25 years of Lessons Learned from the Forensic Data Bank and Fordisc

Data Standards, Archiving, and Analytics in Forensic Anthropology

MERCYHURST UNIVERSITY

Stephen Ousley
Department of Computing and Information Science

Financial Disclosure

I programmed and continue to program Fordisc 3

I receive occasional payments from the University of Tennessee for Fordisc work, travel, and expenses

I made MUCH more money last year from Apple and Microsoft (and Google and Amazon) stock than from Fordisc

My Timeline

I started at UT in 1989

Fordisc 1.0: 1993 (True Basic) DOS

Fordisc 2.0: 1995 (Visual Basic) Windows 95

PhD 1997; I joined Smithsonian in 1998

Fordisc 3: 2005 (Delphi) Windows 98+

I went to Mercyhurst in 2007 (Anthropology/Archaeology and Applied Forensic Sciences Departments)

I am now in the Computing and Information Science Dept

Lessons Learned from Data Standards

1. Repeatedly check current observational standards for accuracy and clarity. They can always be improved.

Challenge: Momentum

- we are all most comfortable using familiar methods
- we need as much continuity with past data
- we need to incorporate improvements

UT Standards
Data Collection Procedures (DCP)

2nd edition: Moore-Jansen and Jantz (1989)

3rd edition: Moore-Jansen, Ousley, Jantz (1994)

DCP 1.0, the basis for the *Standards* (1994)

Craniometrics:

Martin’s standards → Howells (1973)

Rudolf Martin (1864-1926)

- Anatomist
- German, worked in Zurich and Munich
- *Lehrbuch der Anthropologie* (1914, 1928, 1957)
- Martin and Knußmann (1988)
- Established osteometric standards

Measurement Standards based on Martin's *Lehrbuch*

"It is unnecessary, therefore, **in view of the universal recognition of and general acquaintance with Martin's *Lehrbuch***, to do more than indicate that the technique employed in this study is based on his definitions." (Shapiro 1929:1)

Martin's Ectoconchion

Derjenige Punkt an der Umschlagkante des lateralen Orbitarandes, an welchem eine zum Oberrand der Augenhöhle parallel verlaufende Gerade auf jenen Rand- etwa in dessen Mitte- trifft. Um die Umschlagkante genau zu finden, d.h. den Punkt nicht zu tief nach innen noch zu weit nach augen an den Orbitarand zu legen, halte man den Schädel in der *Norma basilaris* - mit dem Gesichtsschädel nach oben - vor sich hin und drehe ihn so weit gegen sich, bis man den äußeren linken Orbitarand in starker Verkürzung sieht. Mit einem Bleistift fahre man dann flach über den linken äußeren Orbitarand und markiere so die Umschlagkante. Das Ektokonchion ist dann der Schnittpunkt dieser Kante mit einer parallel zum Oberrand der Orbita verlaufenden Geraden, die die Orbita in eine obere und untere Hälfte teilt. Dies kann durch Augenmaß genau genug beurteilt werden.

(Martin and Knußmann 1988:167)

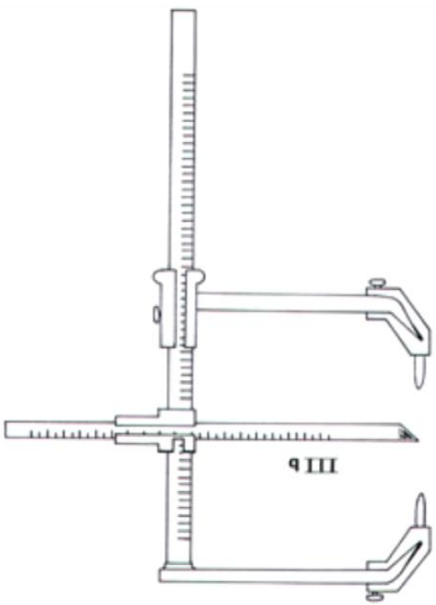
Martin's Ectoconchion

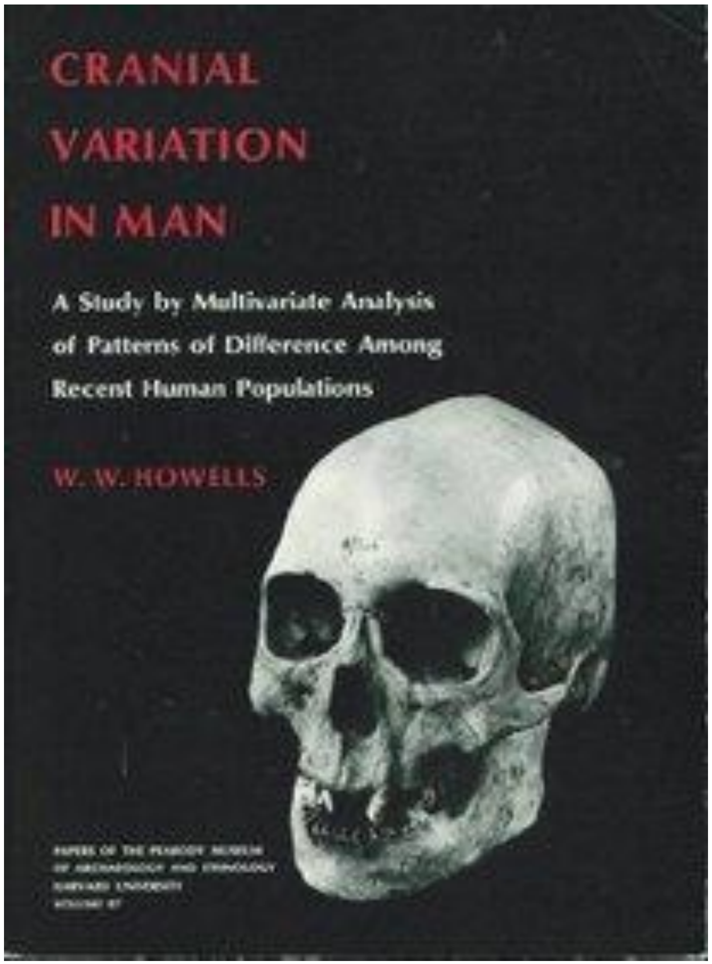
The point on the edge of the lateral orbital margin on which a line parallel to the superior orbital border intersects at about the middle. In order to find the edge not too far in and not too far out on the lateral orbital margin, you hold the skull in front of you, *Norma Basilaris*, with the face pointing upwards, and rotate it until the lateral left orbital margin looks greatly shortened. Mark the edge by moving a pencil flatly along the outer left orbital margin. Ectoconchion is the intersection of the marked edge with a line parallel to the superior orbital border that divides the orbit into equal upper and lower halves. This can be determined well enough by eye.

