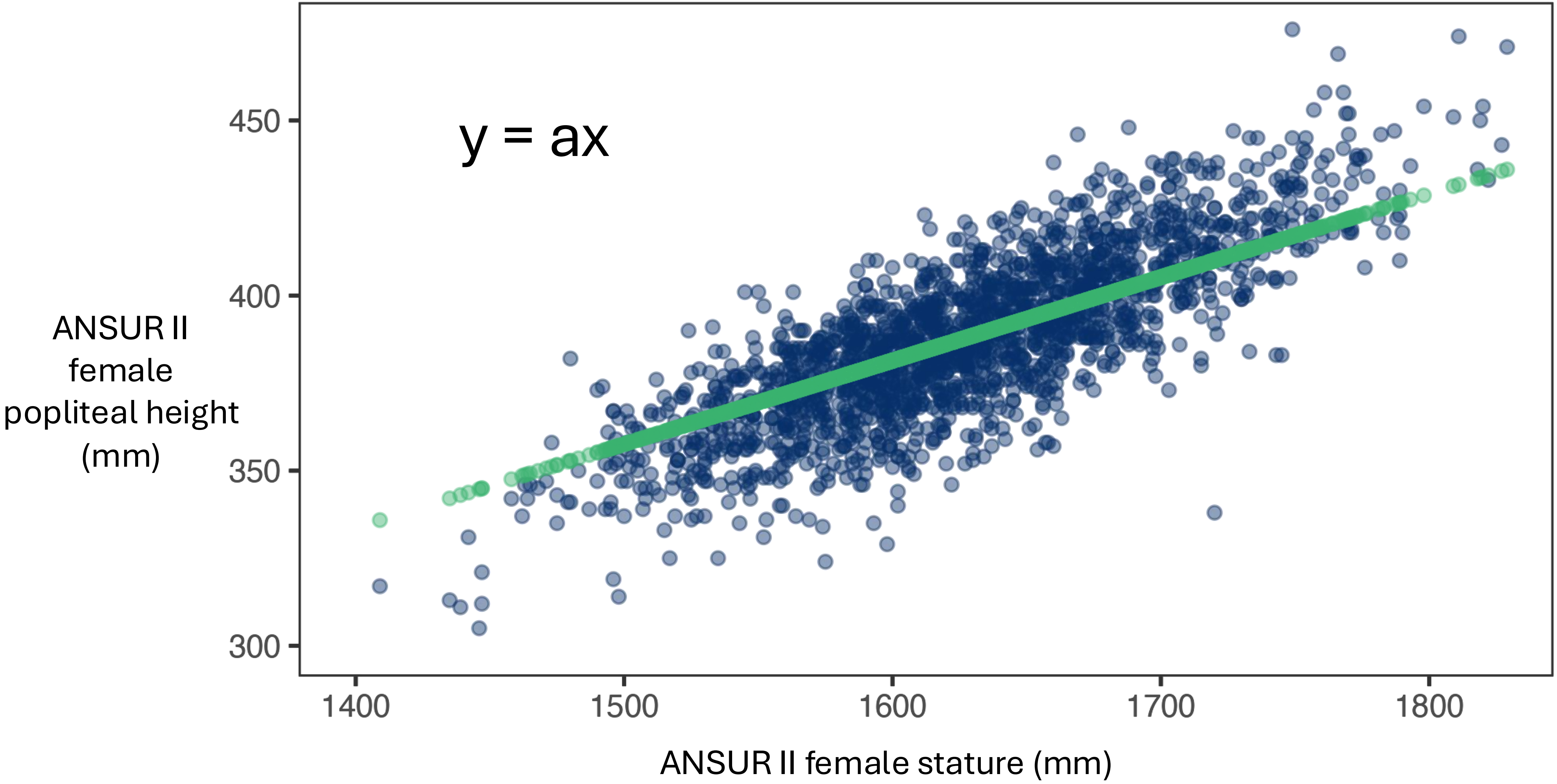
Distribution-based methods


inputs
detailed anthropometry for a reference population
predictor measures for the target population
Regression-based methods
extrapolation
outputs
estimates of detailed anthropometry for a target population

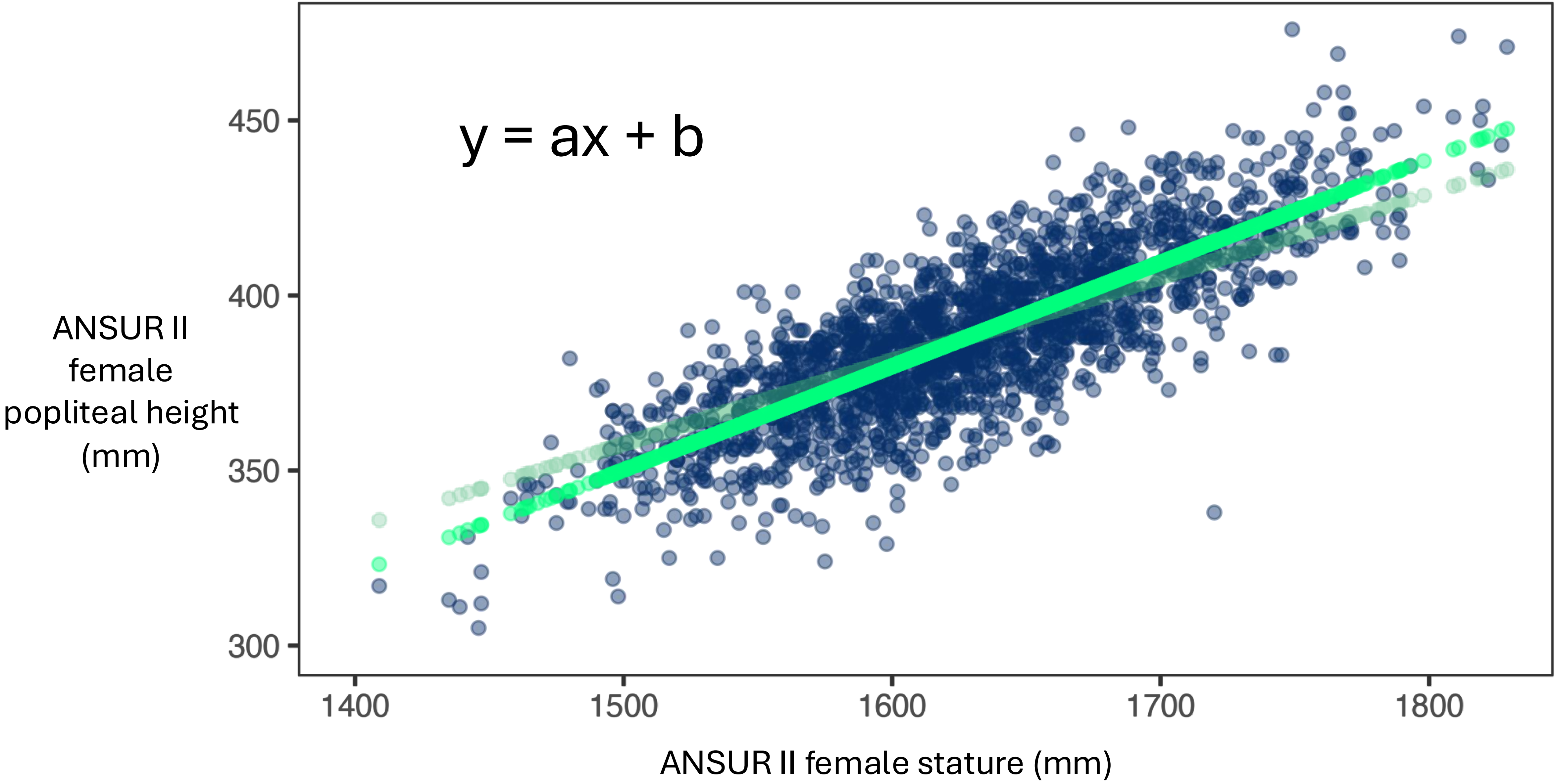
Regression-based methods



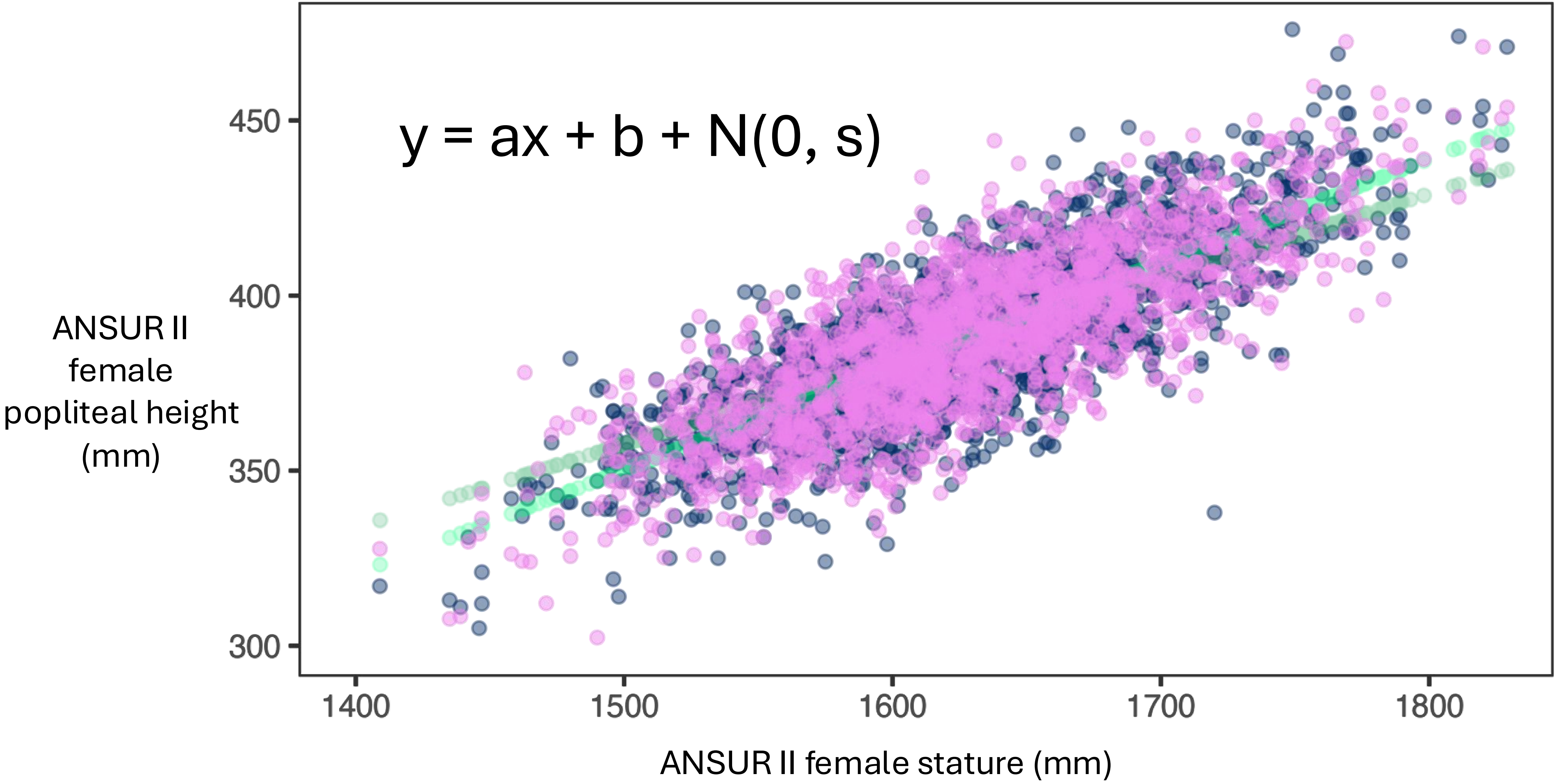
Regression-based methods: Proportionality constant



Regression-based methods: Regression



Regression-based methods: Regression with residual variance

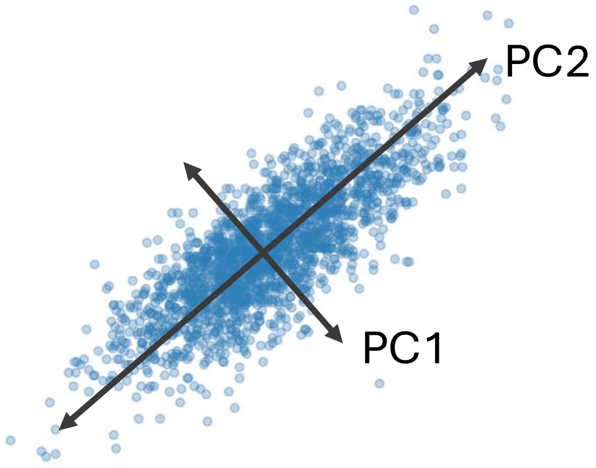


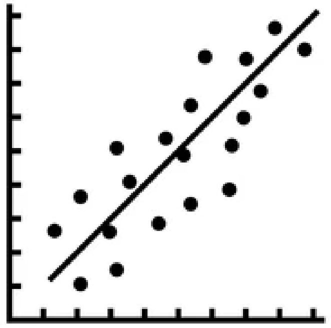
Principal components analysis and regression (PCAR) extends regression to a multivariate space by modeling correlated variation between multiple measures.

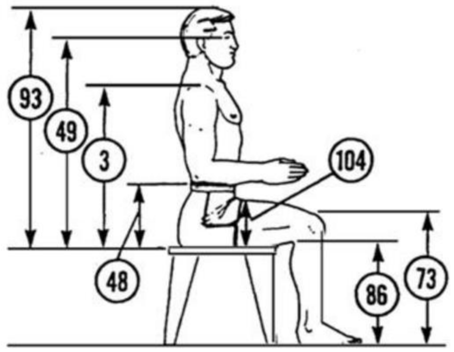
anthropometric variables in an orthogonal basis (PCA)

predict component scores from overall body dimensions (linear regression)

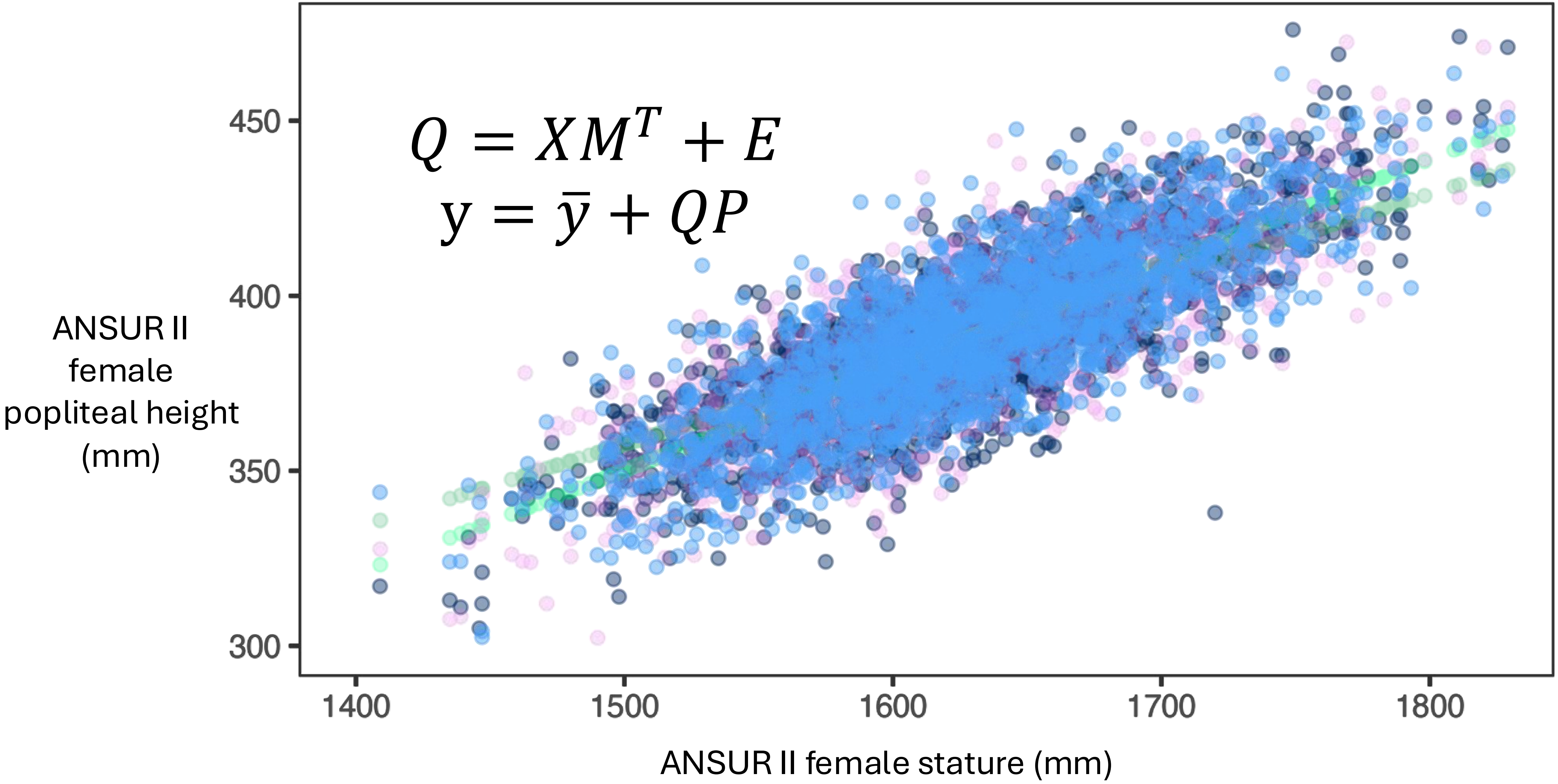
invert PC transformation to estimate a complete set of target population measures







Regression-based methods: Principal components analysis and regression

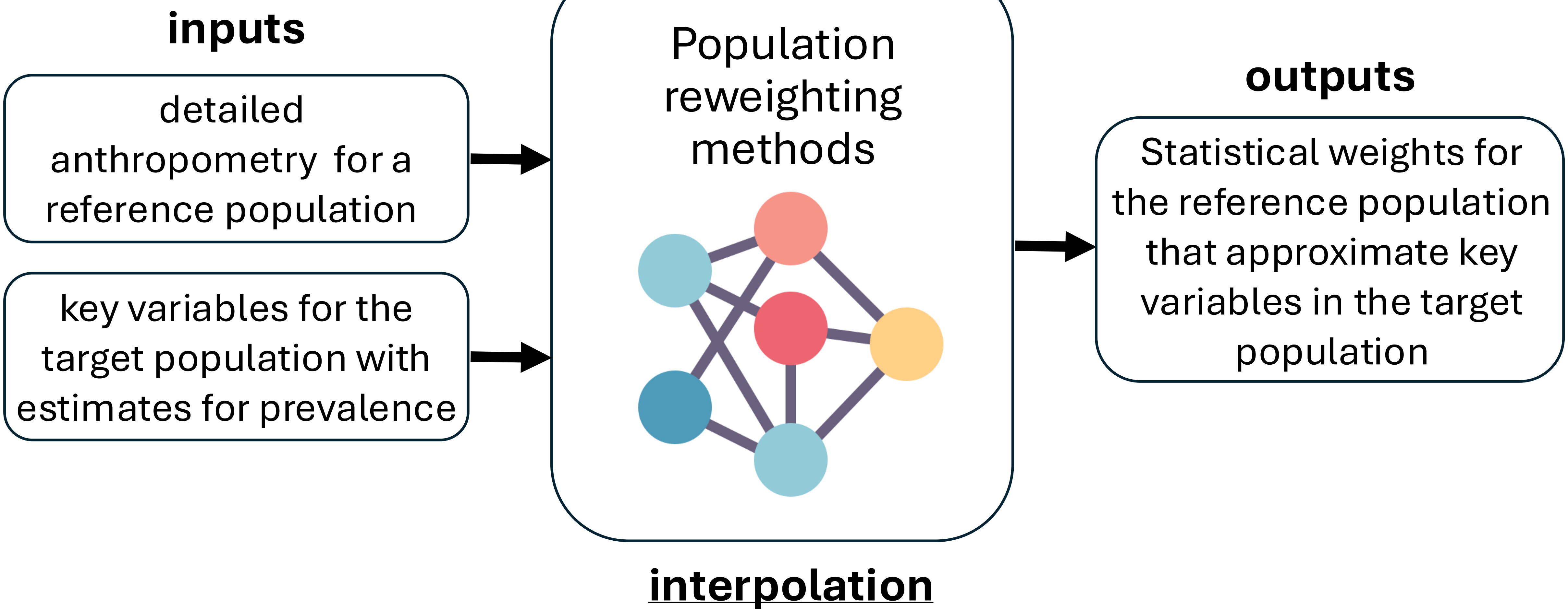


Data synthesis can be broadly categorized into three categories.