(Martin and Knußmann 1988:167)

William White Howells (1908-2005)

Refined and redefined craniometric landmark definitions and established worldwide database (1973)





Howells' Ectoconchion

The intersection of the most anterior surface of the lateral border of the orbit and a line bisecting the orbit along its long axis. Mark both sides with a pencil.

Hold the flat of a pencil lead so that this surface is perpendicular to the median plane of the skull, i.e., tangent to the most anterior curvature of the orbital margin, and use it to draw a line along this crest. Turn the skull so as to be able to sight along the long axis of the orbit, however oblique, and to bisect the orbit visually, with the pencil as a sighting guide. Make a tick with the pencil where this axis appears to intersect the line already made.

(Howells 1973: 168)

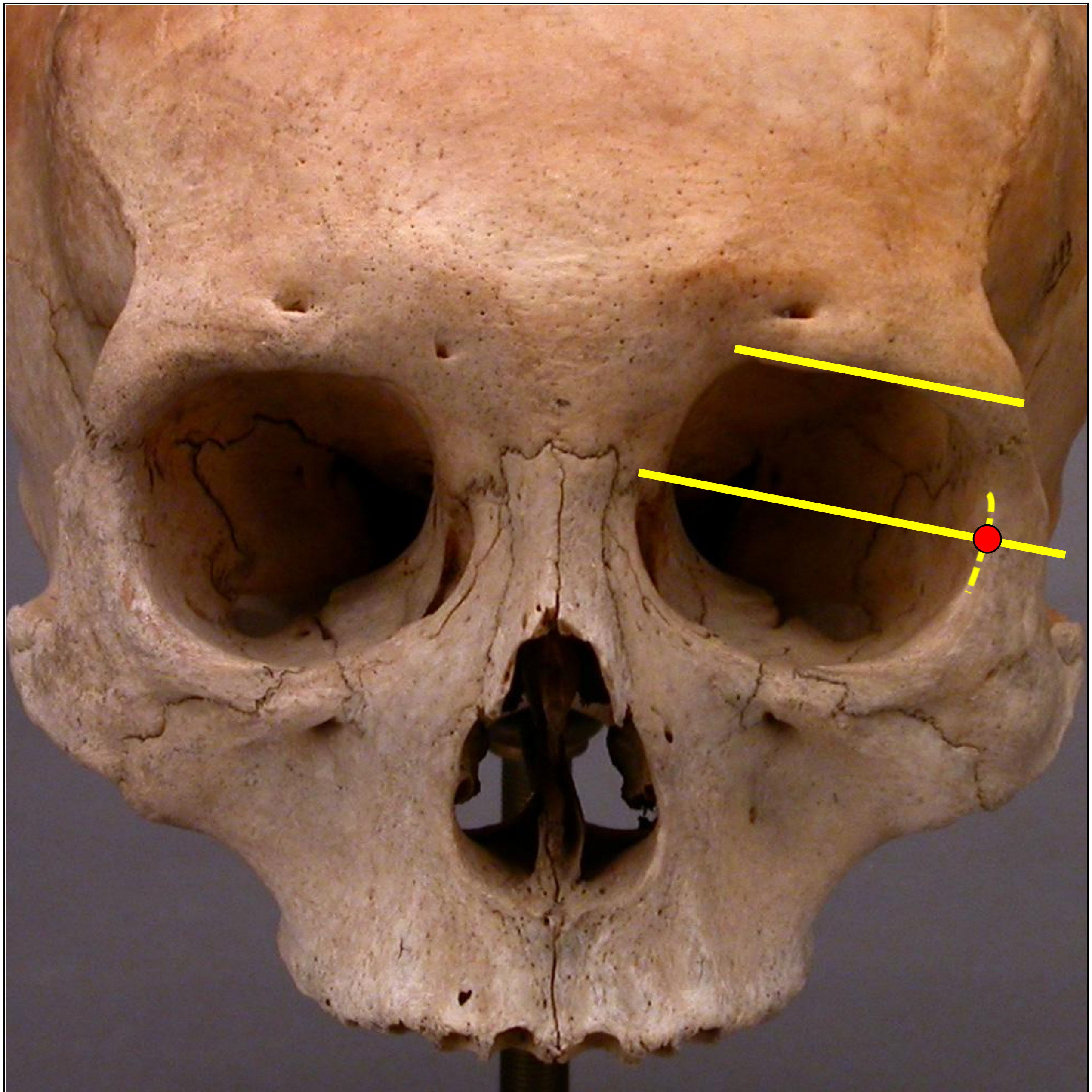
Howells (re)defined the lateral "rim" where ectoconchion is found

The most anterior point on the lateral orbital border.

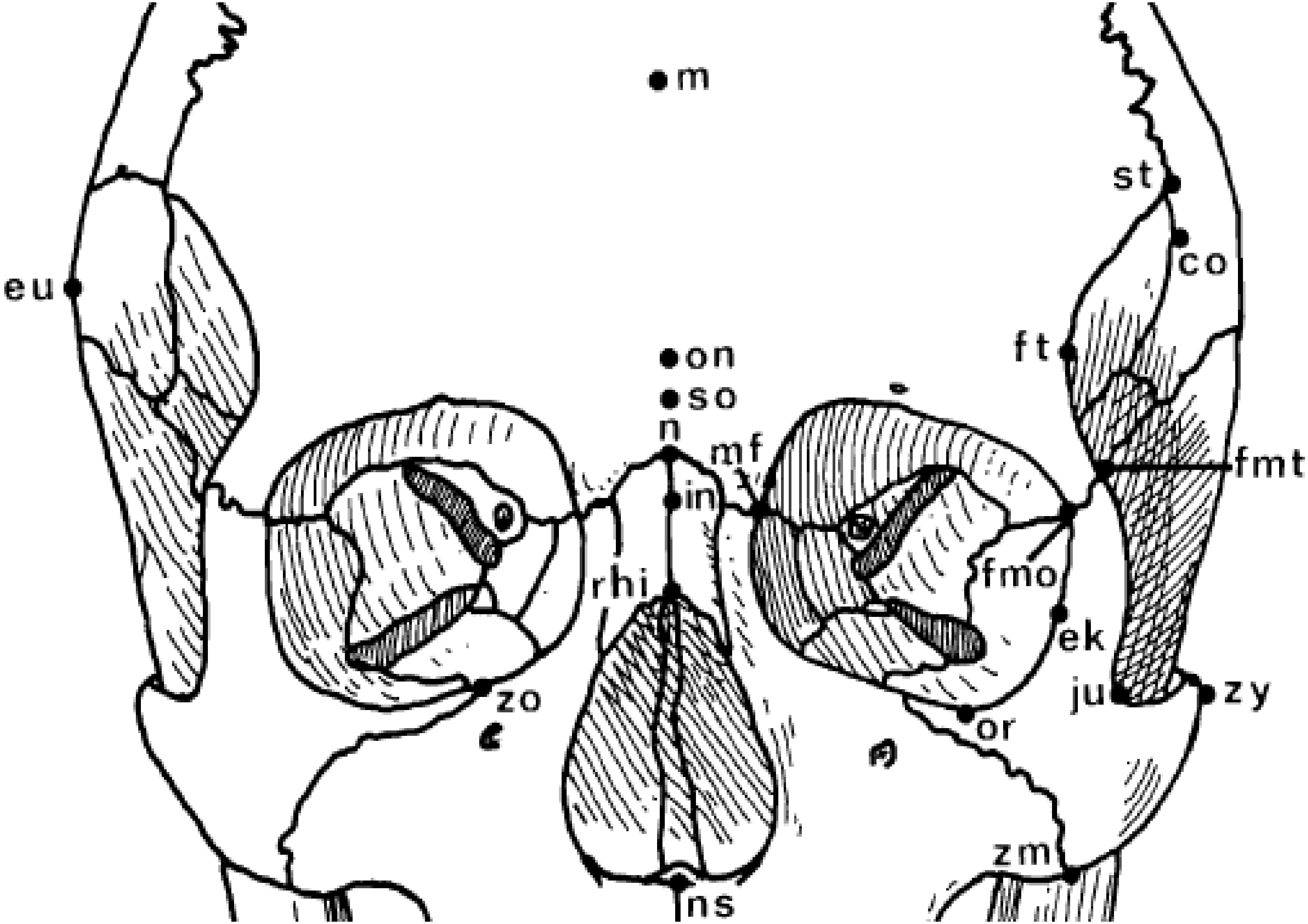
Martin and Knußmann (1998) say the locations differ, but when the orbital rim is sharp they are the same point.

Howells' definition is much more clear!

Ectoconchion



Ectoconchion in Martin's *Lehrbuch*



Martin (1928)