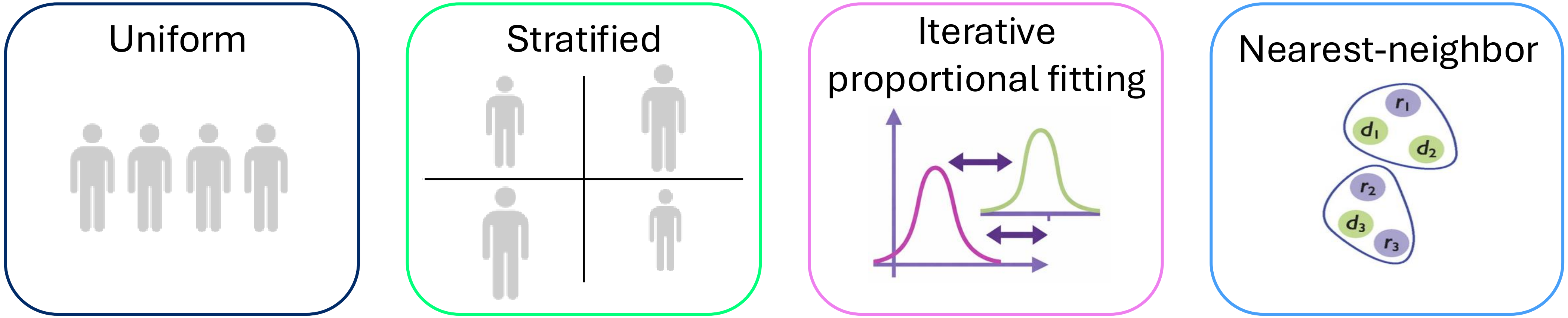
Regression-based methods

Population reweighting methods

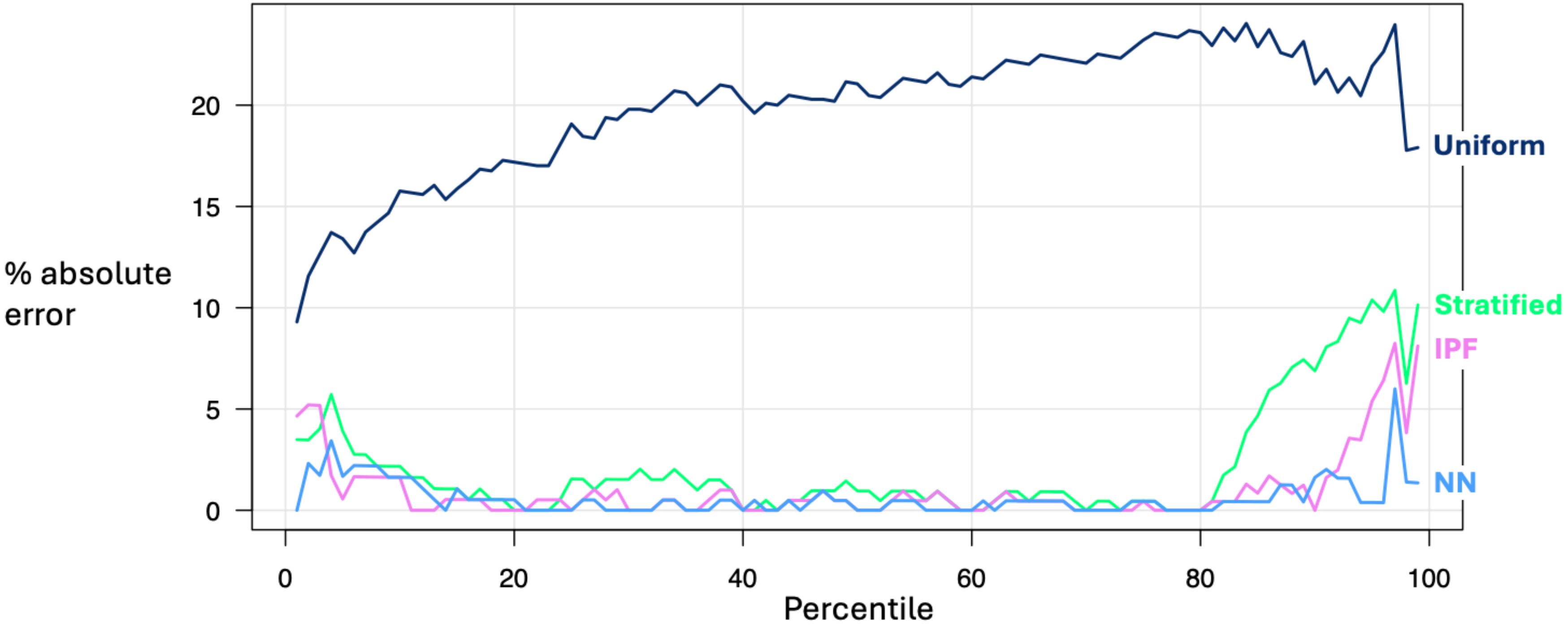
Distribution-based methods


inputs
detailed anthropometry for a reference population
key variables for the target population with estimates for prevalence
Population reweighting methods
interpolation
outputs
Statistical weights for the reference population that approximate key variables in the target population

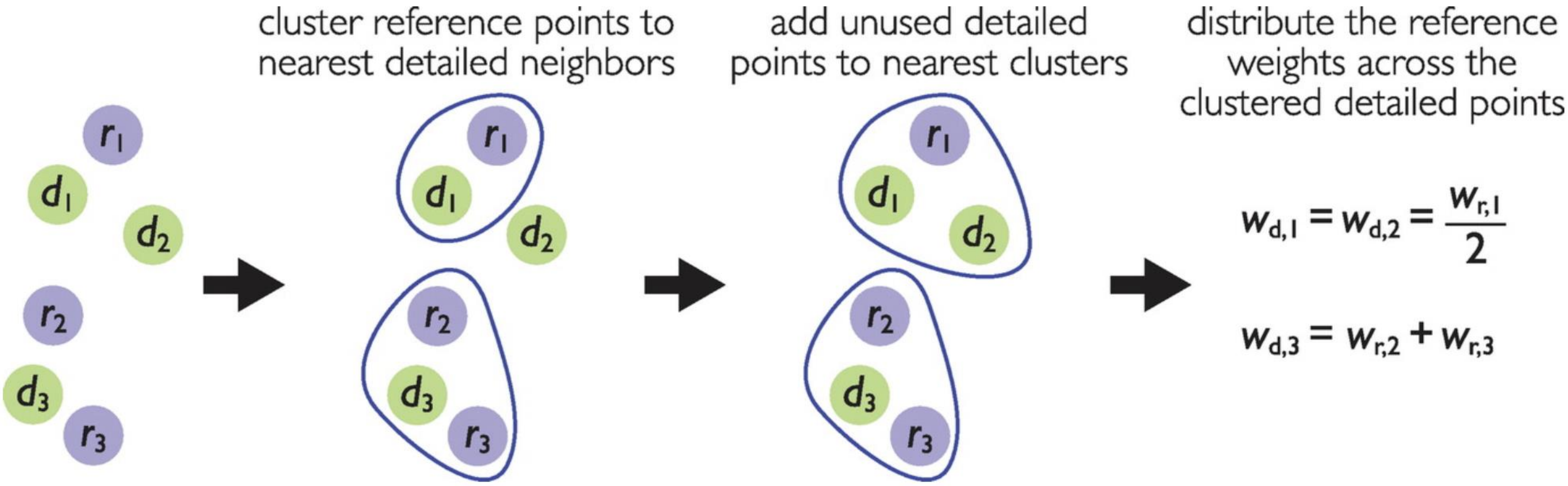
Population reweighting methods



ANSUR II BMI reweighted to represent Japanese females:



Population reweighting methods: Nearest-neighbor weighting



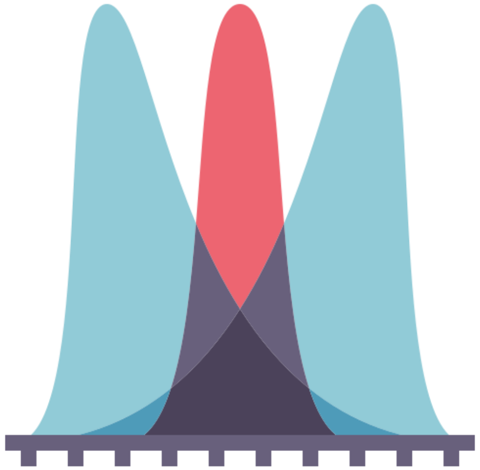
Kumar et al. 2017

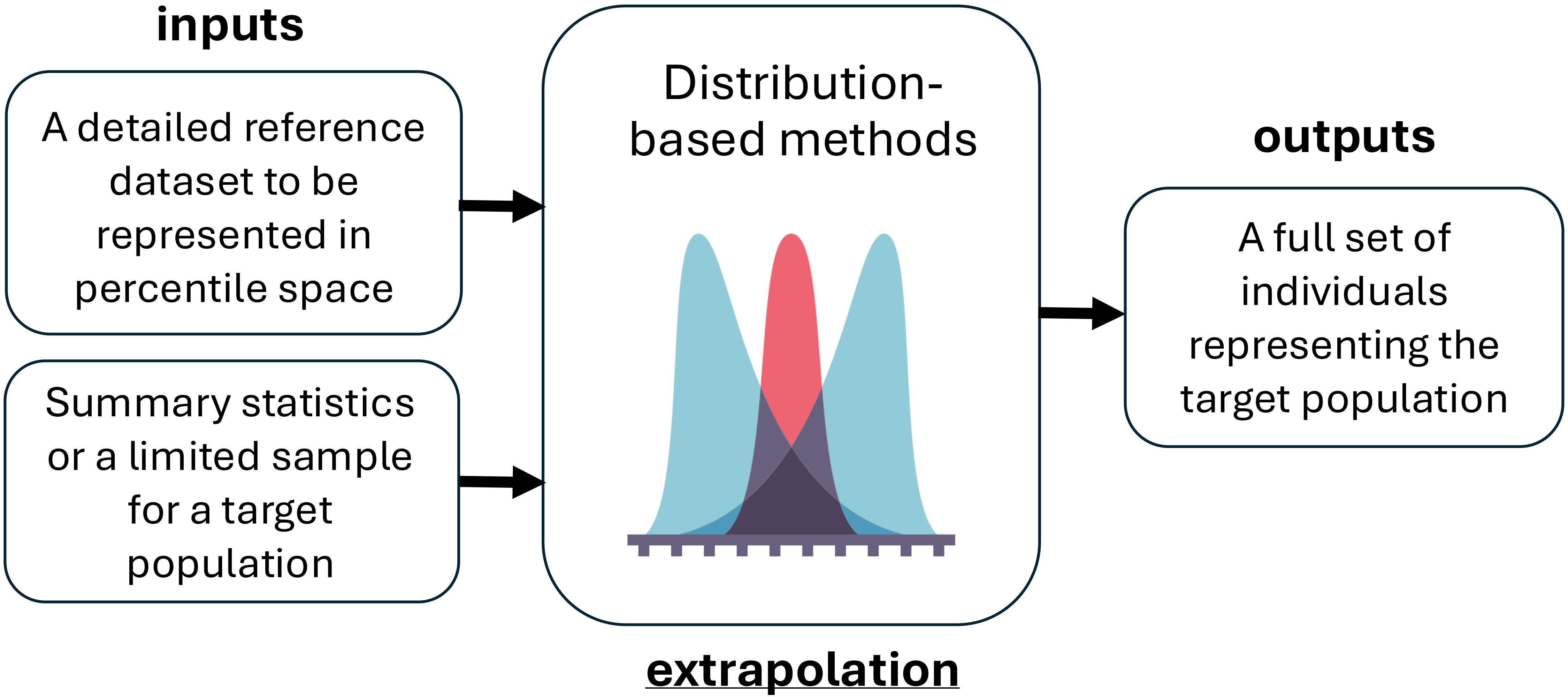
Data synthesis can be broadly categorized into three categories.

Regression-based methods

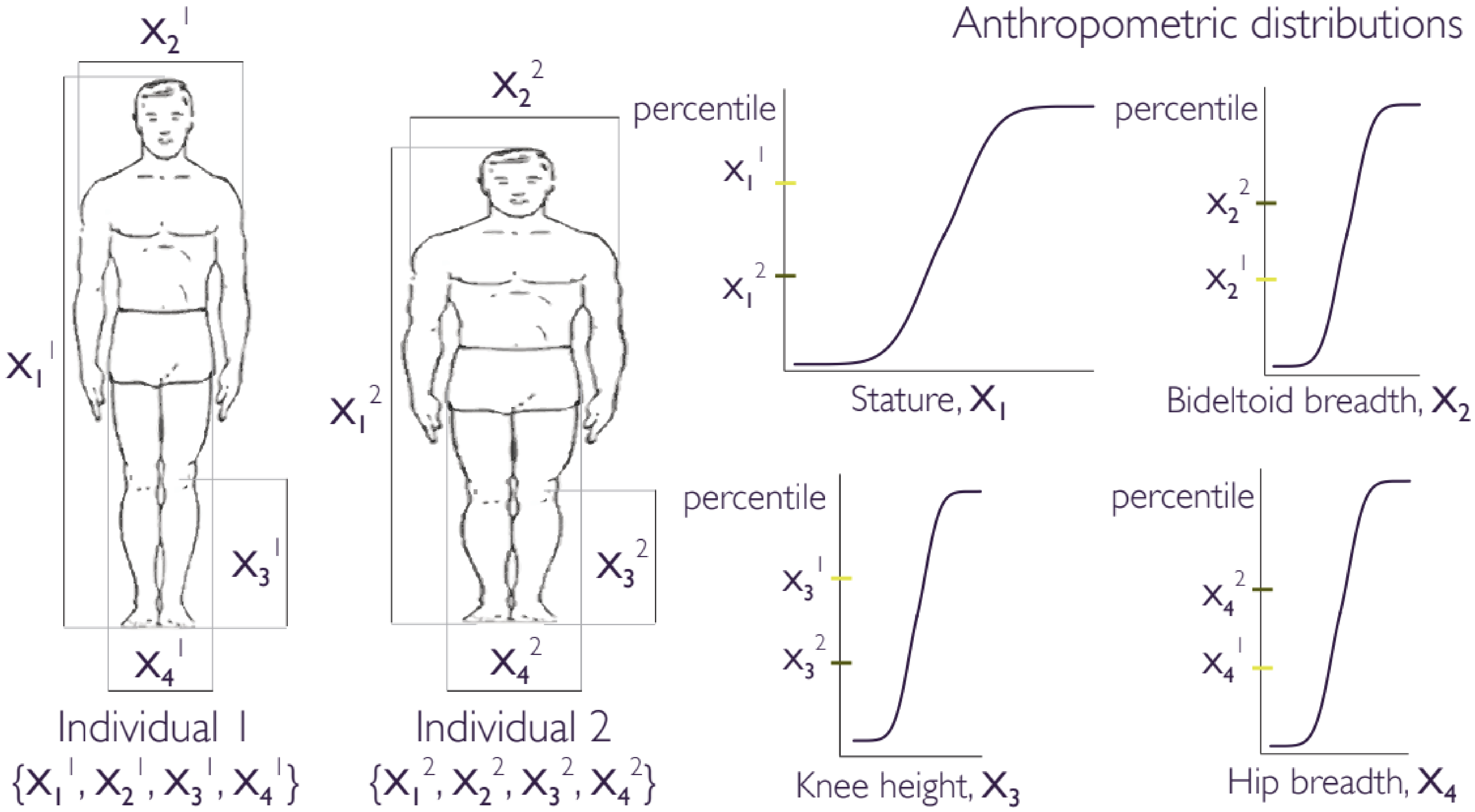
Population reweighting methods

Distribution-based methods




inputs
A detailed reference dataset to be represented in percentile space
Summary statistics or a limited sample for a target population
Distribution-based methods
extrapolation
outputs
A full set of individuals representing the target population

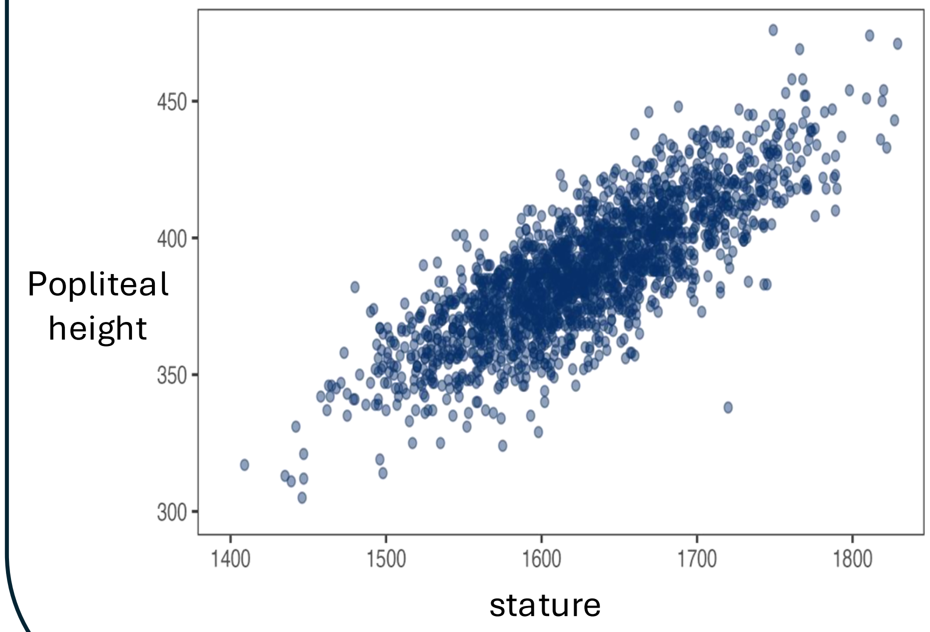
Distribution-based methods operate under the assumption that percentile combinations across two populations are roughly equivalent.



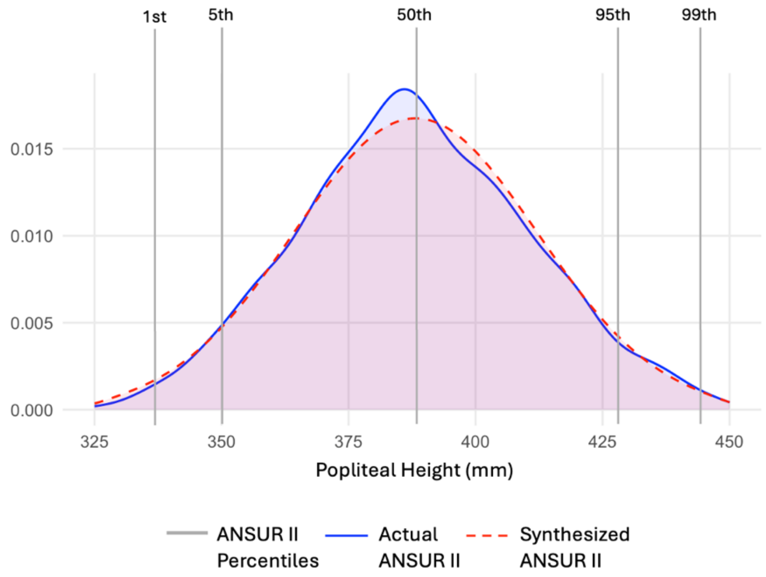
Nadadur et al. 2016

Distribution-based methods

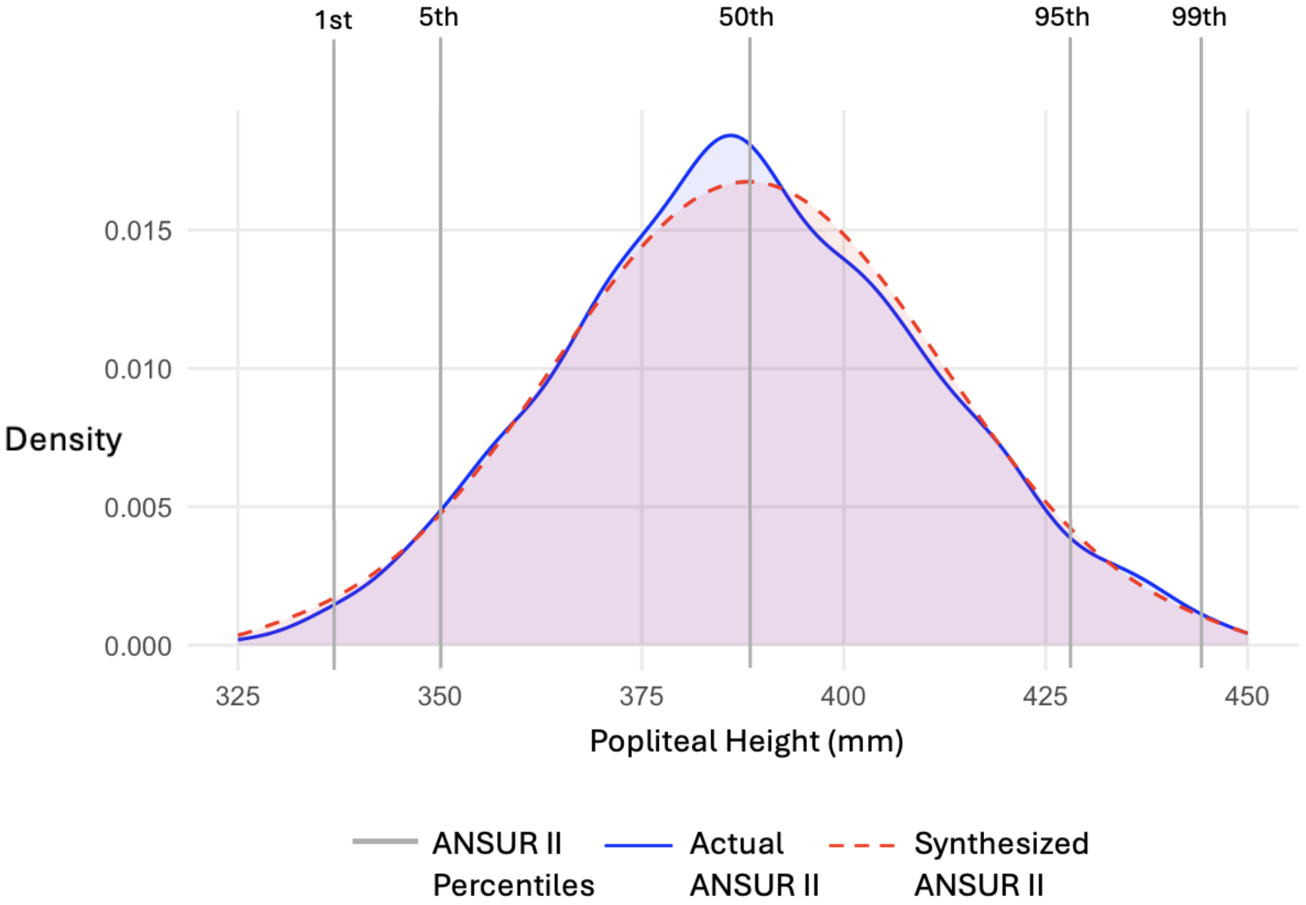
Copulas



Quantile-based

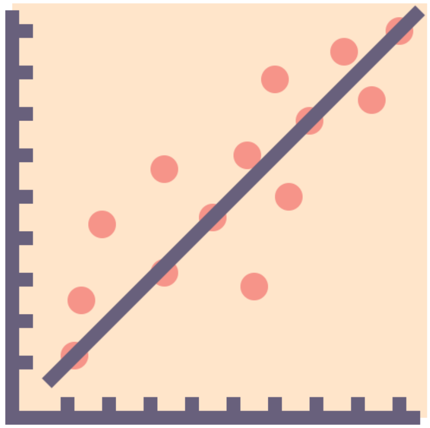


Distribution-based methods: Quantile-based synthesis



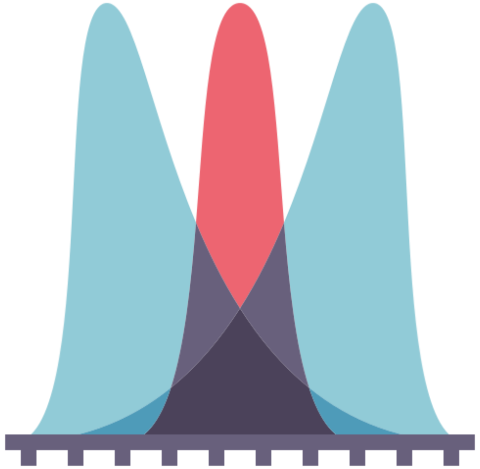
Data synthesis can be broadly categorized into three categories.