Illustrations are also important: *Standards* has errors

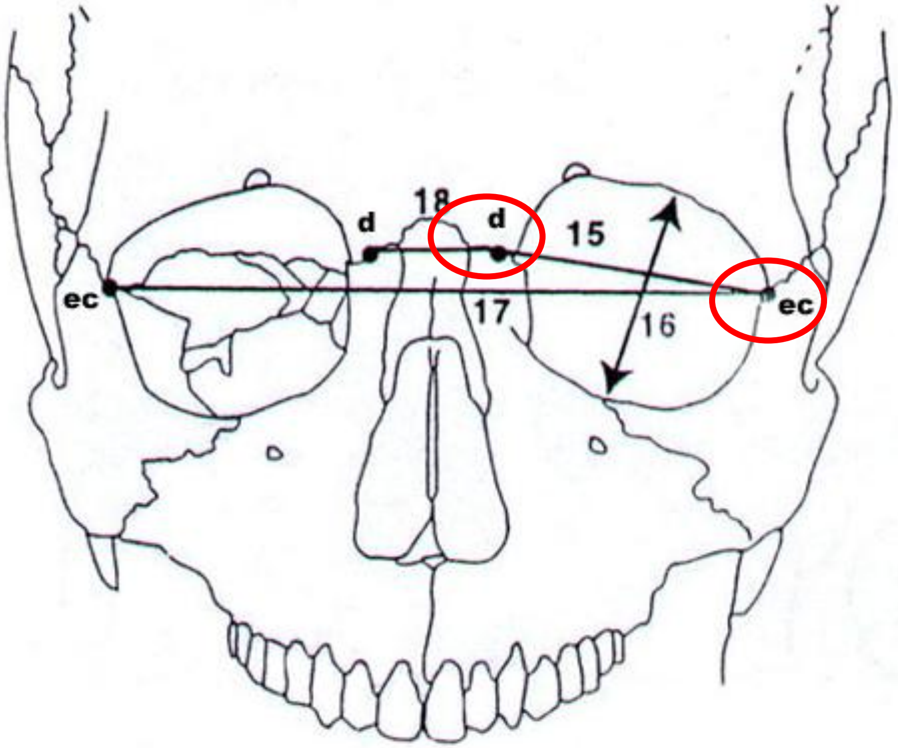
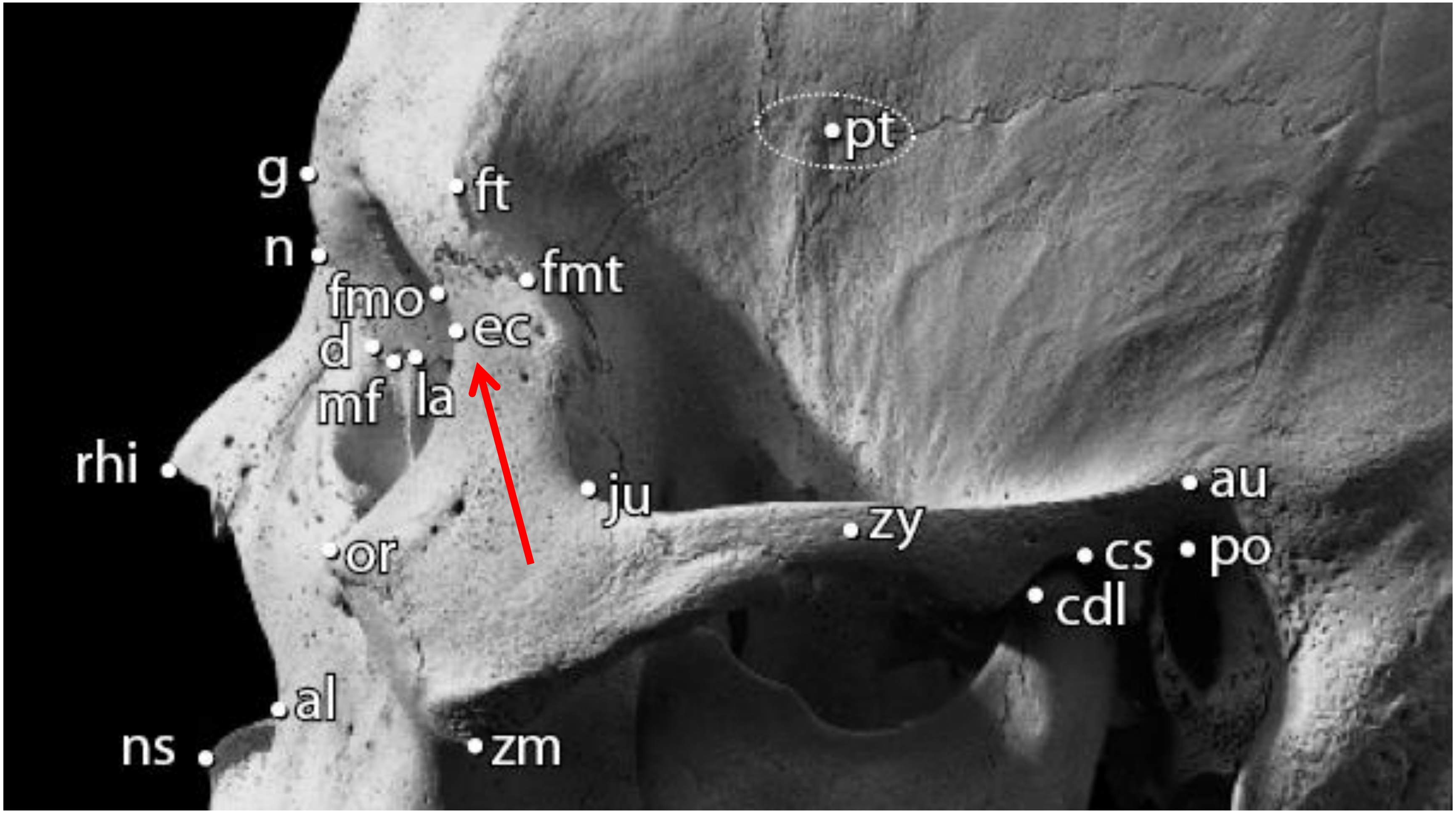


Figure 41: Measurements of the orbital region (after Moore-Jansen et al. 19

Both dacryon and ectoconchion are incorrect here

Ectoconchion in Steele and Bramblett (1988)

Errors in *Human Osteology* (White et al. 2012)



"Ectoconchion is instrumentally determined as the most lateral point on the orbital margin." p.58

Aurel von Török (1842-1912)

Hungarian, University of Budapest

Grundzüge einer systematischen Kraniometrie (Foundations of a Systematic Craniometry) (1890)

von Török (1890)

Defined almost 5,400 cranial measurements

Verhältniszahlen des Gesichtsschädels. 217

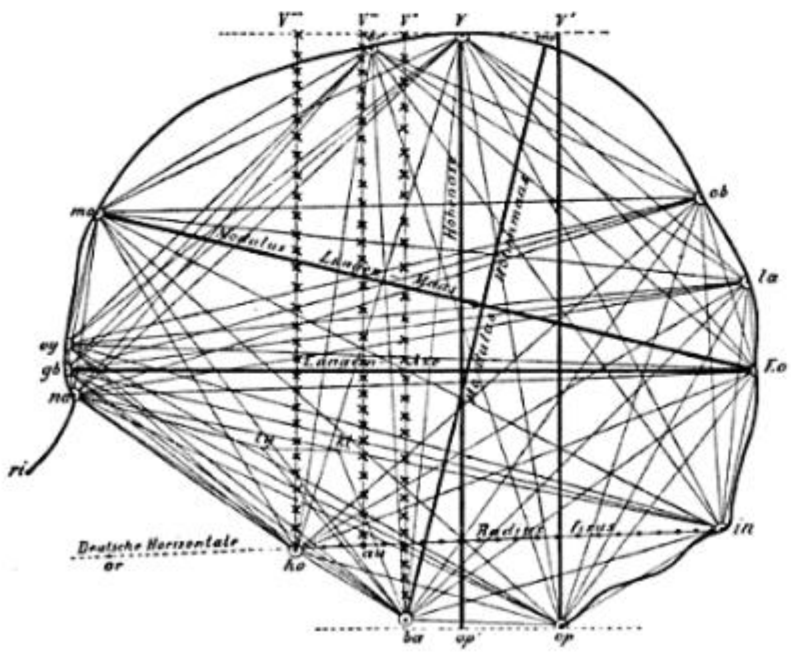
5365—5366. Die lineare Entfernung zwischen dem tiefsten Punkt der Facies orbitalis des Jochbeines und dem Zygomaxillare beiderseits.