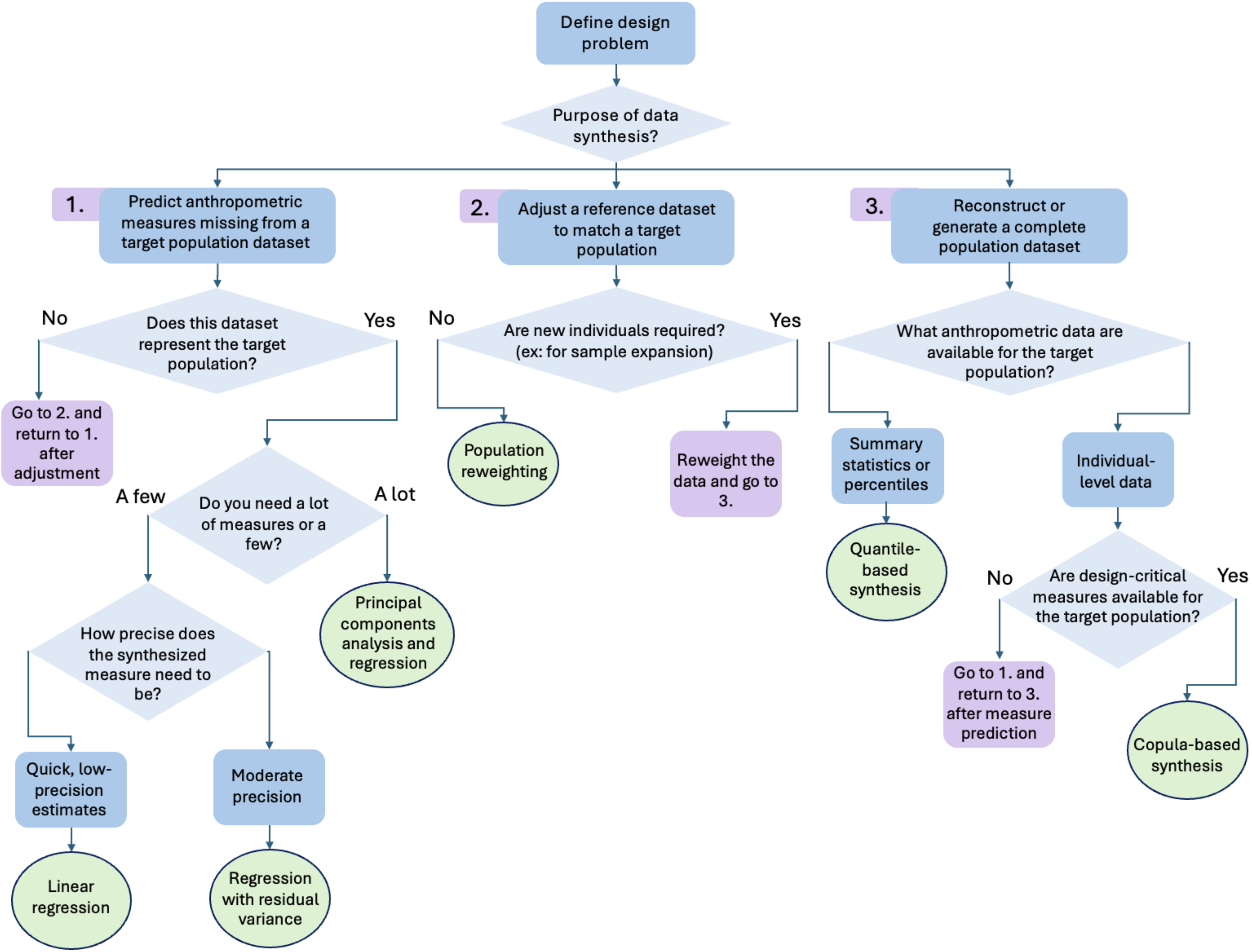
Regression-based methods



Population reweighting methods

Distribution-based methods

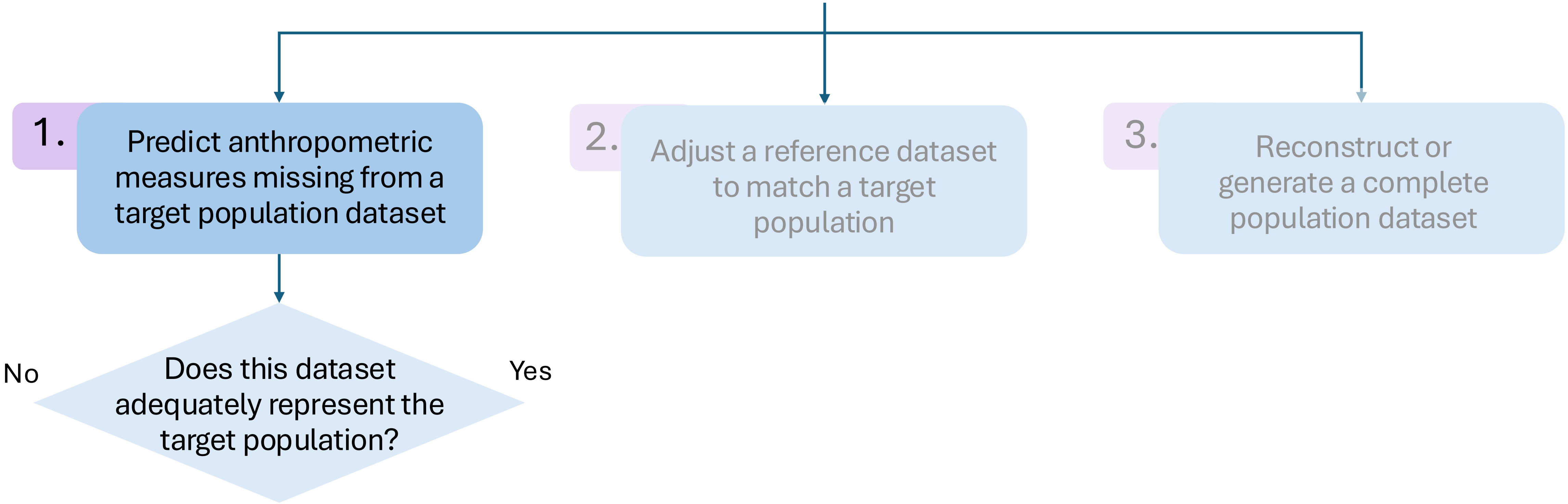



Define design problem
Purpose of data synthesis?
1.
Predict anthropometric measures missing from a target population dataset
2.
Adjust a reference dataset to match a target population
3.
Reconstruct or generate a complete population dataset
No
Does this dataset represent the target population?
Yes
Go to 2. and return to 1. after adjustment
A few
Do you need a lot of measures or a few?
A lot
How precise does the synthesized measure need to be?
Principal components analysis and regression
Quick, low-precision estimates
Moderate precision
Linear regression
Regression with residual variance
No
Are new individuals required? (ex: for sample expansion)
Yes
Population reweighting
Reweight the data and go to 3.
What anthropometric data are available for the target population?
Summary statistics or percentiles
Individual-level data
Quantile-based synthesis
No
Are design-critical measures available for the target population?
Yes
Go to 1. and return to 3. after measure prediction
Copula-based synthesis

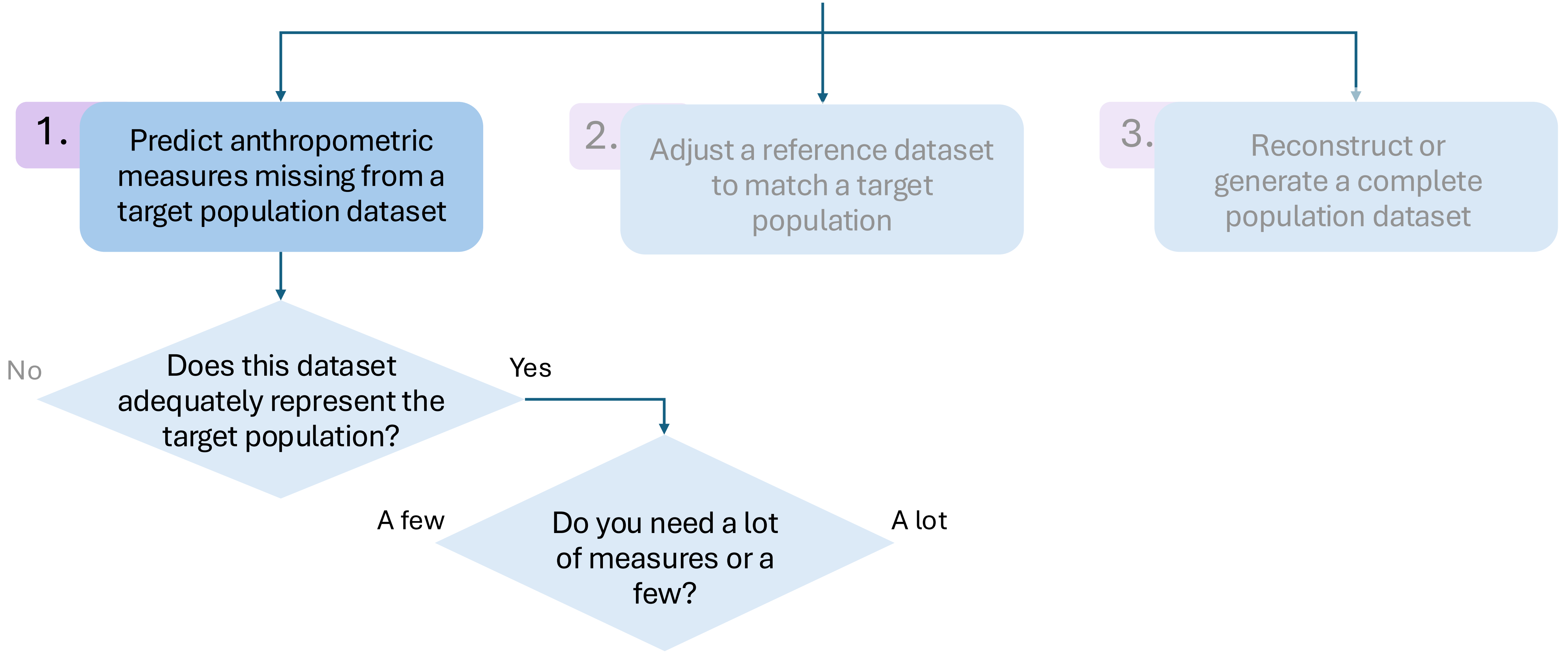
Scenario 1: You want to evaluate a design for the U.S. civilian population. You have nationally-representative data (such as NHANES), but it does not contain the required measures

1. Predict anthropometric measures missing from a target population dataset
2. Adjust a reference dataset to match a target population
3. Reconstruct or generate a complete population dataset

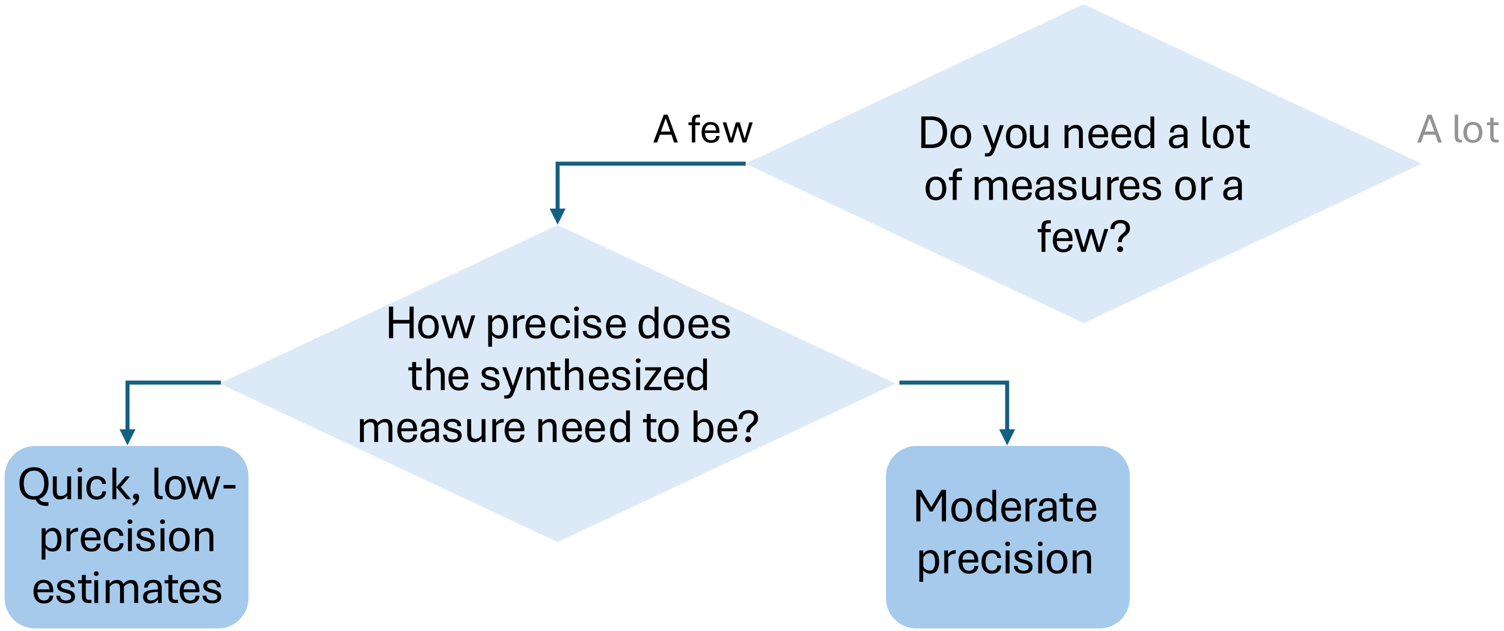
Scenario 1: You want to evaluate a design for the U.S. civilian population. You have nationally-representative data (such as NHANES), but it does not contain the required measures



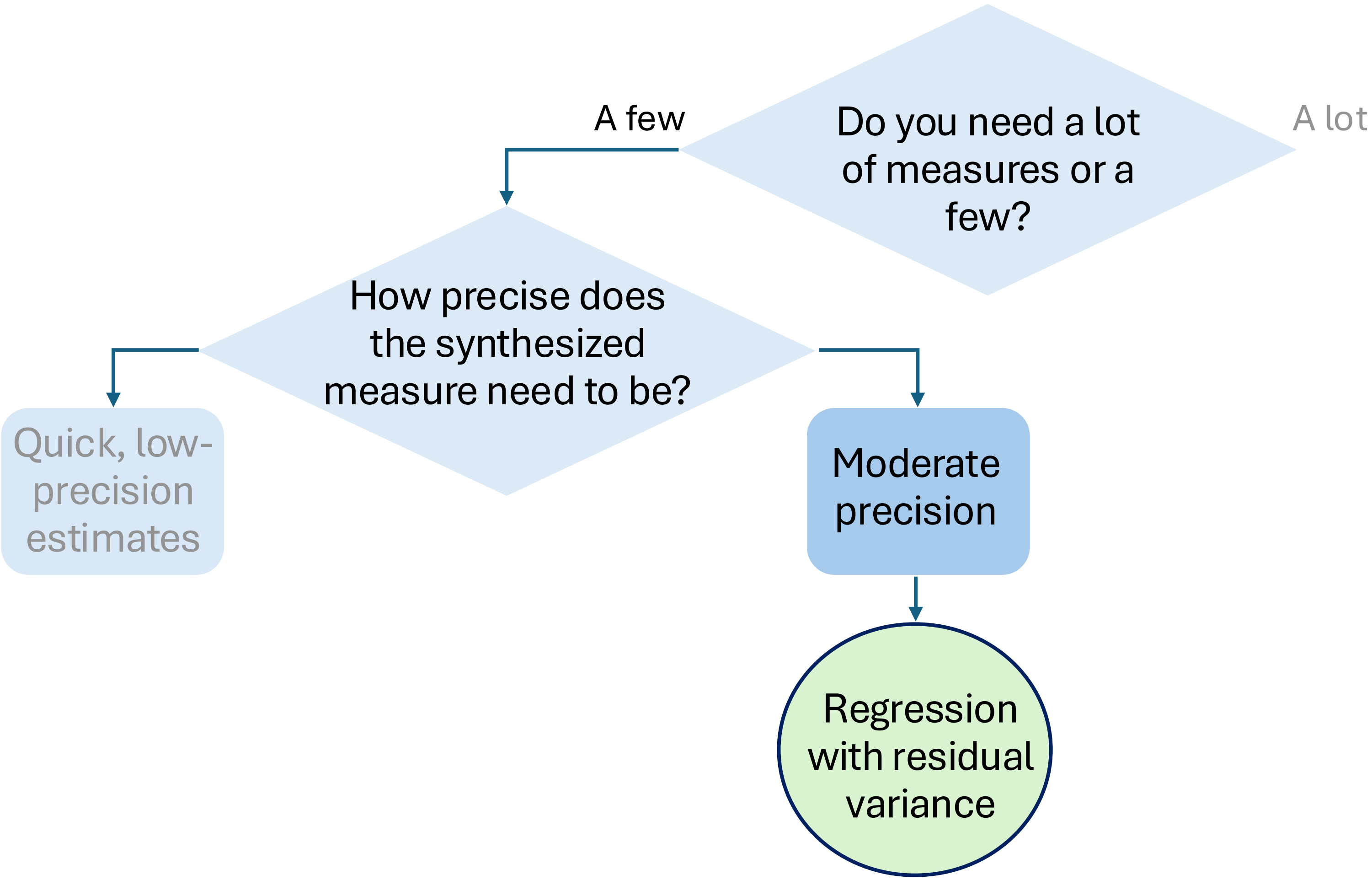
Scenario 1: You want to evaluate a design for the U.S. civilian population. You have nationally-representative data (such as NHANES), but it does not contain the required measures



Scenario 1: You want to evaluate a design for the U.S. civilian population. You have nationally-representative data (such as NHANES), but it does not contain the required measures



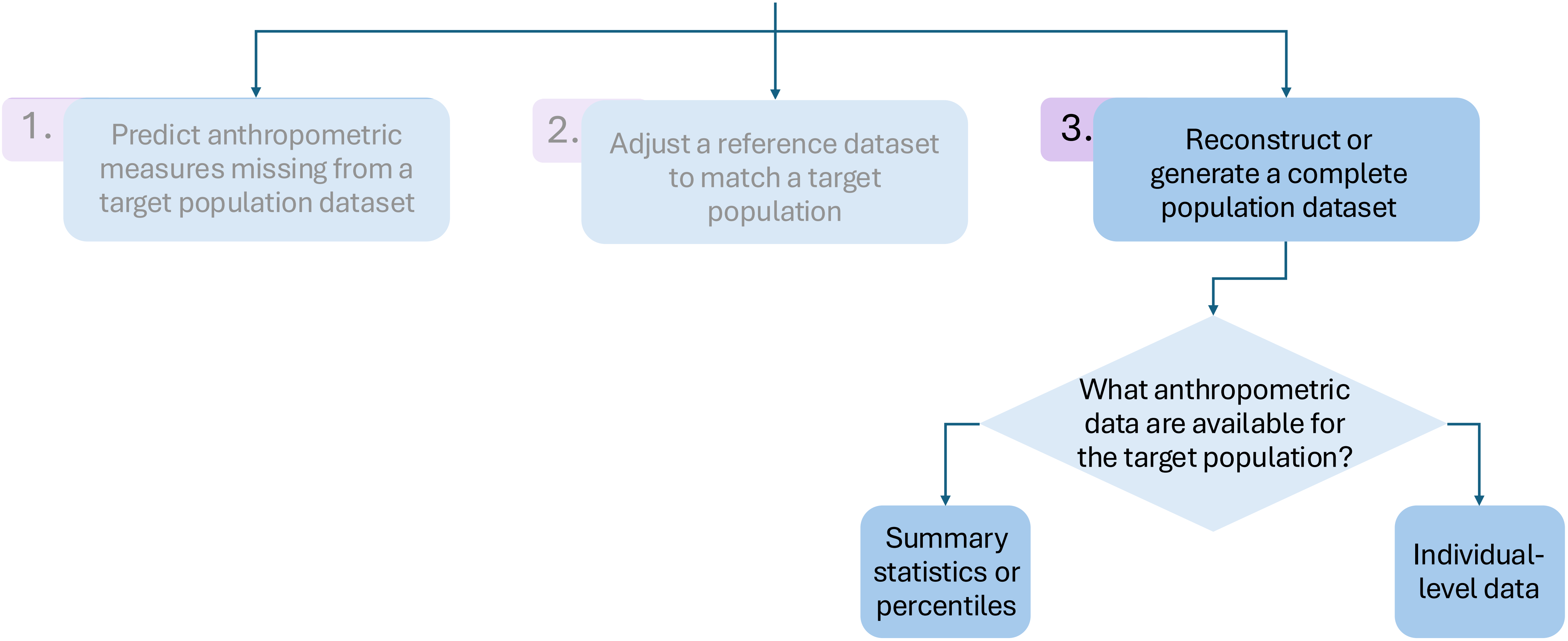
Scenario 1: You want to evaluate a design for the U.S. civilian population. You have nationally-representative data (such as NHANES), but it does not contain the required measures



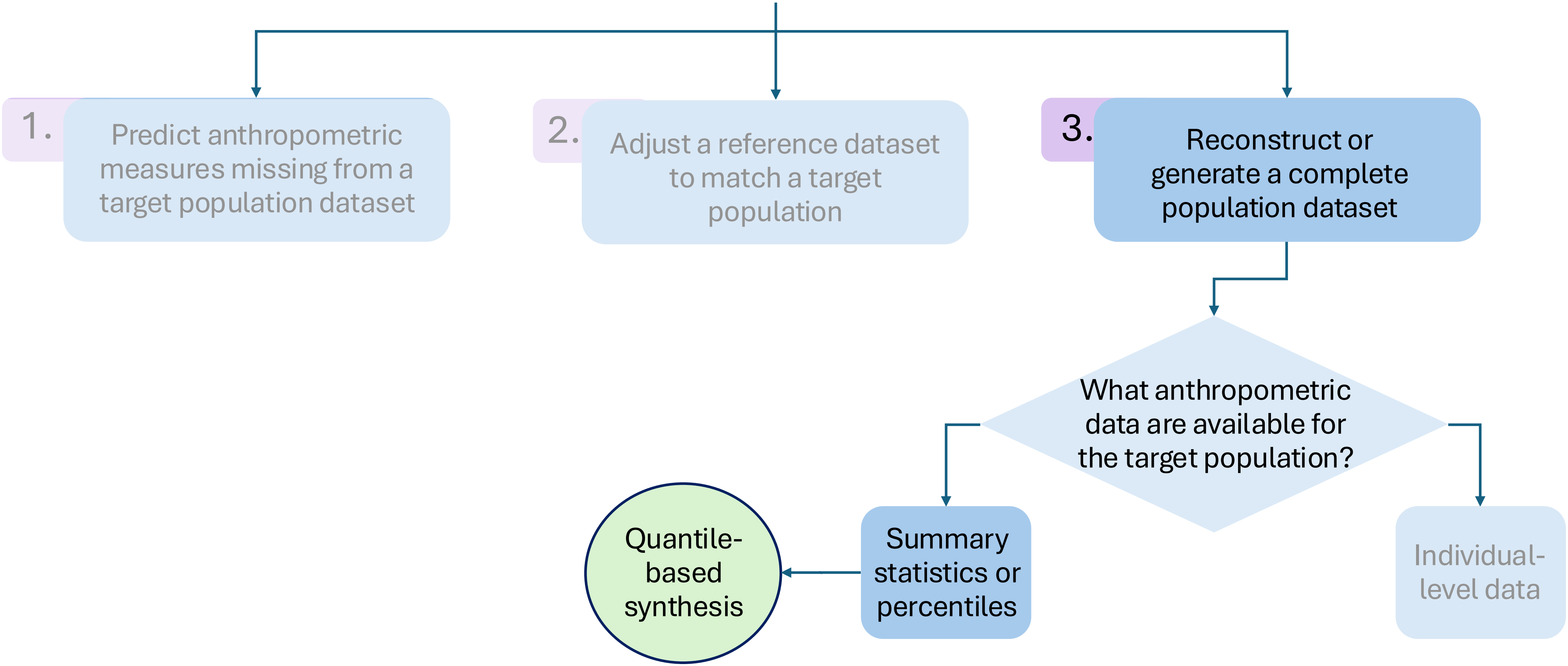
Scenario 2: You want to evaluate a design for the German population. You have nationally-representative detailed data, but it is only reported as summary stats.

1. Predict anthropometric measures missing from a target population dataset
2. Adjust a reference dataset to match a target population
3. Reconstruct or generate a complete population dataset

Scenario 2: You want to evaluate a design for the German population. You have nationally-representative detailed data, but it is only reported as summary stats.



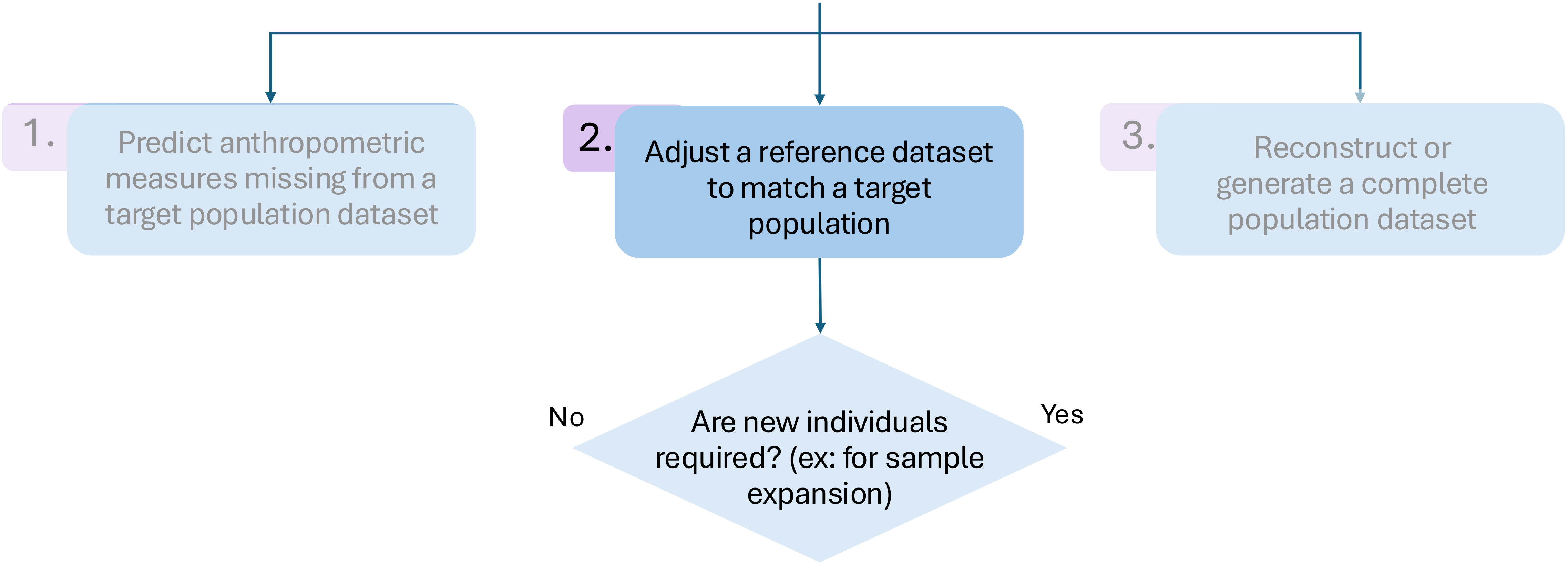
Scenario 2: You want to evaluate a design for the German population. You have nationally-representative detailed data, but it is only reported as summary stats.



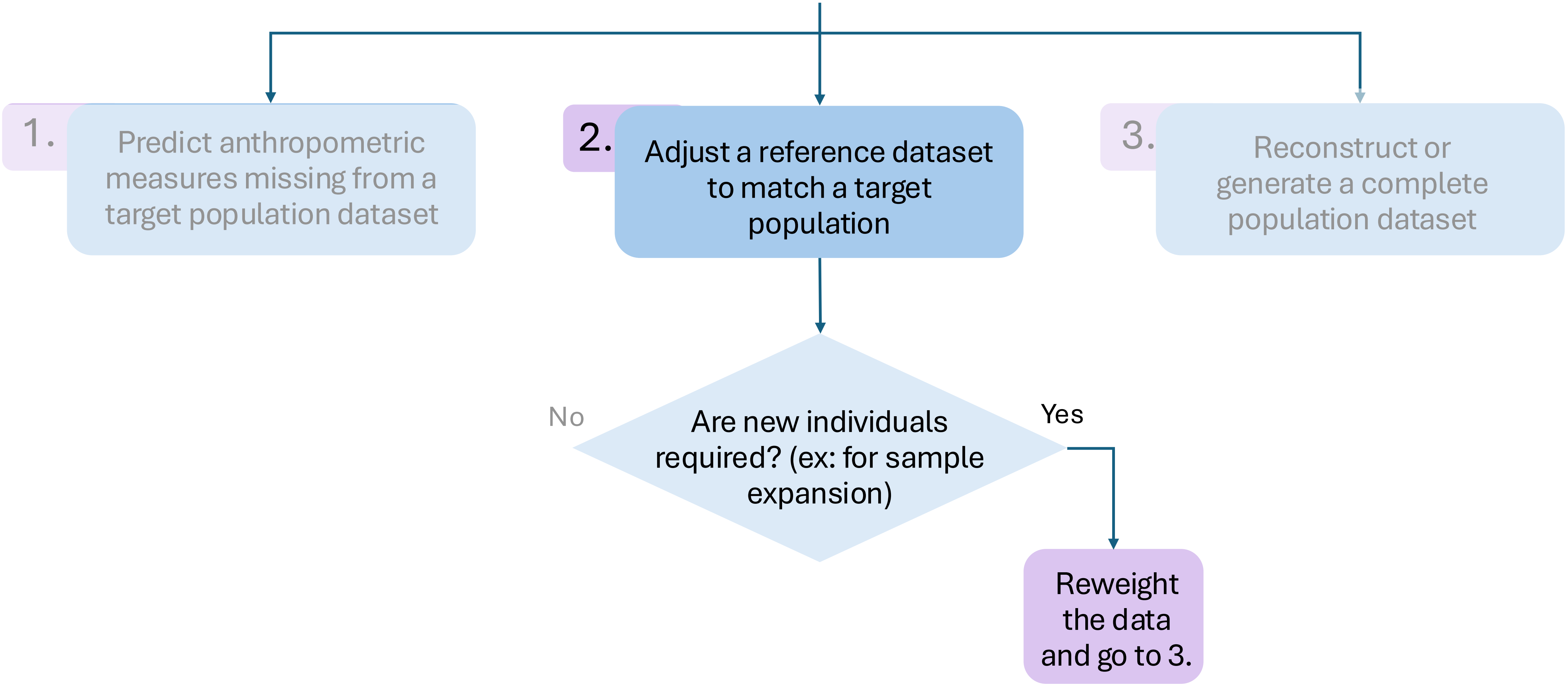
Scenario 3: You have collected a set of data from an experiment that needs to be adjusted to represent the U.S. population and expanded to 1000 people for a virtual fit trial.

1. Predict anthropometric measures missing from a target population dataset
2. Adjust a reference dataset to match a target population
3. Reconstruct or generate a complete population dataset

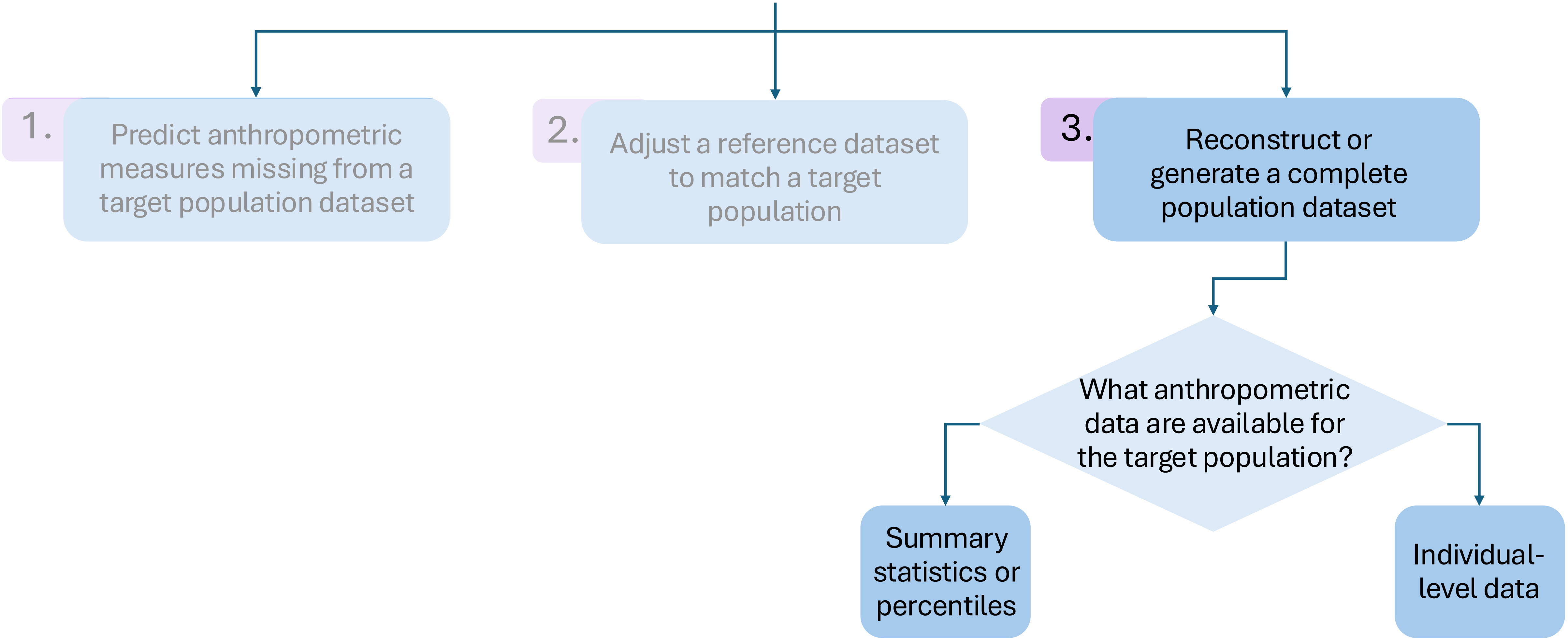
Scenario 3: You have collected a set of data from an experiment that needs to be adjusted to represent the U.S. population and expanded to 1000 people for a virtual fit trial.



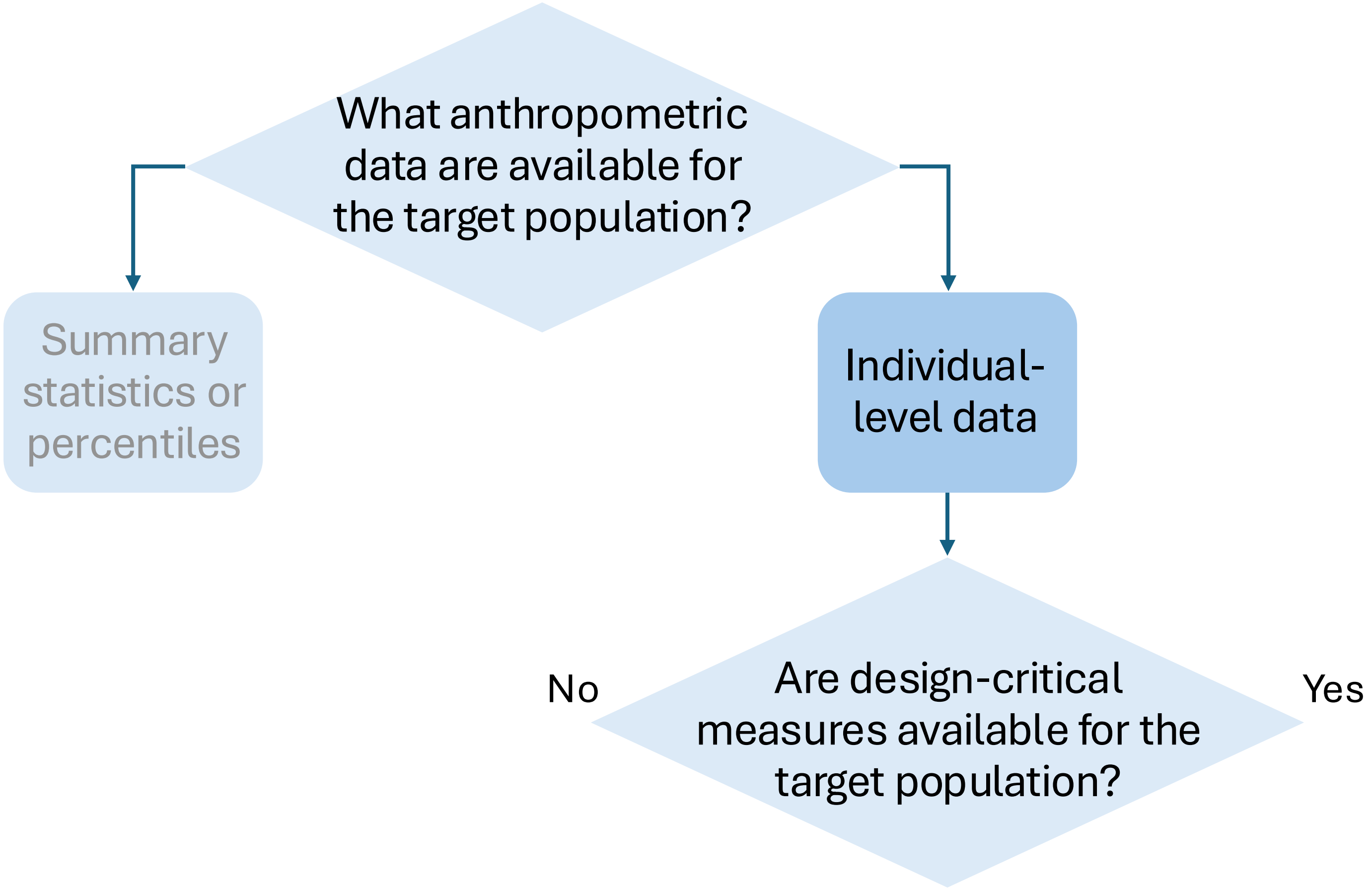
Scenario 3: You have collected a set of data from an experiment that needs to be adjusted to represent the U.S. population and expanded to 1000 people for a virtual fit trial.



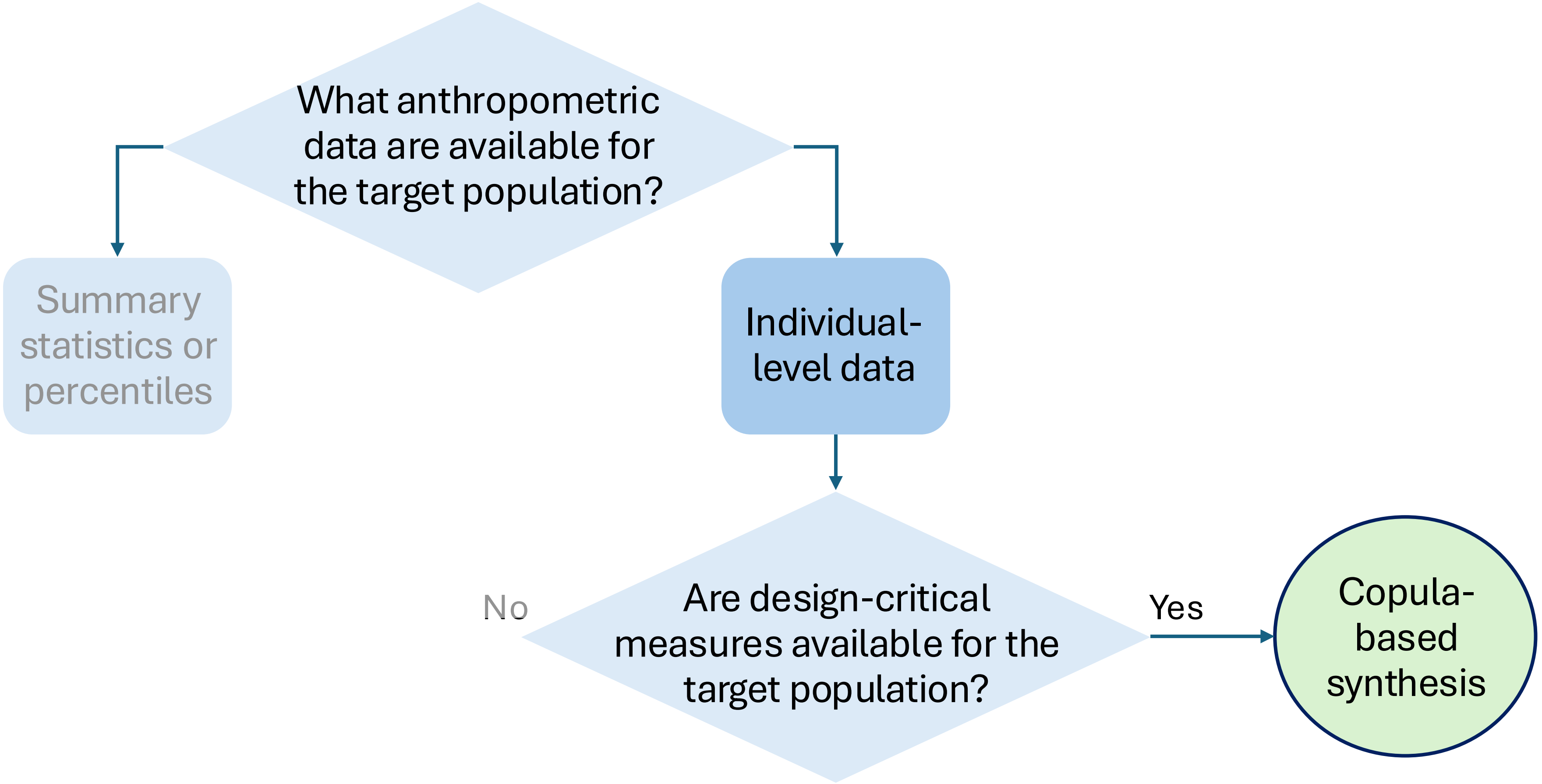
Scenario 3: You have collected a set of data from an experiment that needs to be adjusted to represent the U.S. population and expanded to 1000 people for a virtual fit trial.