5367. Die temporale Orbitalenge, d. i. die lineare ektoorbitale Breite zwischen der beiderseitigen lateralen Orbitalwandung.

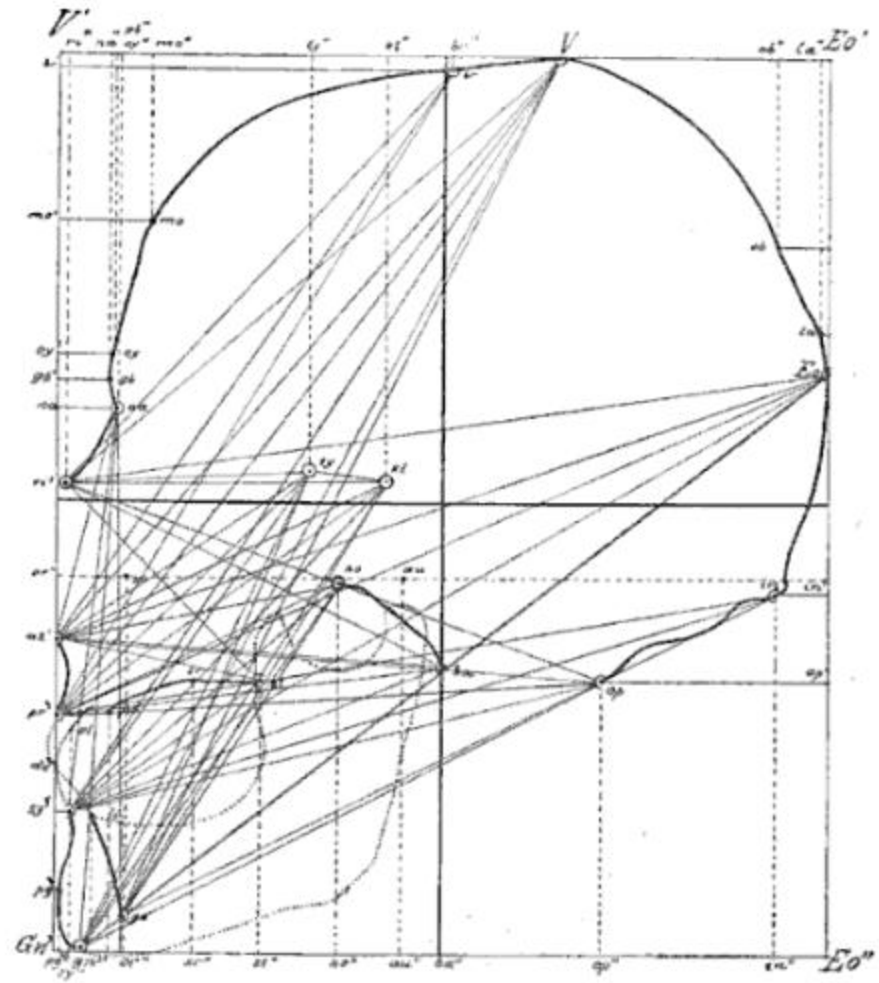
5368—5369. Die Länge des Sulcus lacrymalis zwischen dem Dakryon und dem tiefsten Randpunkte der orbitalen Oeffnung des Canalis nasolacrymalis beiderseits.

5370—5371. Die grösste Breite des Sulcus lacrymalis zwischen der Crista lacrymalis anterior und posterior beiderseits.

von Török (1890)



Tafel 16.
Direkte Linearmessungen, sowie Projektionsmessungen in der Medianebene.



Tafel 17.
Das Umrissviereck der Norma mediana und seine Linearmaasse [1]).

Necessary Craniometric Instruments

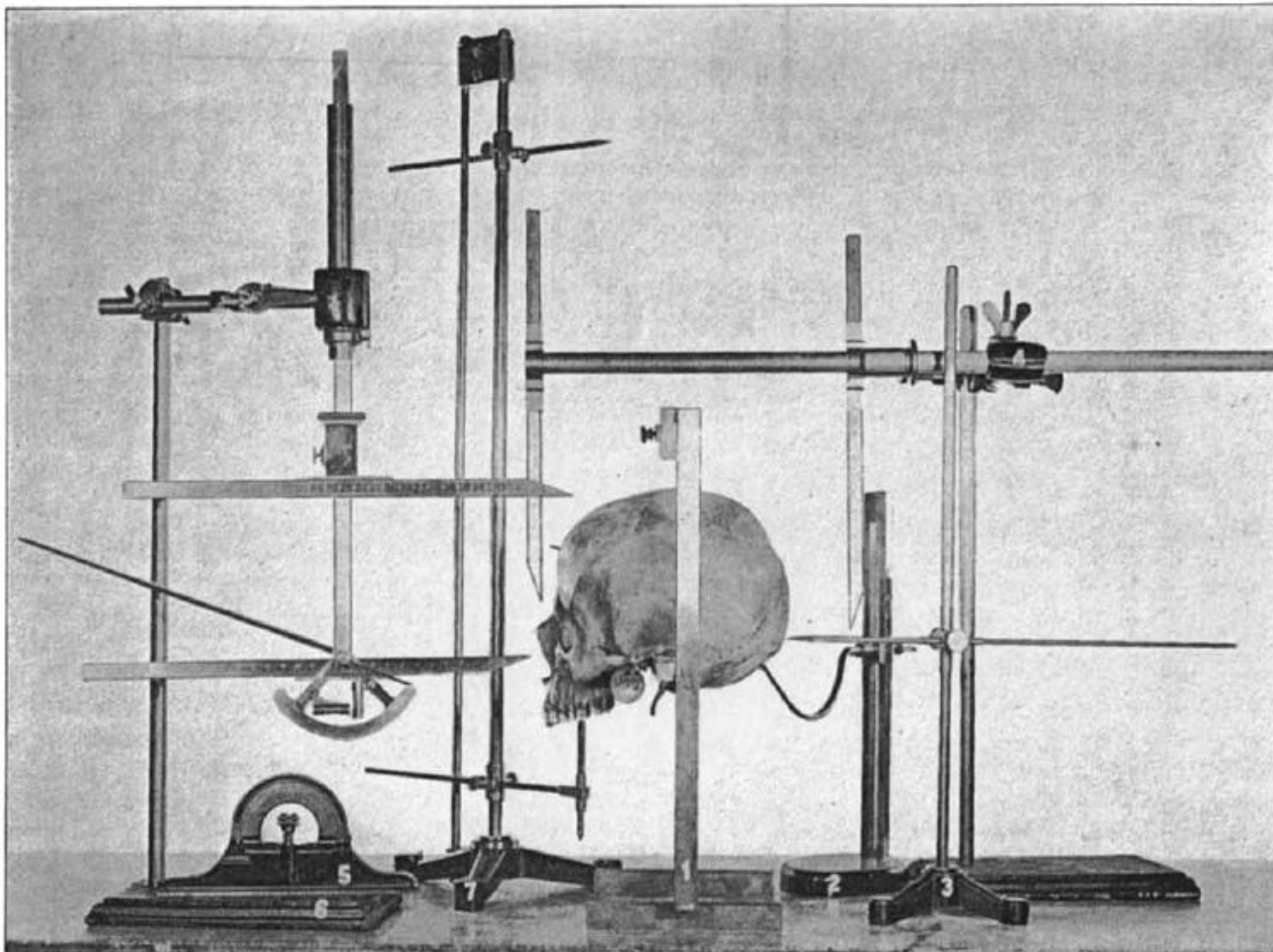


FIG. 1 Apparatus used in measuring lengths and auricular heights of skulls oriented in Frankfort plane, called in this work the "Old Apparatus." 1. Reserve head-frame; 2. Diagraph used to support occiput; 3. Horizontal needle; 4. Stangenzirkel held in jaws of osteophore and used to determine greatest length; 5. Spirit level; 6. Stativgoniometer used for measuring auricular height; 7. Parallelograph as employed to support palate.

Microscribe 3-D Digitizer

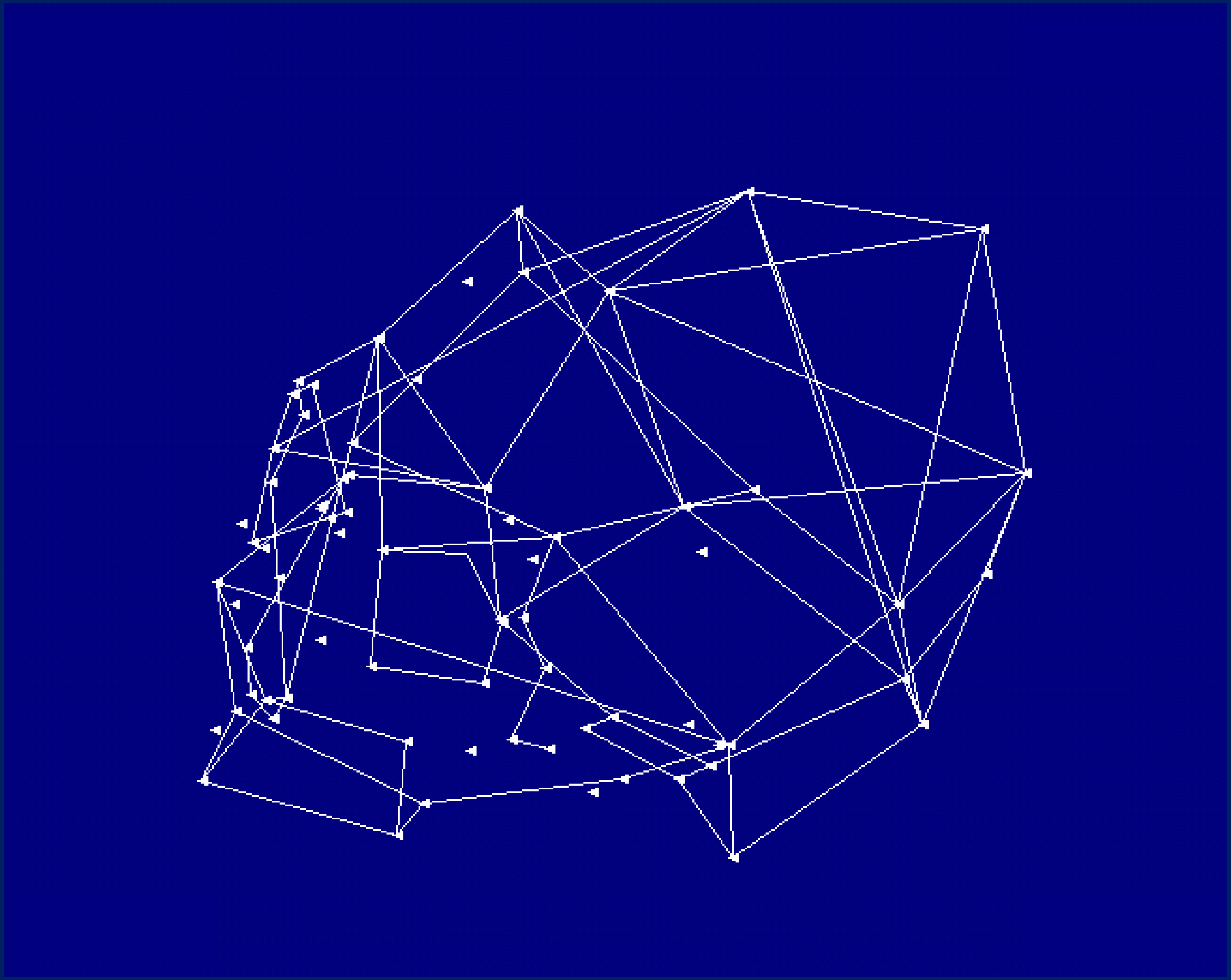
Records x, y, and z landmark coordinates