Scenario 3: You have collected a set of data from an experiment that needs to be adjusted to represent the U.S. population and expanded to 1000 people for a virtual fit trial.



Scenario 3: You have collected a set of data from an experiment that needs to be adjusted to represent the U.S. population and expanded to 1000 people for a virtual fit trial.



Thanks!

open
design lab

Escaping the 1D Bottleneck: An Online Framework for Applying Population-Level 3D Anthropometry to Design

Matt Reed
Daniel Park
Biosciences Group

March 2026

UMTRI TRANSPORTATION RESEARCH INSTITUTE
UNIVERSITY OF MICHIGAN

Disclosures

The University of Michigan has licensed intellectual property described in this presentation to commercial entities.

Dr. Reed and Dr. Park receive royalties from these licenses.

This research was supported by multiple organizations and collaborators, including:

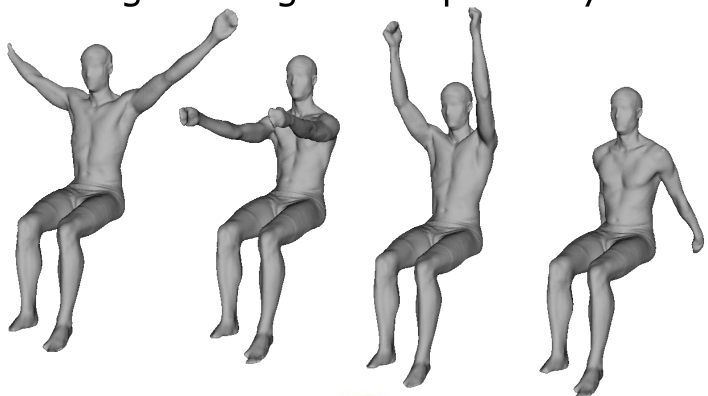
Amazon Inc.
Gentex Corp.
Meta Platforms, Inc.
U.S. Army
U.S. Navy
U.S. Air Force
U.S. Marine Corps
SENTIR Labs (STI-TEC)

Matthew Reed, PhD
mreed@umich.edu
http://mreed.umtri.umich.edu

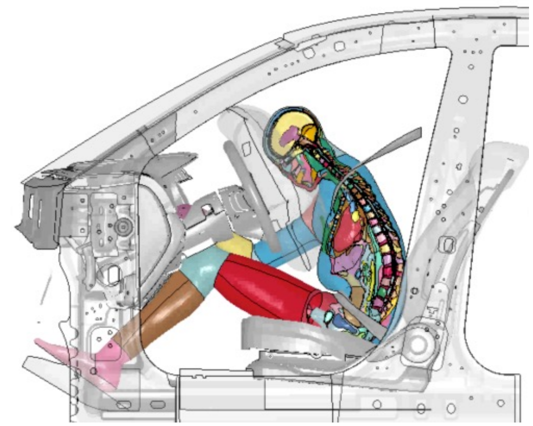
M | UMTRI

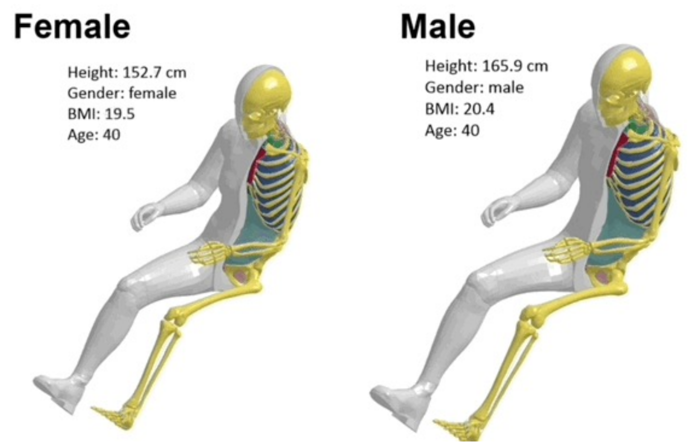
Biosciences Group: Major Research Areas

Engineering Anthropometry

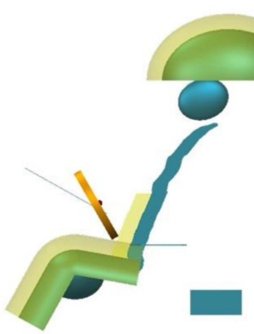


Occupant Crash Protection



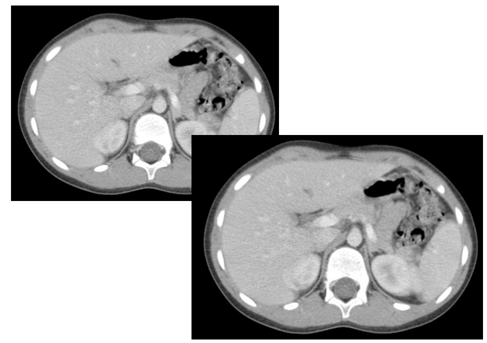


Computational Human Body Modeling



Vehicle Occupant Posture Prediction and Accommodation Modeling

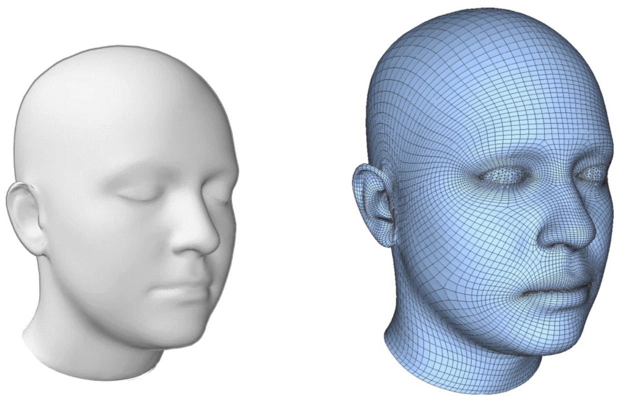
Over 25 years of research with U.S. Department of Defense



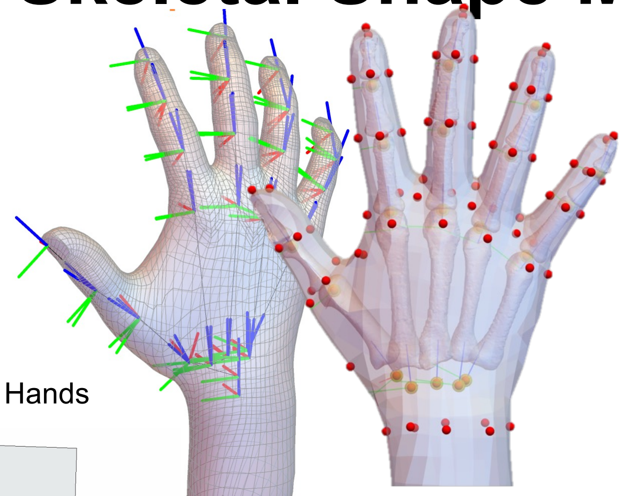
Injury Biomechanics

UMTRI

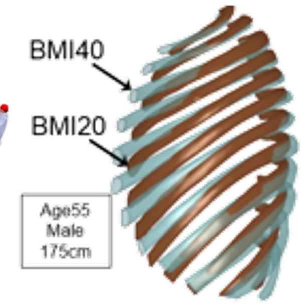
Regional and Skeletal Shape Modeling

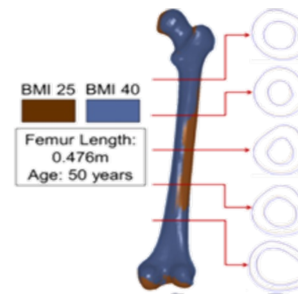


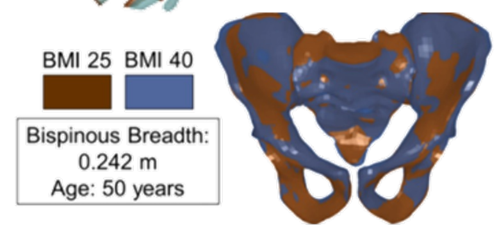
Head with accurate scalp

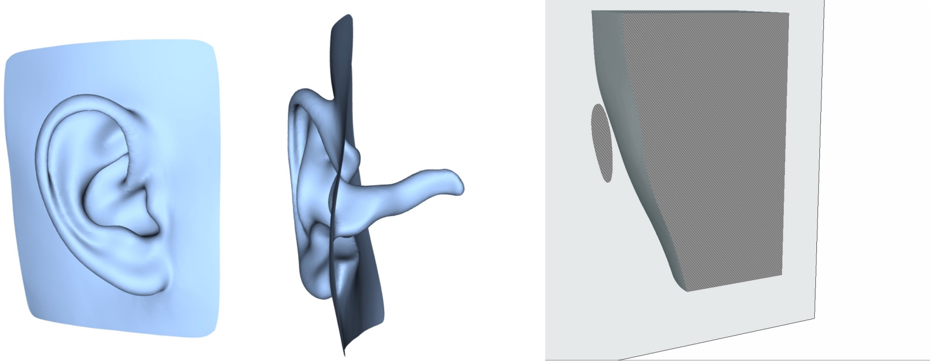


Hands

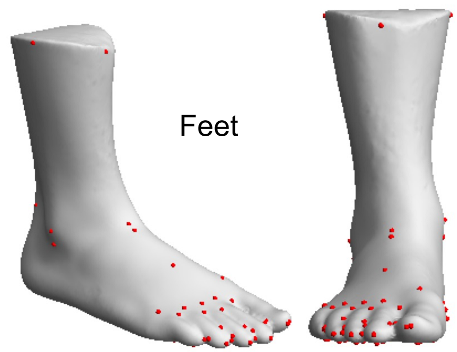




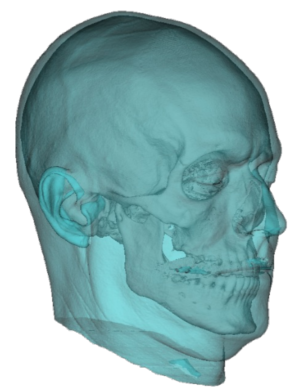




Detailed ear geometry with canal



Feet



skeletal structures

UMTRI

Online models at http://HumanShape.org/

Applied Anthropometry is Stuck in the 1950s

How do nearly all product designers find and use anthropometric data?

- do a Google search; when that fails, ask AI! or
- look in a reference text for a dimension that looks like it might be relevant
- select a 5th or 95th percentile value

or

- use a manikin, template, headform, or some other pre-defined reference: even when 3D forms are used, usually fewer than 6 and often **only one 3D form** is used

What are the problems with this approach?

- univariate
- one-dimensional data
- wrong dimension
- wrong population
- wrong size (manikin approach)

UMTRI

Surely things are better for military applications?

Not much better:

- In the U.S., detailed **3D** anthropometric information for warfighters (e.g., ANSUR II) is generally not available to product designers; 10s of thousands of warfighters have been scanned, but only a few people have access due to privacy concerns
- Most designers are using univariate statistics from published reports
- Tools for actually using population-level 3D data in design are not available and hence aren't in acquisition requirements

What are the problems with this approach?

- ✓ univariate
- ✓ one-dimensional data
- wrong dimension (limited to what's been measured previously)
- wrong population (surface Navy?)
- wrong size (manikin approach)*

often

*The so-called "50th percentile male" is never a limiting case

UMTRI

Thought Experiment: How would you do product design…

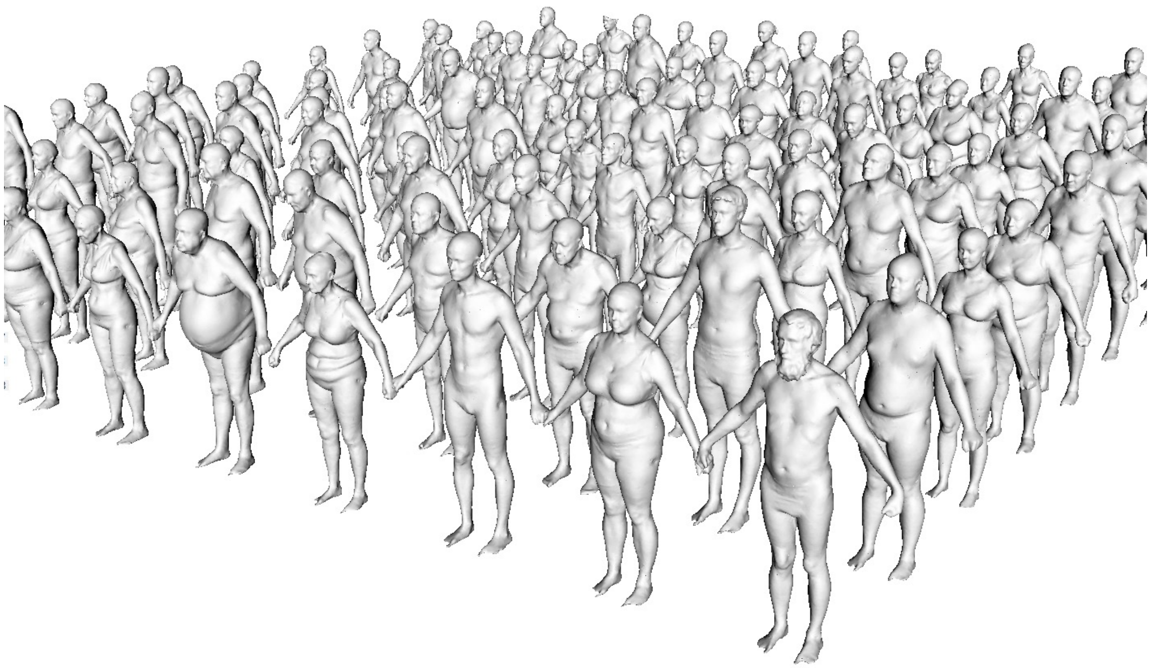
If you had infinite resources ….

and a time machine?

M | UMTRI

Fantasy Design World

- Infinite physical protoypes!
- Try everything on everyone!
- Optimize designs (size/shape/adjustability/sizing etc.) across the whole population by doing physical try-ons!



UMTRI

Back in the Real World: Virtual Fit Testing

Individual-level population characteristics that affect product interaction

Model of product interaction

Concept of fit: how is the product supposed to fit and how do you know if it fits

Cost functions that convert product interaction outputs into a quantitative score

Simple idea: count the number of users in a population who obtain “good fit”

VFT Data Requirements

Population data:

- correct population
- correct variables
- can be more than anthropometry: strength, age, etc – anything we can know about the population that influences the interaction with the product

Product interaction data:

- objective and subjective outcomes from parametric, structured interaction (fit/accommodation) testing
- statistical relationships between outcomes and population variables

Need data for **individuals** who **represent*** the **population**

* can be weighted or unweighted

UMTRI

Simple VFT

individual-level 1D anthropometric dimensions → assumed 1:1 relationship between anthro and fit/accommodation → Multivariate binary cost functions (fits/doesn't fit)

Example: HFES Multivariate Accommodation Test Tool (MATT)

https://my.hfes.org/online-store/publications

Multivariate Accommodation Testing Tool

IMPORTANT - Change only the **green** cells. The **red** cells will update automatically in response.

Fraction Male **0.3** (between 0 and 1)

	Quantiles - Men			Quantiles - Women			Associated Dimension Value		Percent Accommodated		
Measures	5%	50%	95%	5%	50%	95%	Low	High	Men	Women	Both
Stature (mm)	1634	1758	1882	1501	1621	1741					
BMI (kg/m^2)	21	28	39	20	28	43					
Seated Measures											
Abdominal Extension Depth (mm)	228	290	420	212	288	424					
Buttock-Knee Length (mm)	559	616	679	532	591	656					
Buttock-Popliteal Length (mm)	440	486	548	420	474	529					
Elbow Rest Height, Sitting (mm)	193	242	290	192	235	281					
Eye Height, Sitting (mm)	736	802	859	686	748	805					
Forearm-Forearm Breadth (mm)	482	566	681	410	491	633					
Hip Breadth, Sitting (mm)	333	382	455	353	421	530		530	99.4%	94.9%	96.2%
Knee Height (mm)	508	557	606	459	504	550		600	91.7%	99.7%	97.3%
Popliteal Height (mm)	390	434	478	338	382	426	340	470	90.1%	93.7%	92.6%
Shoulder Breadth (mm)	444	498	562	387	436	524					
Thigh Clearance (mm)	138	168	204	124	150	196					
Elbow Height Above Floor Estimated (mm)	615	679	739	554	618	681					
Thigh Clearance Above Floor Estimated (mm)	540	602	662	480	536	594					
Knee Depth Estimated (mm)	214	319	387	187	299	371					
Foot Depth Estimated (mm)	334	460	544	283	422	511					
Standing Measures											
Eye Height, Standing (mm)	1517	1640	1762	1400	1512	1631					
Elbow-Fingertip Length (mm)	431	470	512	386	421	462					
Elbow Rest Height, Standing (mm)	997	1084	1177	914	1001	1090					
Foot Length (mm)	244	266	291	220	239	260					
							Total Accommodation		**86.8%**	**89.9%**	**89.0%**

UMTRI

But What About 3D?

To a reasonably approximation, <1% of 3D body shape data that have been gathered have ever been used in product design

Typical Approaches to 3D Databases

Faster tape measure: extract the same 1D dimensions that would have measured by hand, proceed as if 1975 [VFT on 1D measures]

PCA on the wrong variables: pick some overall body dimensions (stature, sitting height, …) do a PCA & generate boundary cases, find nearest neighbors in 3D, ignore the rest of the 3D data while using the boundary cases for specific analyses not related to these variables

Why Haven't 3D Data Helped (much)?

Scan data is rarely and minimally used. Why?

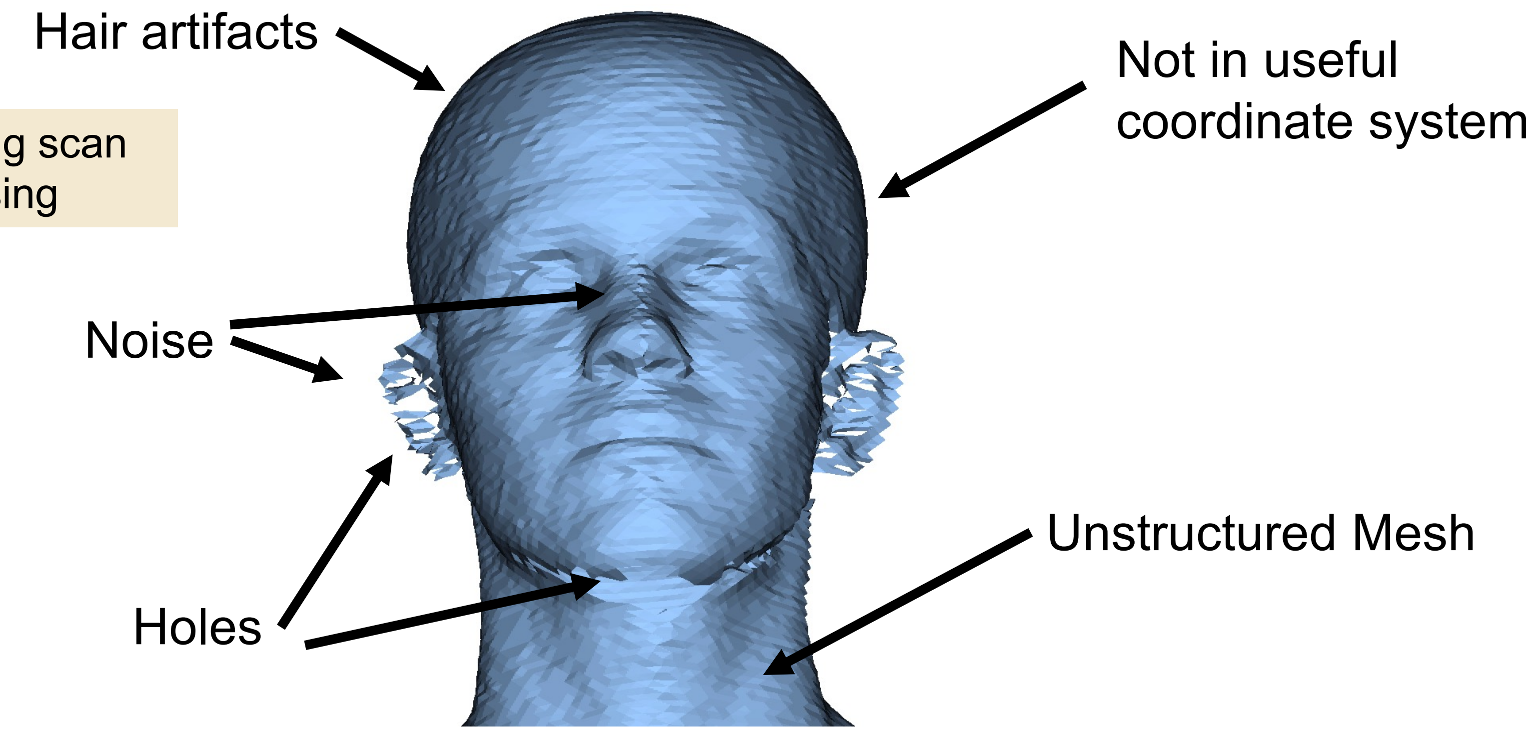
- scan files are hard to use, effectively impossible for most designers; template fitting is essential but uncommon
- privacy limits access
- doing good scanning is hard (and still expensive)
- sampling bias is worse than with 1D dimensions
- statistical knowledge needed for shape analysis is not broadly distributed
- **few software tools available to do useful things with 3D data**
- **even fewer tools available to useful things at a population level**

Some good 3D work has been done for clothing, body armor, helmets, etc. by DOD researchers, but this has not resulted in widely used methods or software, even within the goverment

UMTRI

Some of the Problems with 3D Scan Data

Typical standing scan before processing



No widely available software well-suited for using 3D scan data

What is Needed to Use 3D Data?