3Skull Software

Calculates traditional craniometrics

Saves craniometrics and coordinate data in tables

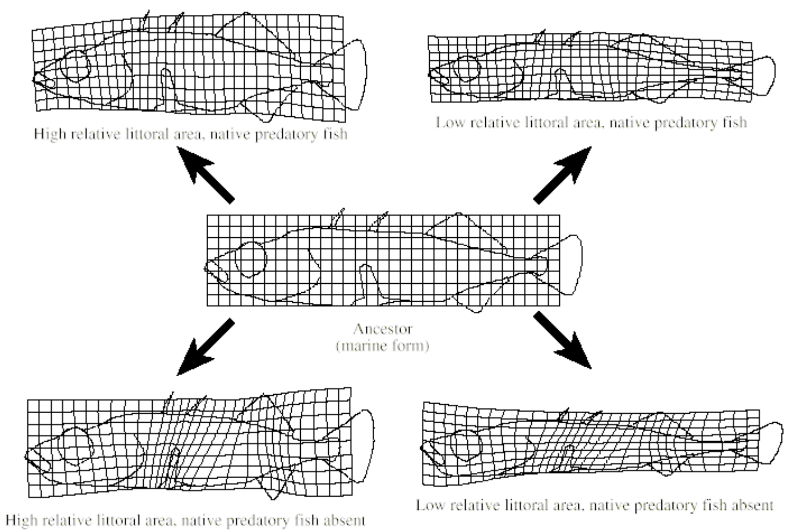
Digitizers record richer morphological data



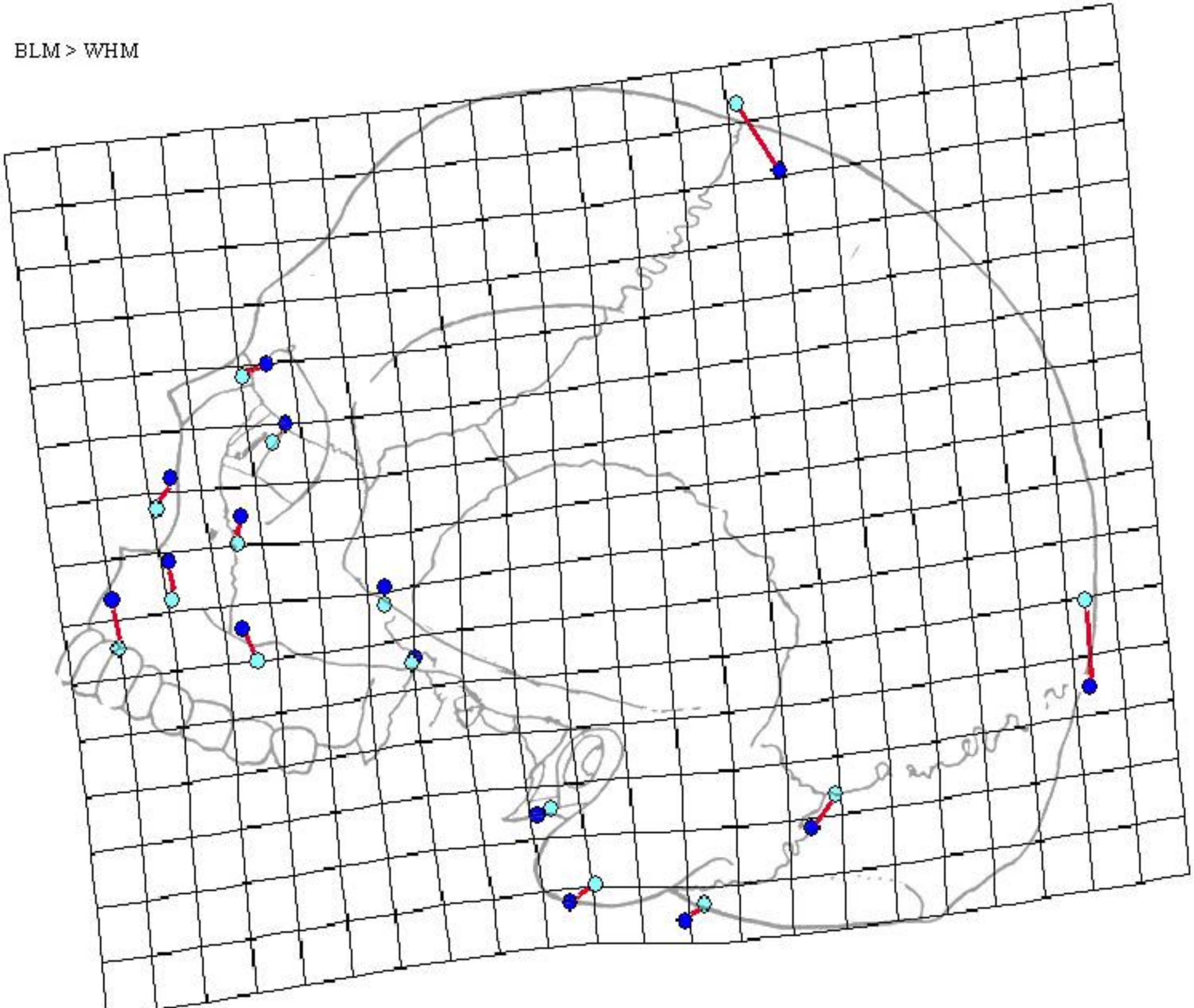
Fred Bookstein: a Biomathematician

Father of Modern Geometric Morphometrics
Morphometric Tools for Landmark Data (1991)