1. **Software**
2. Design methods
3. Education

Preliminaries: Database Development

High-quality 3D data from relevant population in relevant poses, with covariates

process/clean data (e.g., remove props); landmarking as needed

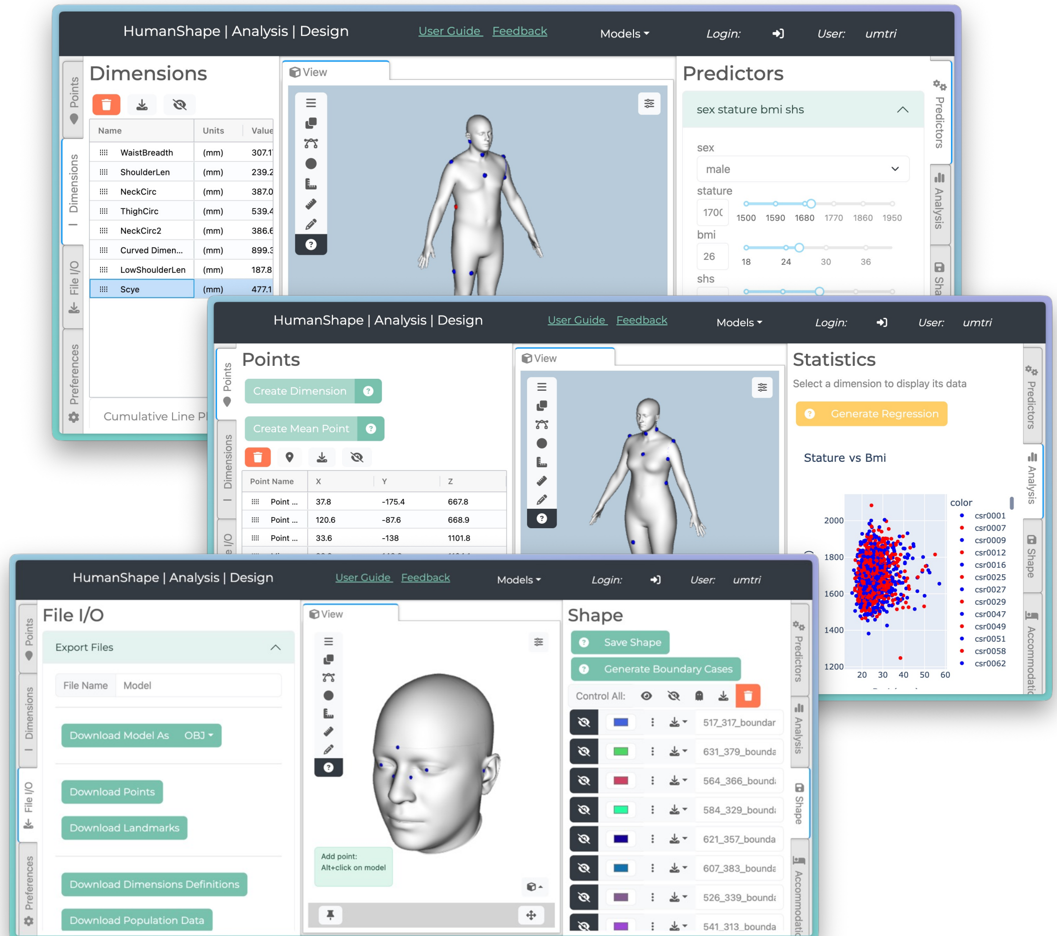
template fitting: all data on the same homologous mesh (non-rigid registration)

statistical analysis (typically PCA on vertex coordinates, but can be any parametric shape description)

HumanShape App: Online Human-Centered Design

- Unified interface for working with all types of human body shape data, from whole-body to anatomical regions: hands, heads, ears, feet
- Online app means no software installation problems and accessible from anywhere
- Working with large populations of virtual avatars, not raw scan data
- Synthesis techniques ensure accuracy while eliminating privacy-related constraints

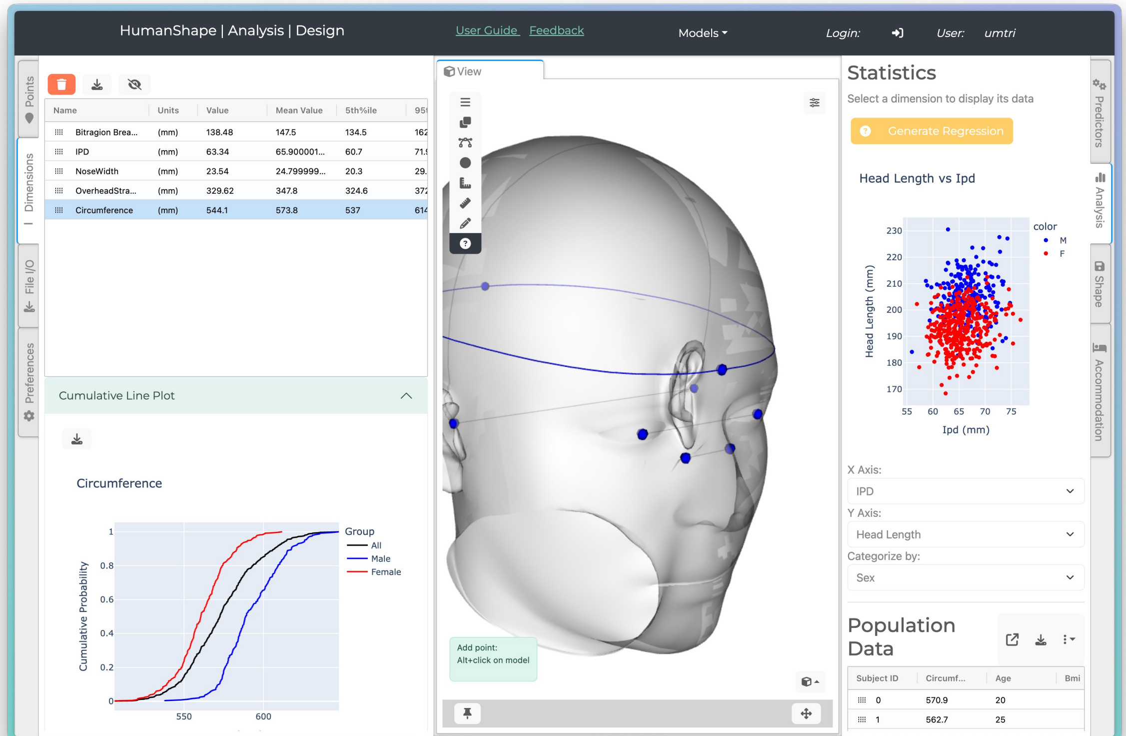
https://HumanShape.app/



UMTRI

HumanShape App: Online Human-Centered Design

- User-defined measurements – not limited to predefined measures
- Point-to-point dimensions, single-axis measurements, curves, circumferences, angles, centroids…
- All dimensions are automatically computed for every individual in the population
- Statistical analyses of population distributions of user-defined measures for male & female



UMTRI

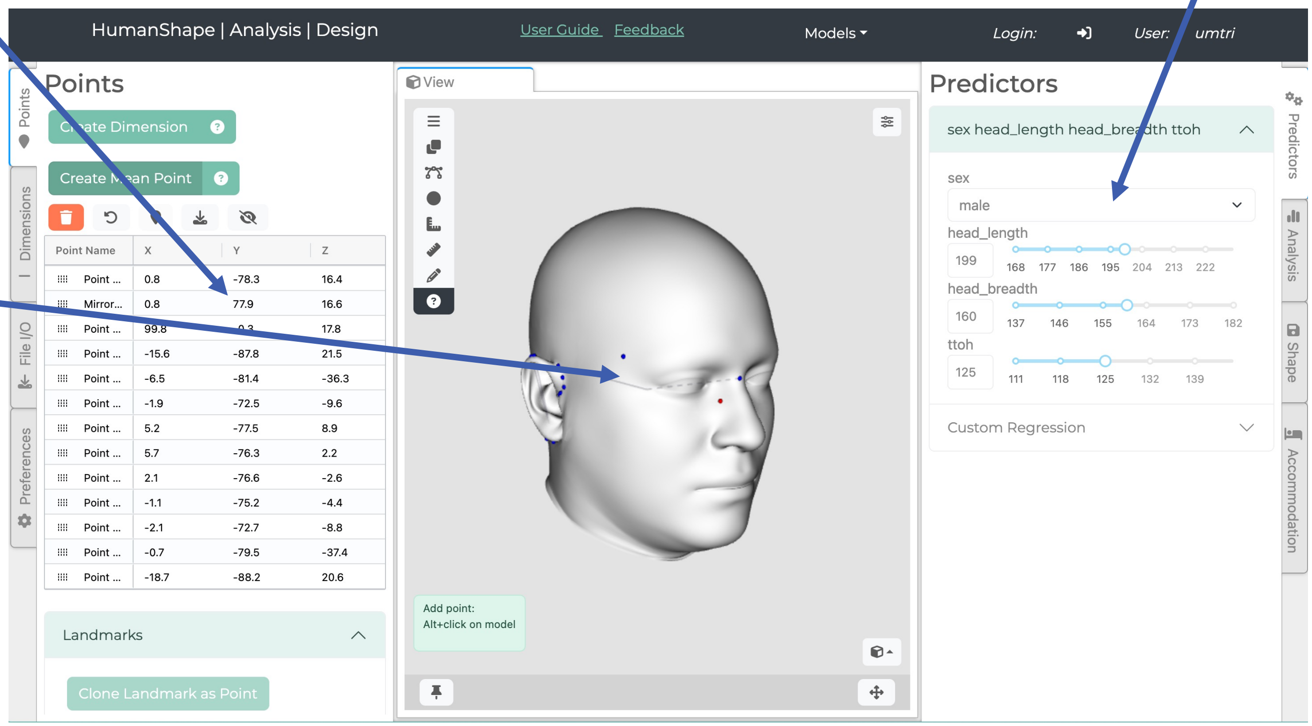
US Navy Aircrew head model available

HumanShape App: Functionality

User defined points

User defined dimensions: point to point, circumferences, surface curves, angles…

Pre-defined and user-defined regressions



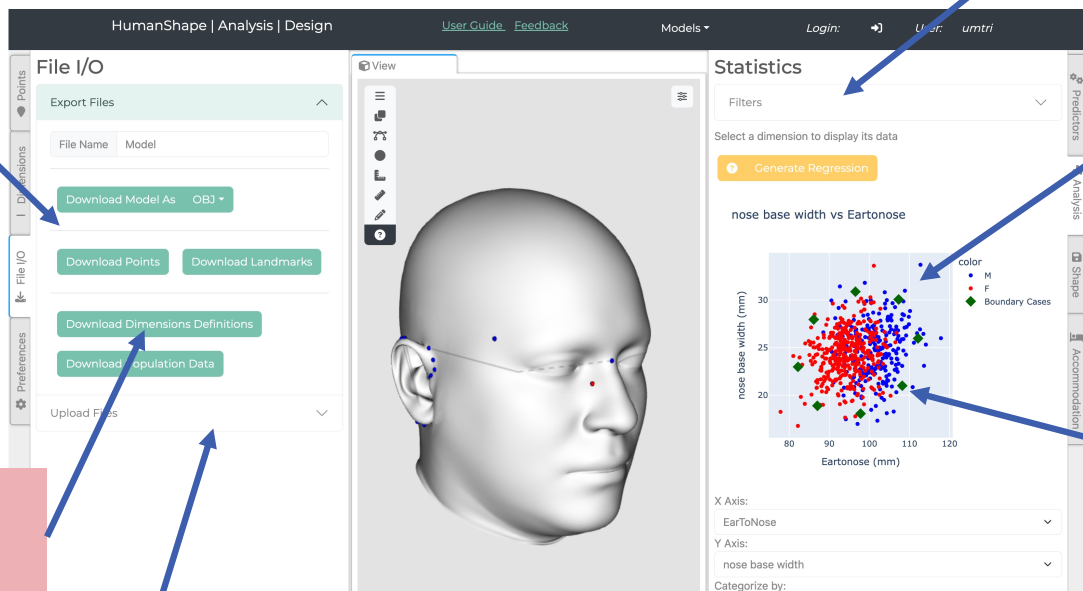
UMTRI

HumanShape App: Functionality

Apply range filters

Download shapes

Bivariate plots



Boundary case generation on user-defined variables

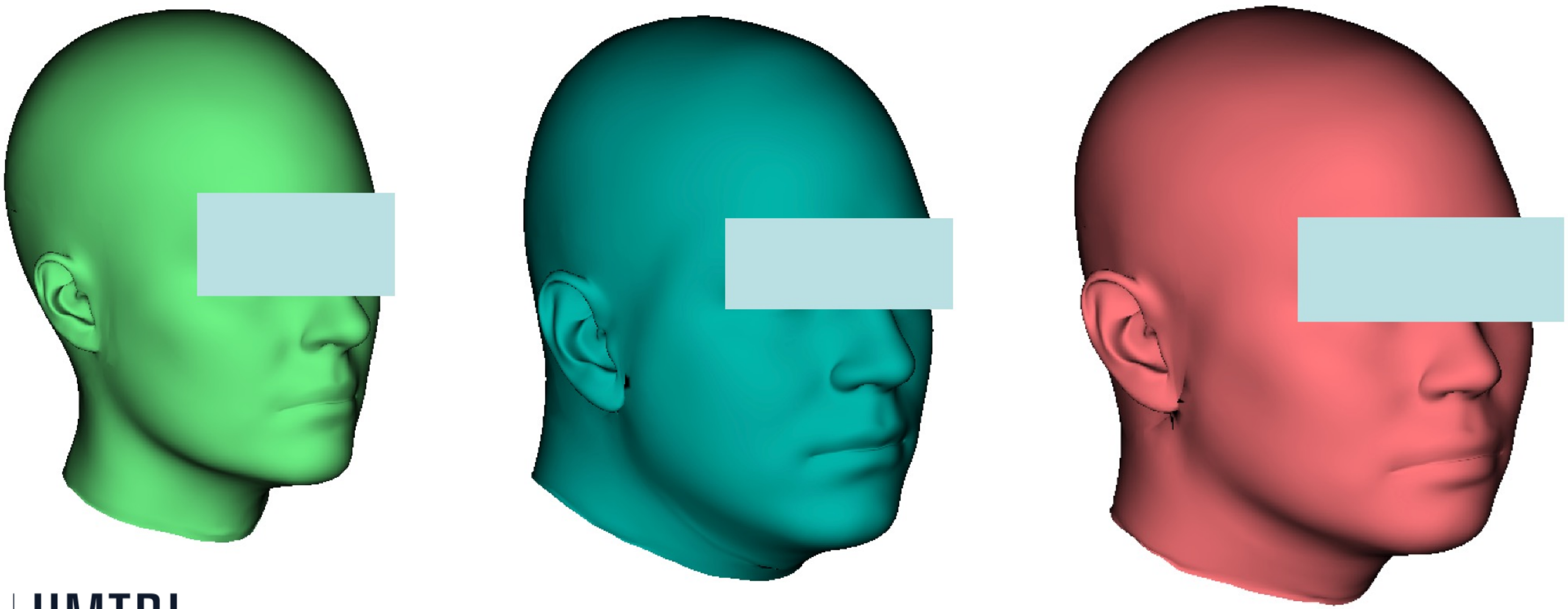
Download points, landmarks, dimensions, …

Upload geometry

UMTRI

Sidebar: De-Identification

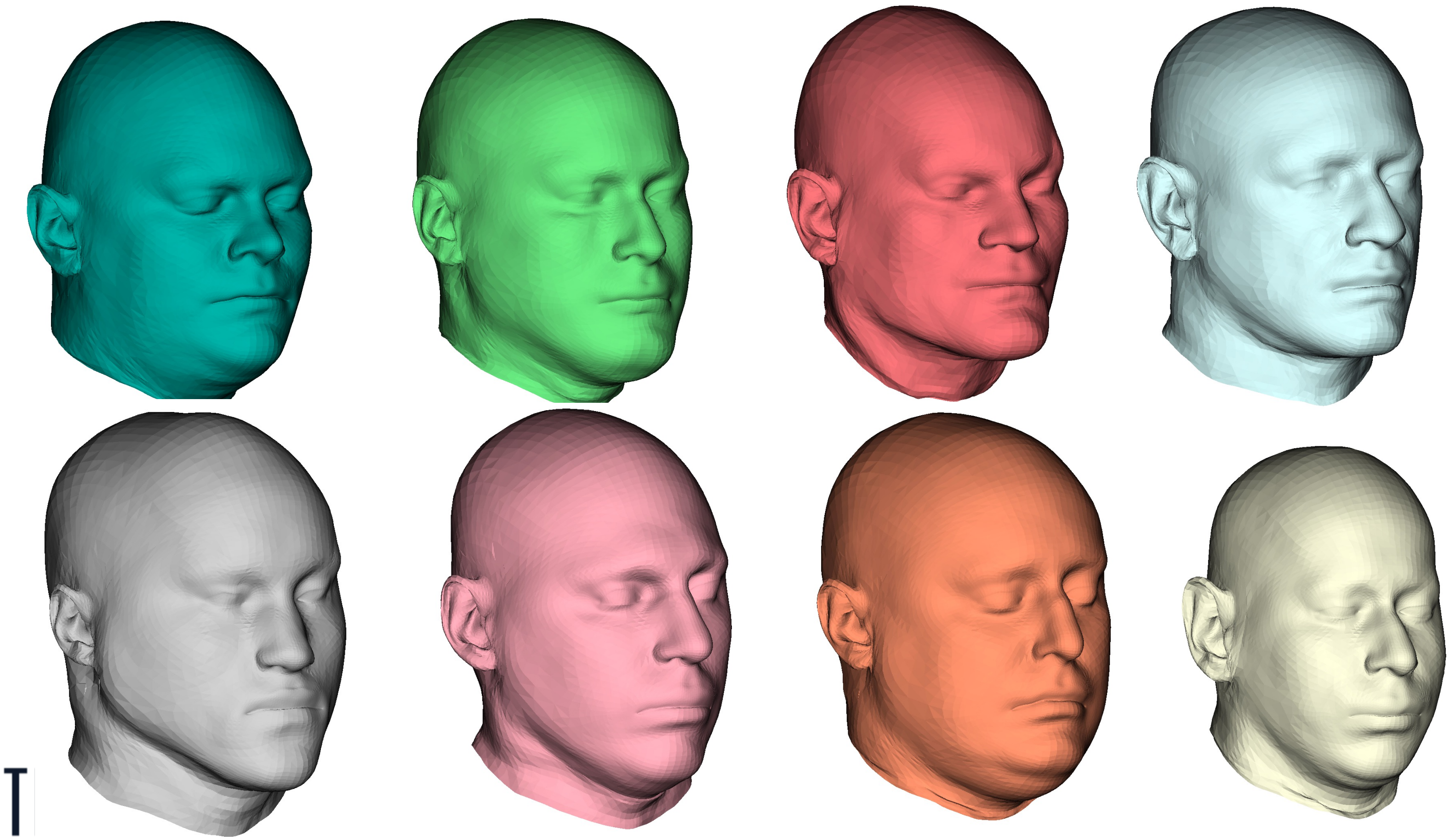
Problem: privacy issues greatly restrict the use of head/face scan data



M | UMTRI

De-Identification through Synthesis

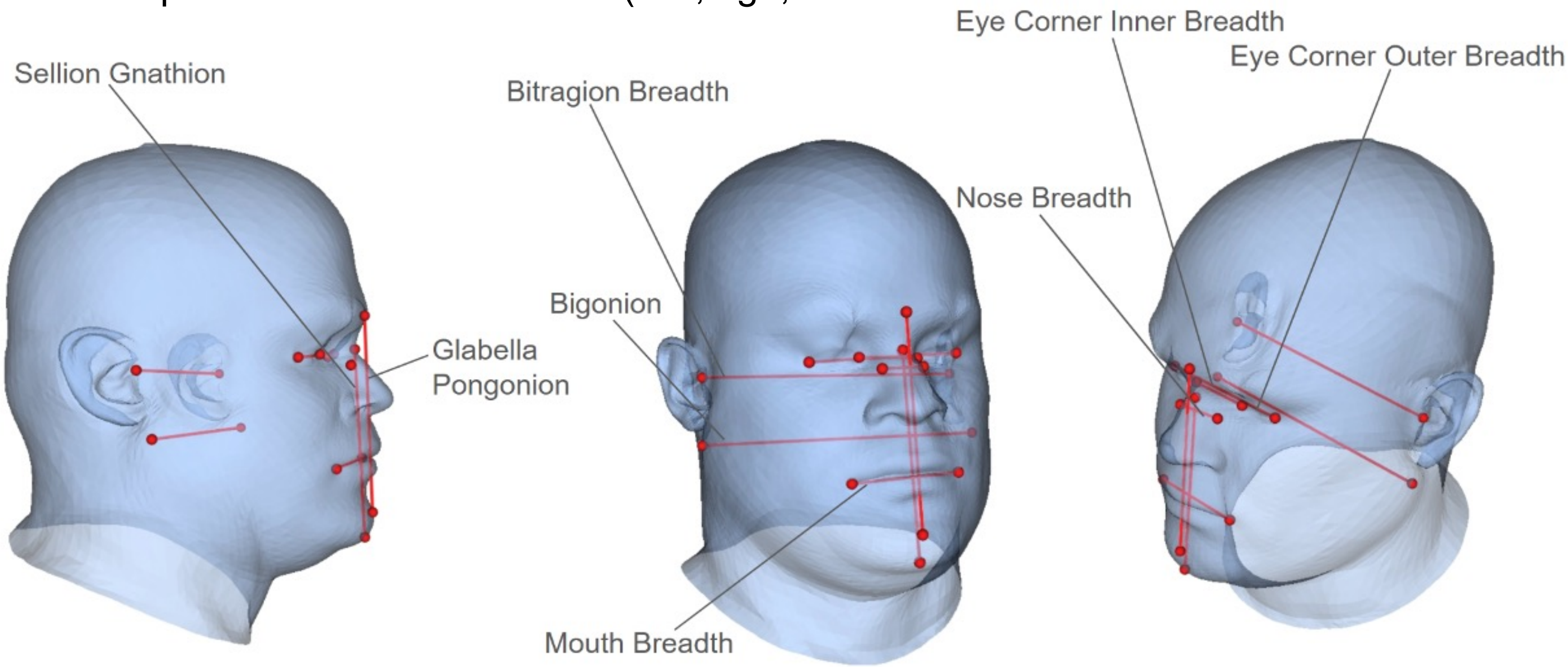
Synthesized versions of the same subject (none of these is an actual person)



M | UMT

Head Synthesis Validation

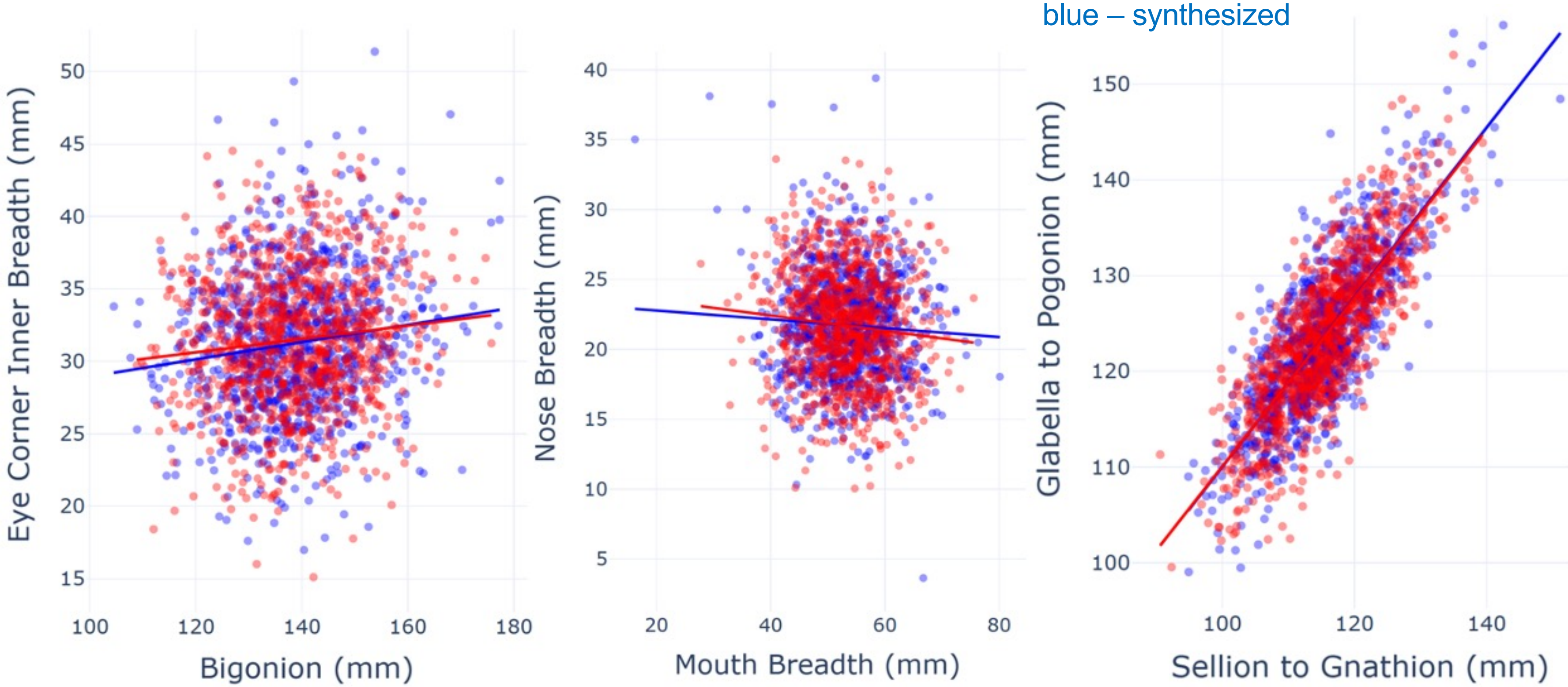
But does it work? The method must retain:
1. Summary statistics on head dimensions for population
2. Relationships among dimensions
3. Relationships with external covariates (sex, age, etc.



Head Synthesis Validation

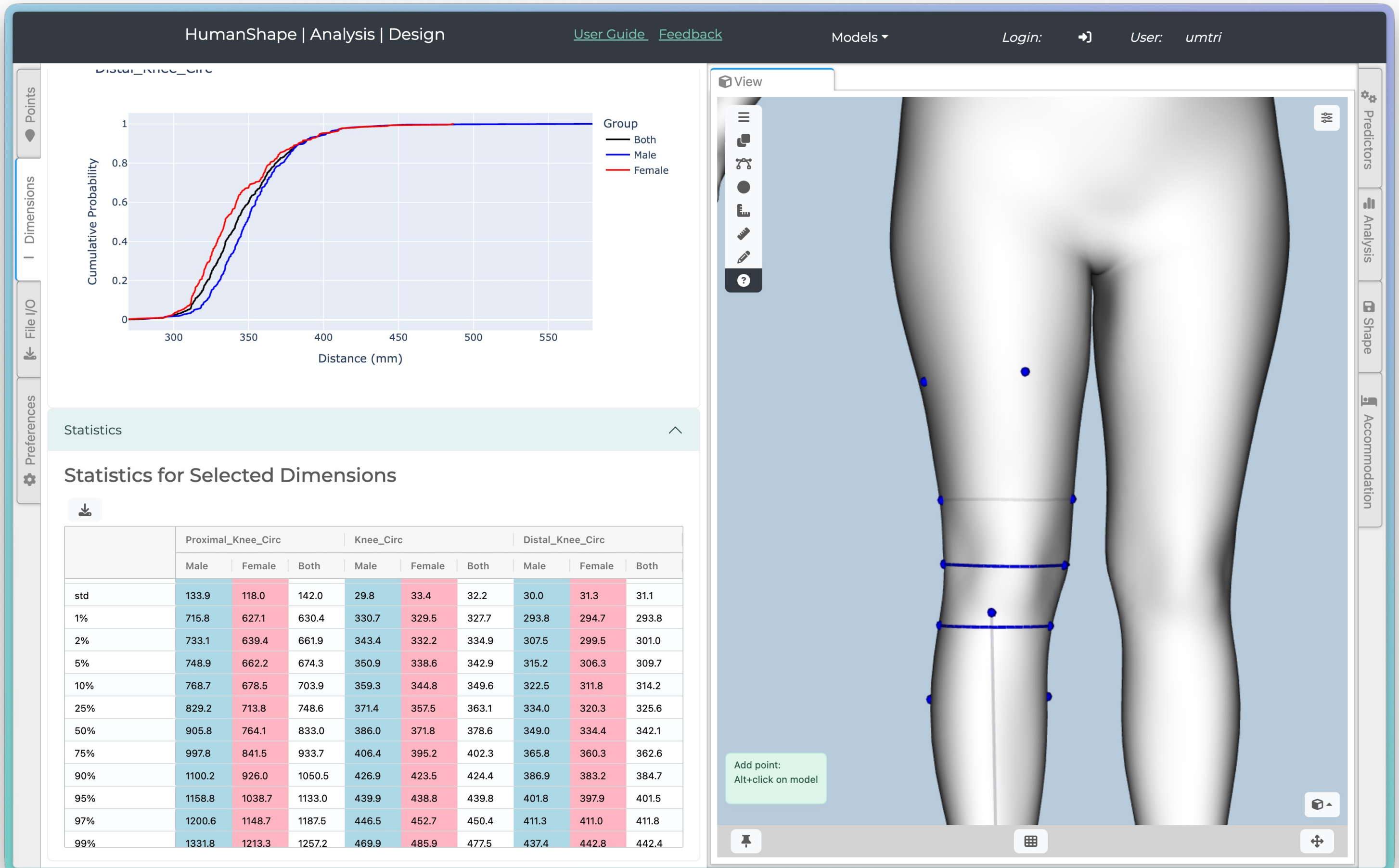
Statistical analyses for de-identified population are essentially identical

red – original
blue – synthesized



Case Study: Knee Brace

Extract novel measures across the population

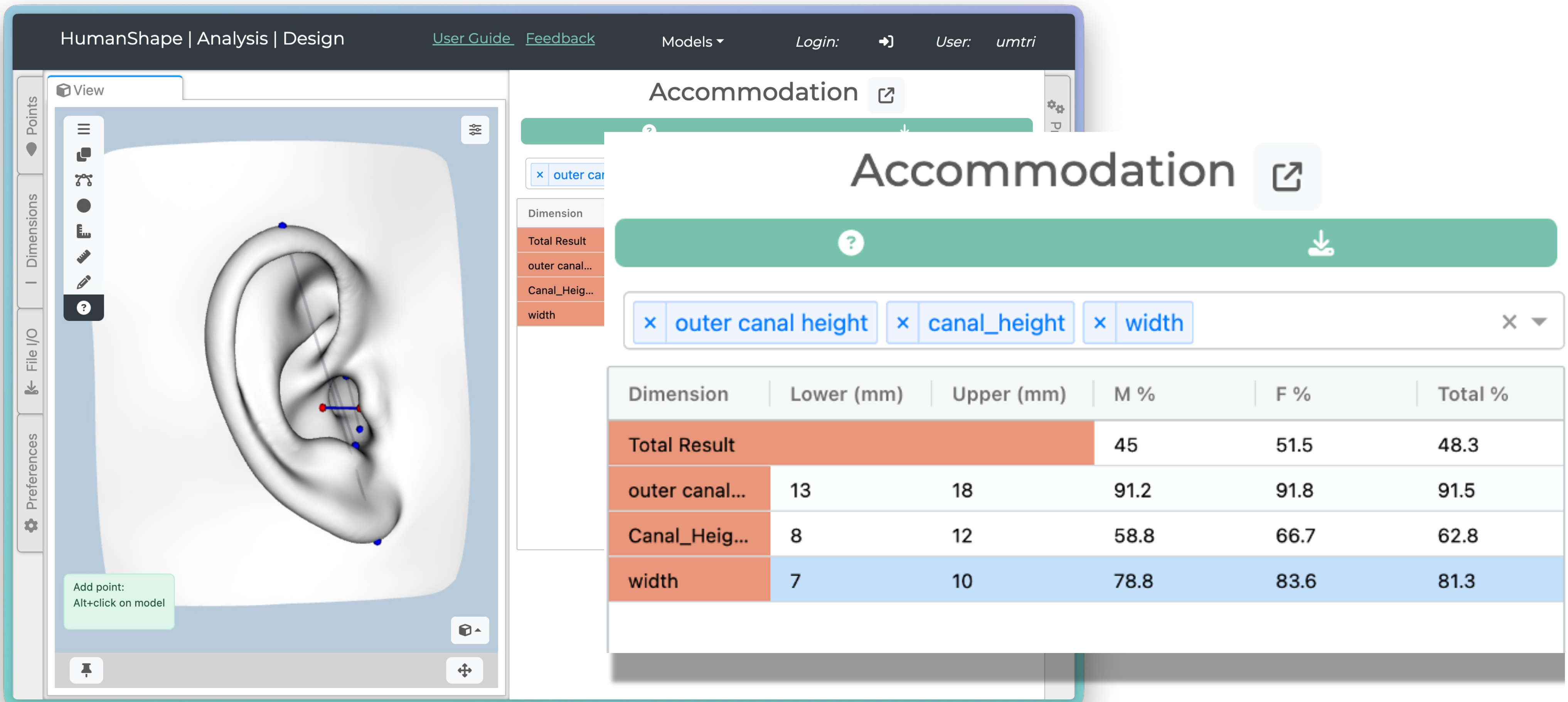


	Proximal_Knee_Circ			Knee_Circ			Distal_Knee_Circ		
	Male	Female	Both	Male	Female	Both	Male	Female	Both
std	133.9	118.0	142.0	29.8	33.4	32.2	30.0	31.3	31.1
1%	715.8	627.1	630.4	330.7	329.5	327.7	293.8	294.7	293.8
2%	733.1	639.4	661.9	343.4	332.2	334.9	307.5	299.5	301.0
5%	748.9	662.2	674.3	350.9	338.6	342.9	315.2	306.3	309.7
10%	768.7	678.5	703.9	359.3	344.8	349.6	322.5	311.8	314.2
25%	829.2	713.8	748.6	371.4	357.5	363.1	334.0	320.3	325.6
50%	905.8	764.1	833.0	386.0	371.8	378.6	349.0	334.4	342.1
75%	997.8	841.5	933.7	406.4	395.2	402.3	365.8	360.3	362.6
90%	1100.2	926.0	1050.5	426.9	423.5	424.4	386.9	383.2	384.7
95%	1158.8	1038.7	1133.0	439.9	438.8	439.8	401.8	397.9	401.5
97%	1200.6	1148.7	1187.5	446.5	452.7	450.4	411.3	411.0	411.8
99%	1331.8	1213.3	1257.2	469.9	485.9	477.5	437.4	442.8	442.4

UMTRI

Case Study: In-Ear Audio

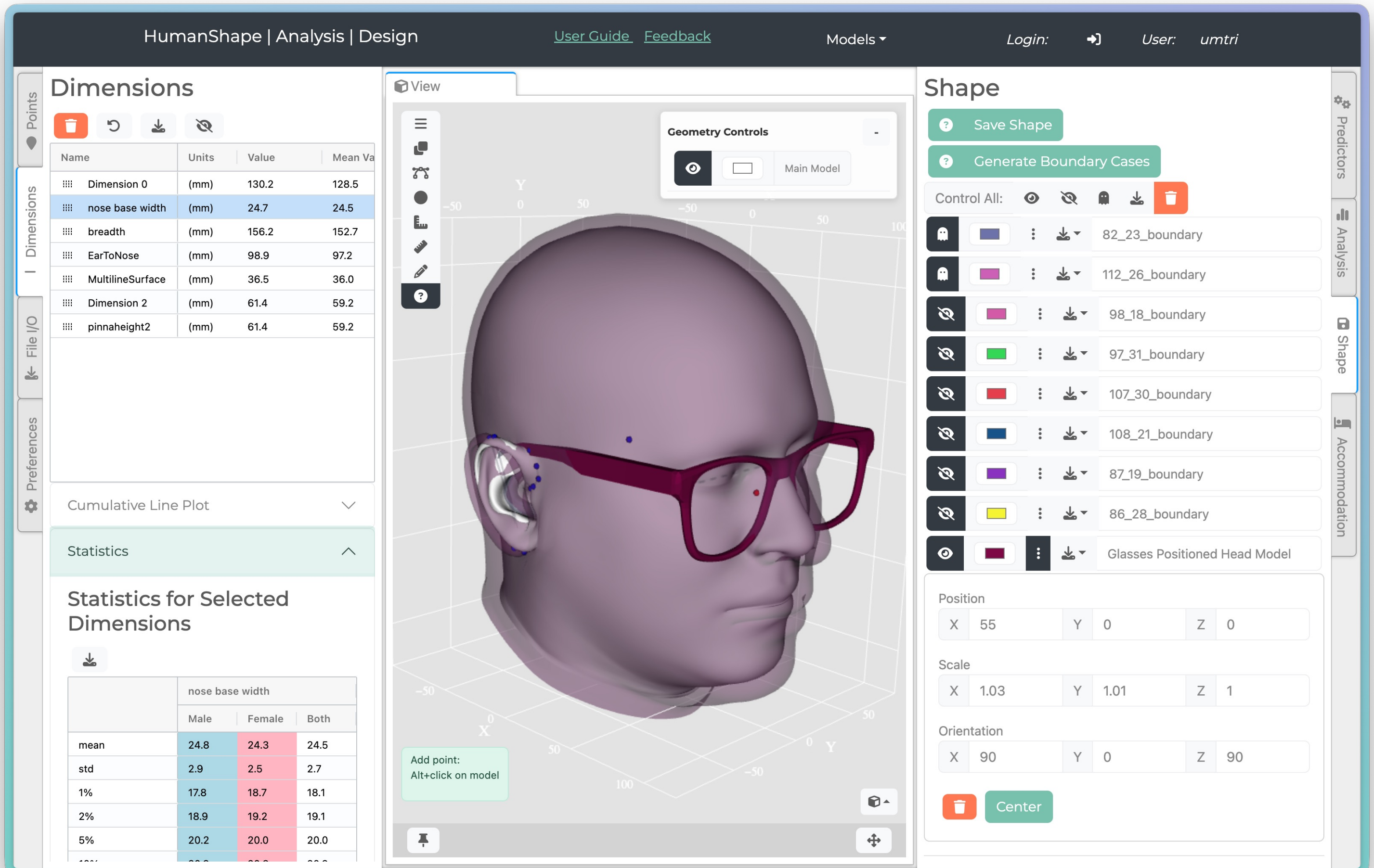
Multivariate accommodation



Dimension	Lower (mm)	Upper (mm)	M %	F %	Total %
Total Result			45	51.5	48.3
outer canal...	13	18	91.2	91.8	91.5
Canal_Heig...	8	12	58.8	66.7	62.8
width	7	10	78.8	83.6	81.3

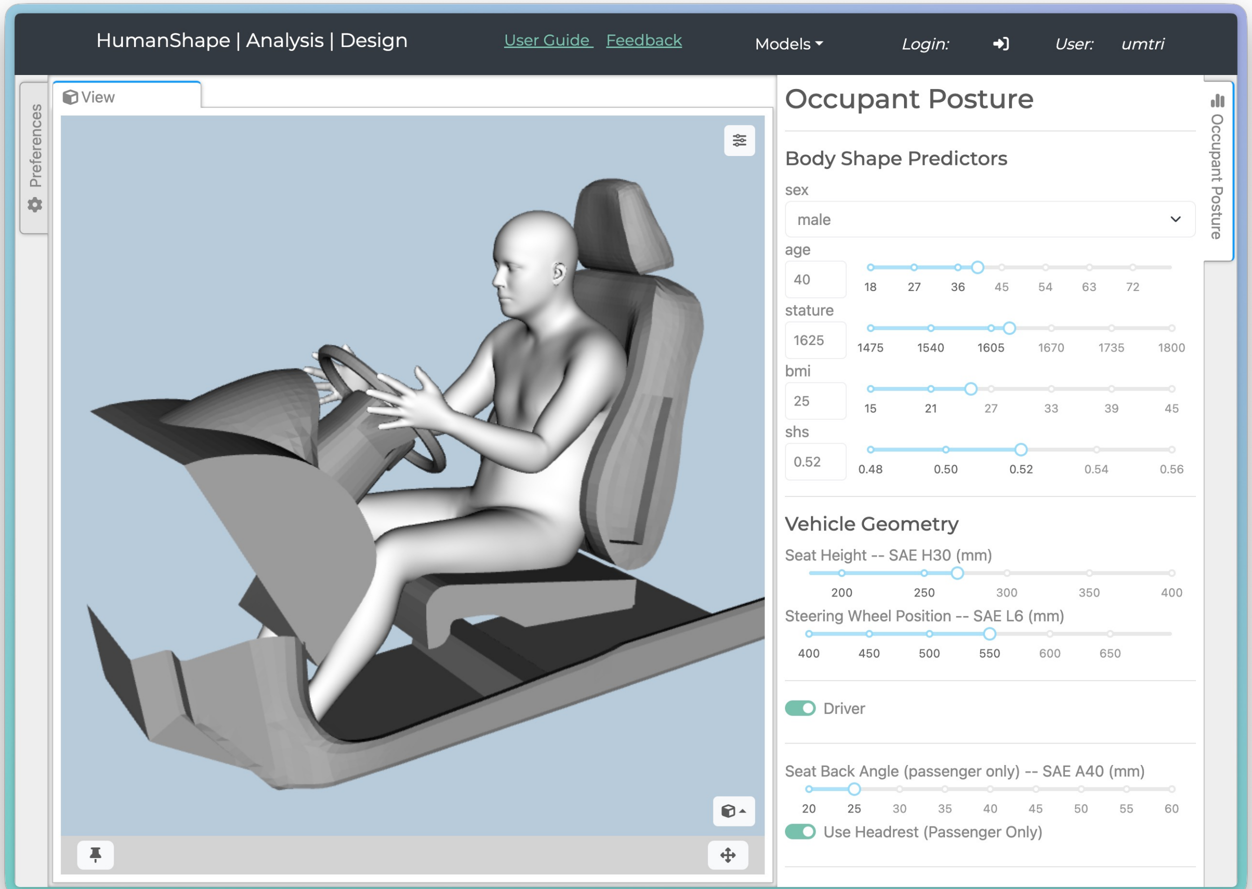
Case Study: Smart Glasses

Explore correlations and create boundary cases using regression or PCA



Case Study: Vehicle Interior Analysis

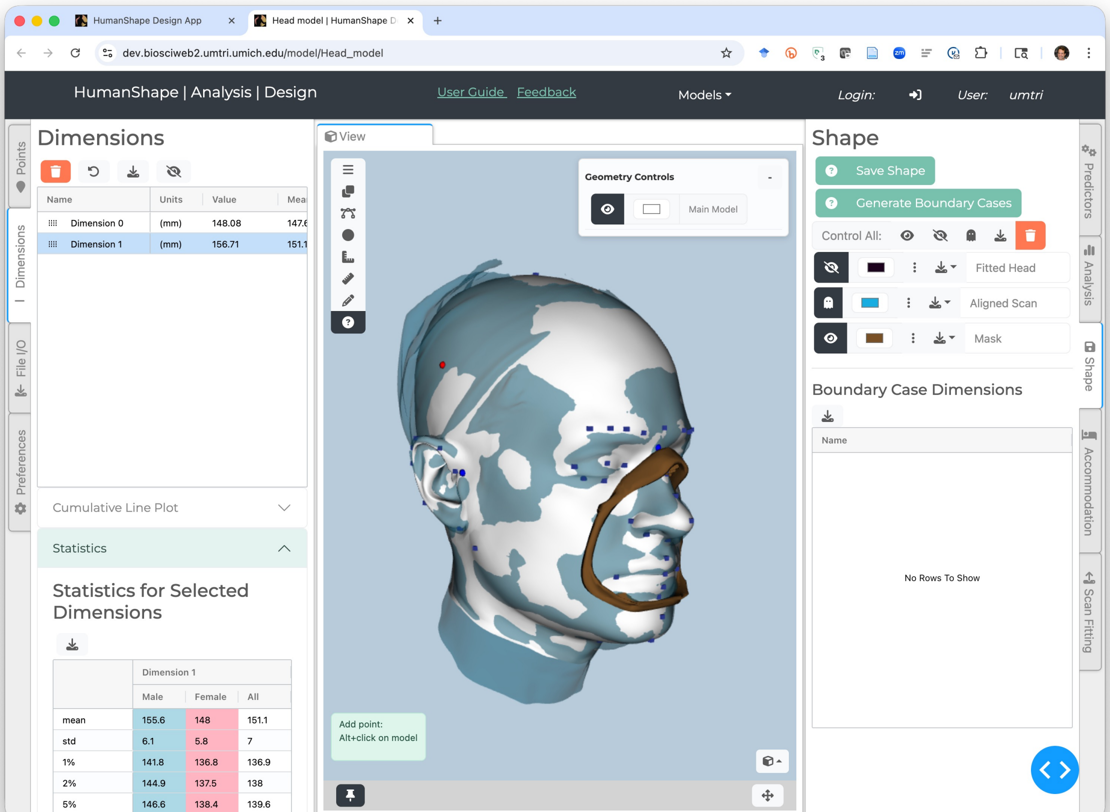
Generate body shapes in realistic postures as a function of vehicle layout



UMTRI

Case Study: Scan-to-Measurement

Upload a head scan, fit it, instantly get landmarks and all predefined measures, measuree anything else of interest, and place it within a population



Bonus: get recommended mask size!

UMTRI

Future Functionality: Parametric Product Models

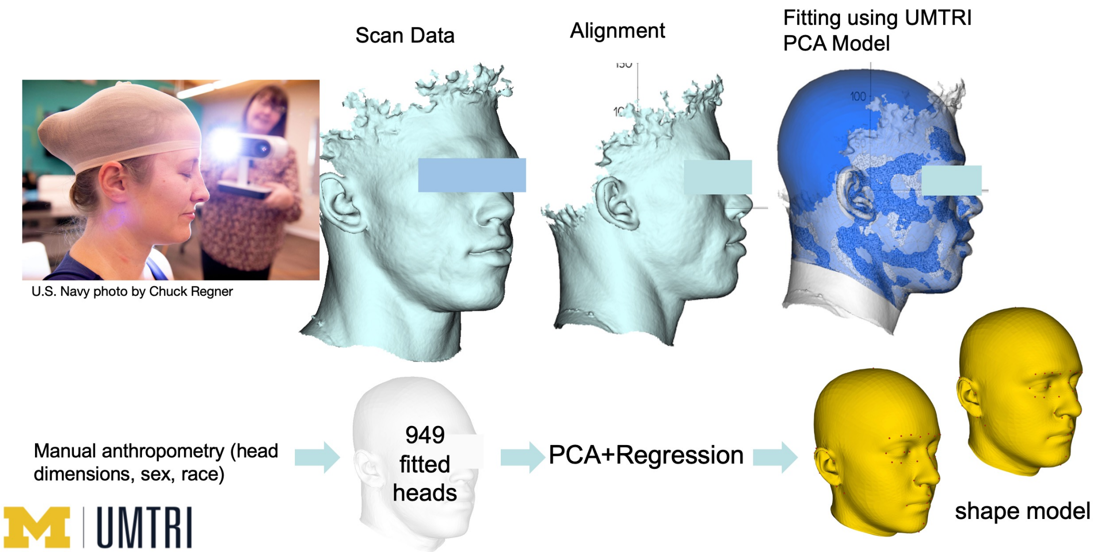
Design wizards for common analyses

Why: Designers should not have to decide which aspects of body size & shape matter for their design, and figure out how to convert between body dimensions and product dimensions

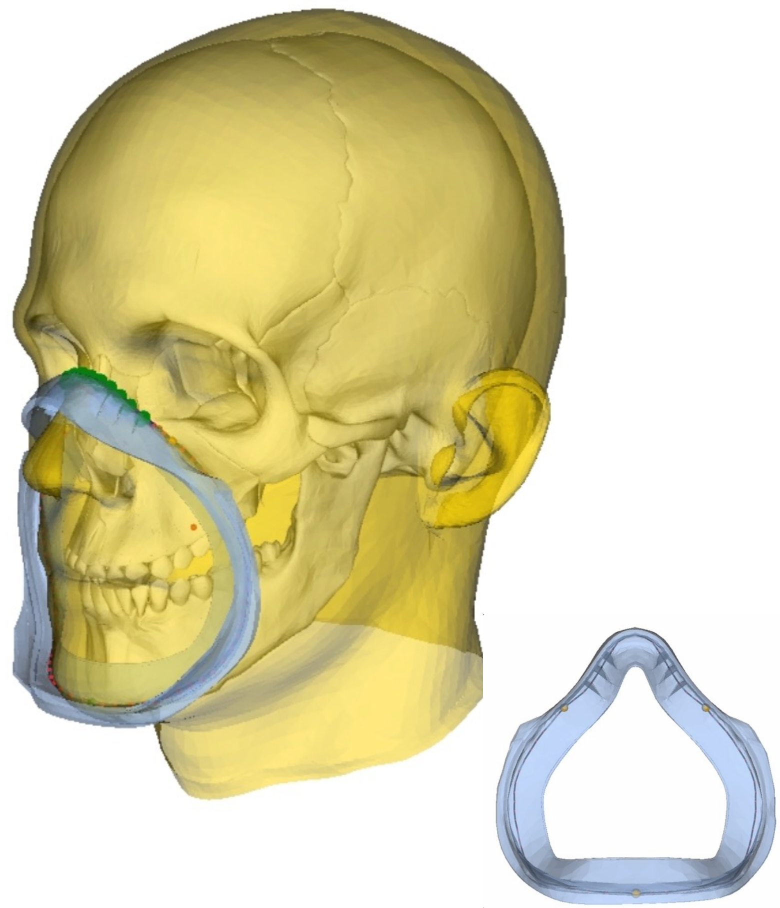
What: Generic, parametric design models for common analyses; input your parameters → get your targets for product **design variables rather than anthropometric variables**

Case Study: Pilot Oxygen Masks (work in progress)

Navy Dataset Data from Lori Basham @ NAVAIR DREAMS lab under CRADA



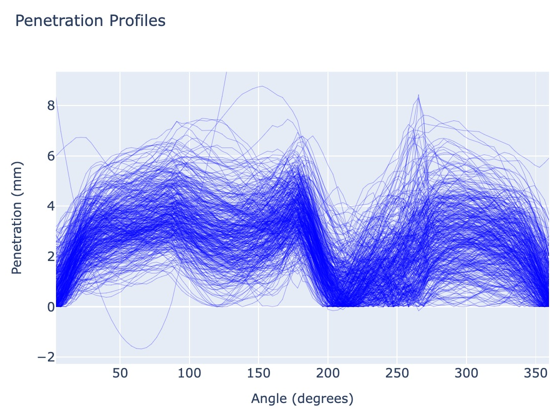
Collaboration with Gentex funded by Navy

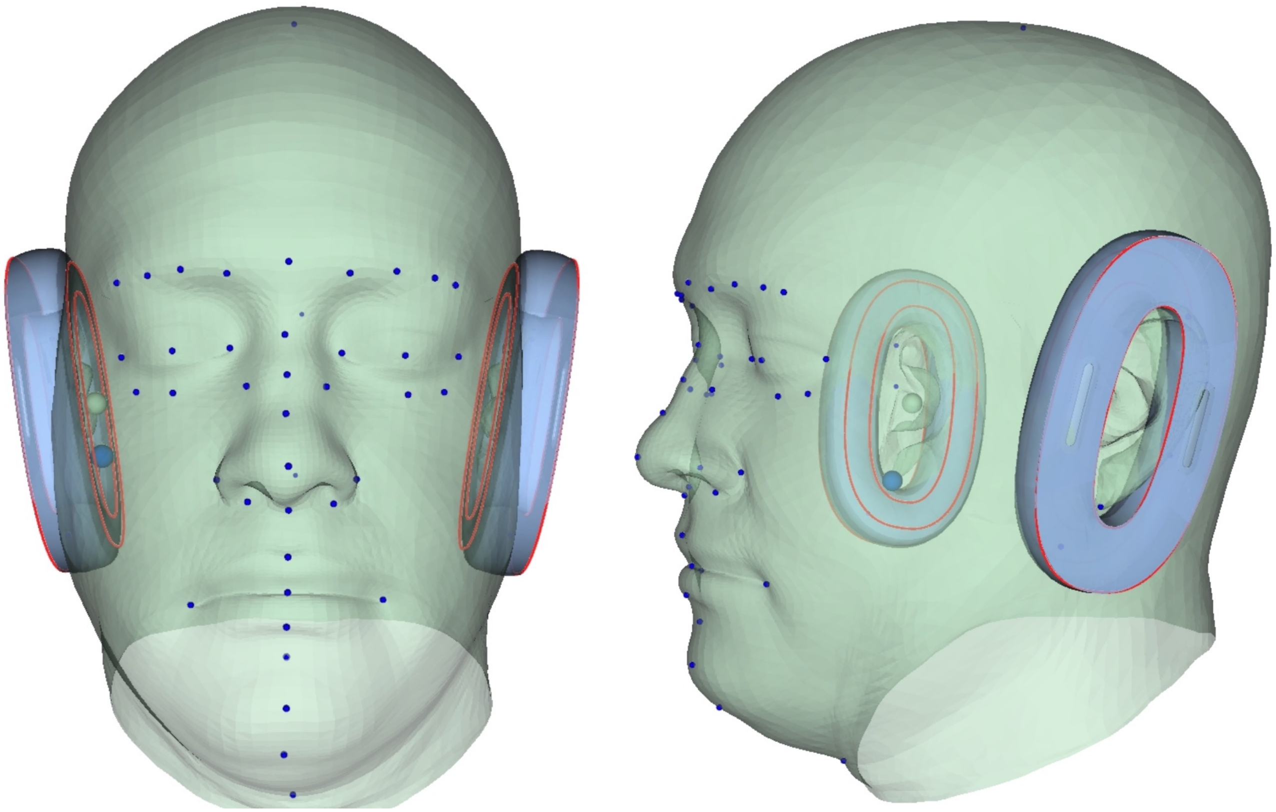


UMTRI

Case Study: Earcups (work in progress)

- Define parametric product model
- Conduct VFT across population
- Optimize design using quantitative metrics





Collaboration with Gentex

What about *my* population?

- We can fit any existing scan data, synthesize a matching, de-identified dataset, and add it to the app (U.S. Navy aircrew heads – thanks Lori!)
- Because we have statistical shape models, we can readily synthesize new populations that lie within shape spaces for which we have data.
- With data from thousands of U.S. and European military and civilian adults 18 to 80 years, and U.S. children ages 3 to 17 years, we can synthesize most “Western” populations of interest, including all U.S. military populations for which we have demographic descriptors (sex, age, race/ethnicity, stature & BMI if available)

Limitations and Future Work

Still an anthropometry tool: needs to be enable the user to work in design variables, with the software creating the link to anthrometry

Posing: our BioHuman avatar is fully posable, but creating a good UI for posing is challenging; posing is necessary, but often creates more problems than it solves (valid posture prediction is much better)

More anatomy: we have linked skeletal geometry, organ geometry available

Many more models: foot, hand with grasp, head/neck with face/neck movement…

Linkage to FE and musculoskeletal models: We can immediately generate runnable models

Product geometry interaction: users can upload polygonal geometry, but the software doesn't play well with native CAD geometry; CAD system implementations?

Customized populations:

- currently no interface; requires us to generate the model (but this is programmatic)
- legal roadblocks due to addressable privacy concerns

UMTRI

Acknowledgements

Special thanks to Lori Basham and team at U.S. Navy NAWC-AD

Mask research is a close collaboration with Gentex Corporation

This research was supported by multiple organizations and collaborators, including:

Amazon Inc.
Gentex Corp.
Meta Platforms, Inc.
U.S. Army
U.S. Navy
U.S. Air Force
U.S. Marine Corps
SENTIR Labs (STI-TEC)

Matthew Reed, PhD

mreed@umich.edu

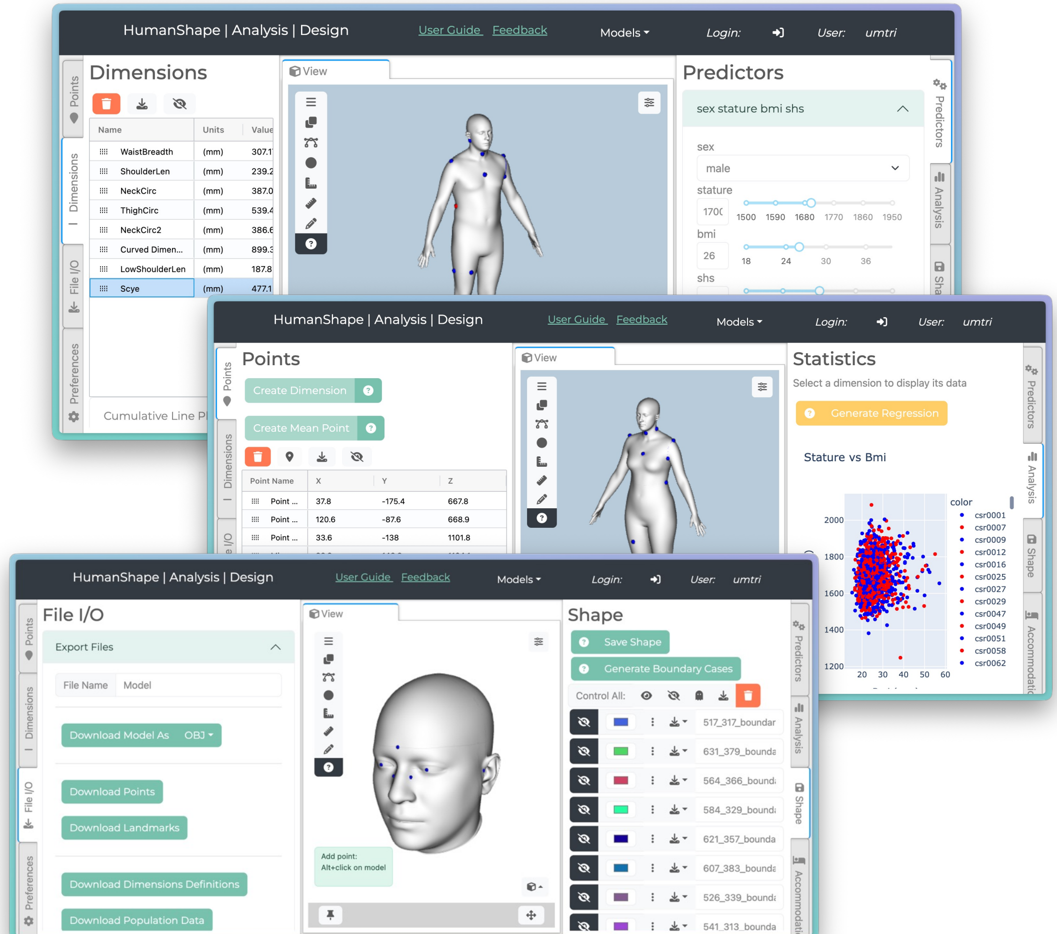
http://mreed.umtri.umich.edu

UMTRI

HumanShape App: Online Human-Centered Design

- Unified interface for working with all types of human body shape data, from whole-body to anatomical regions: hands, heads, ears, feet
- Online app means no software installation problems and accessible from anywhere
- Working with large populations of virtual avatars, not raw scan data
- Synthesis techniques ensure accuracy while eliminating privacy-related constraints

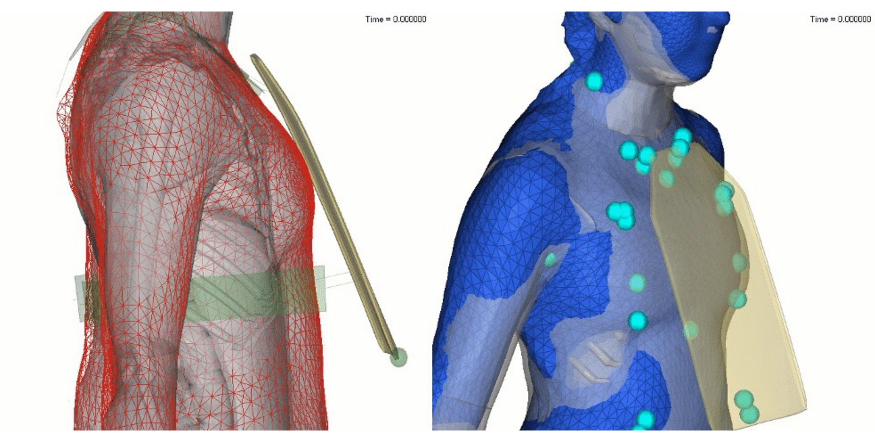
https://HumanShape.app/



UMTRI

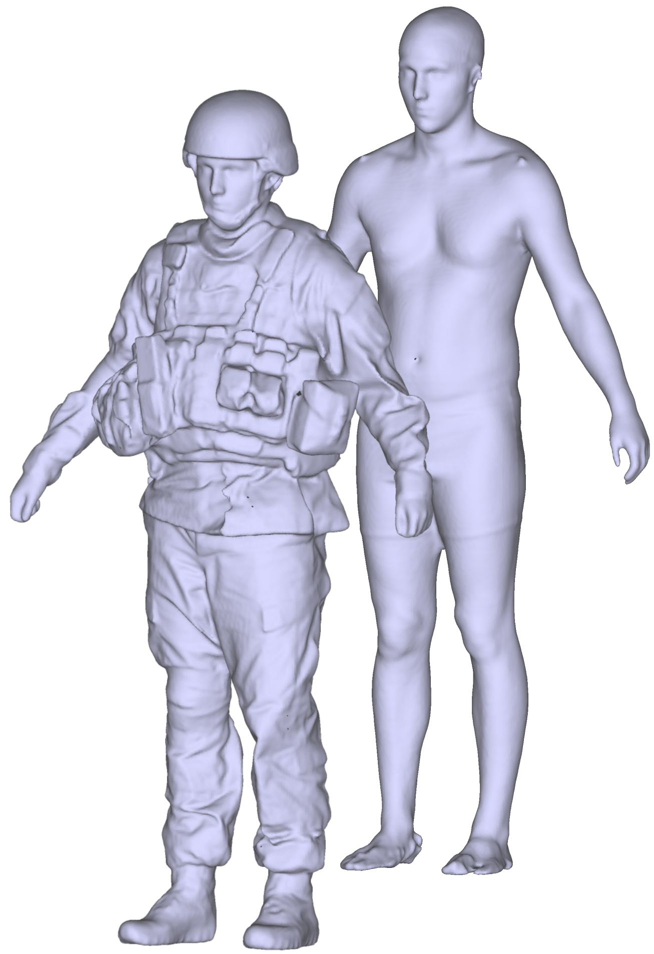
Additional Applications

Anything that fits close to the body:
- Glasses and goggles
- Head-worn displays
- Hearing protection
- Body armor
- Gloves
- Splints and other medical devices



Simulated plate interaction with parametric human model

UMTRI



Scans from UMTRI studies