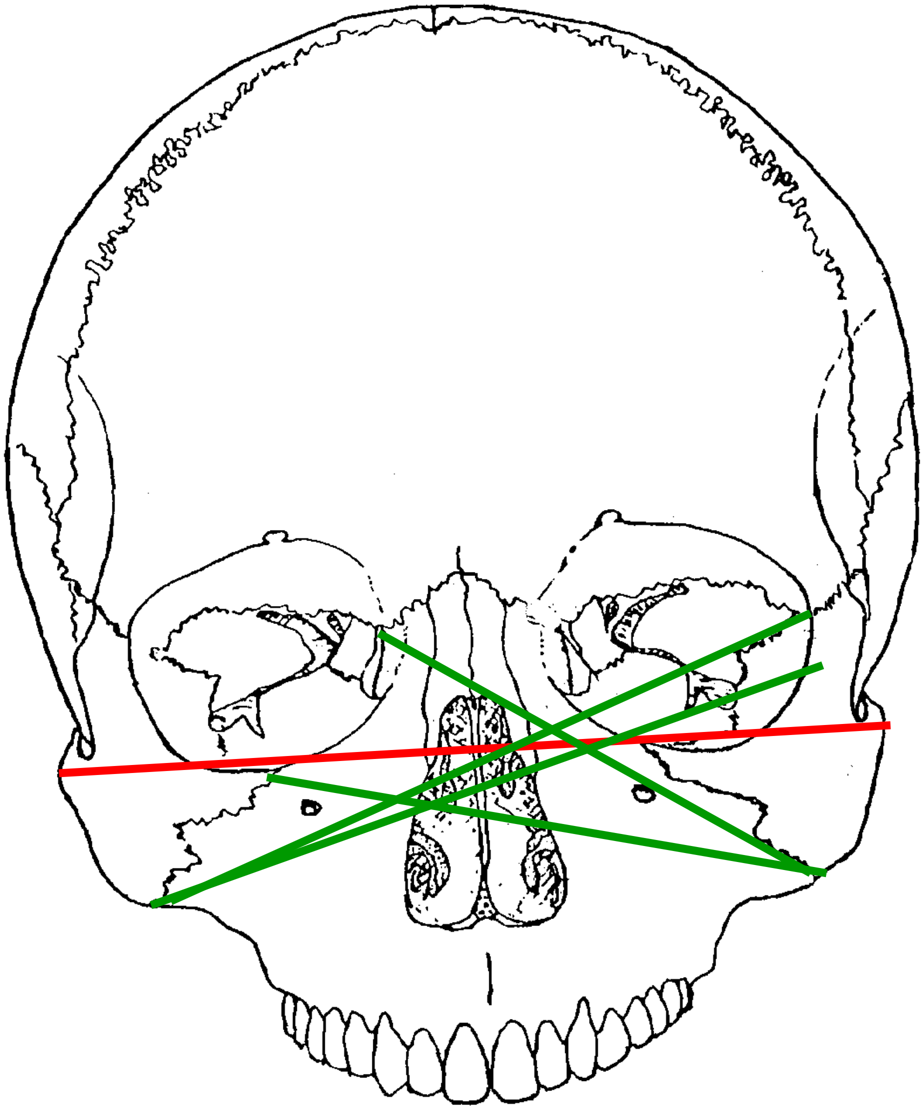
Geometric Morphometrics:
statistical analysis of shape



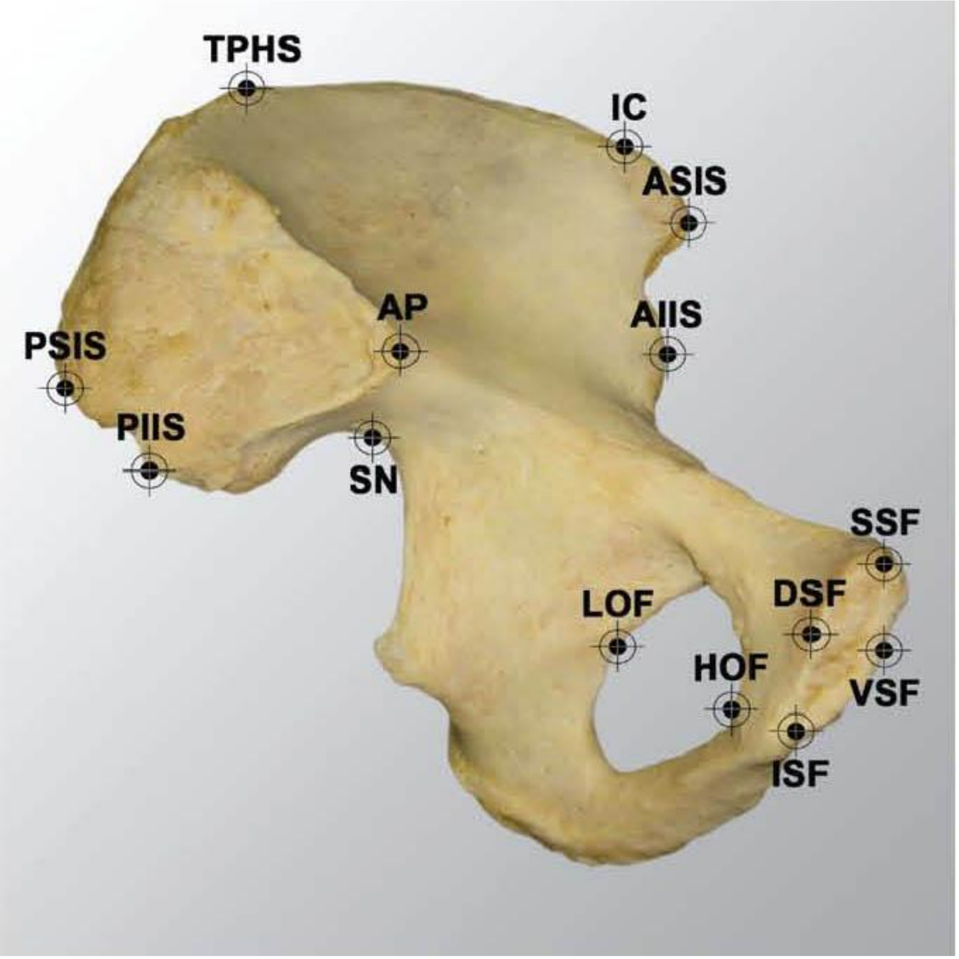
Three-Dimensional Landmark Comparisons (Ousley and McKeown 2003)

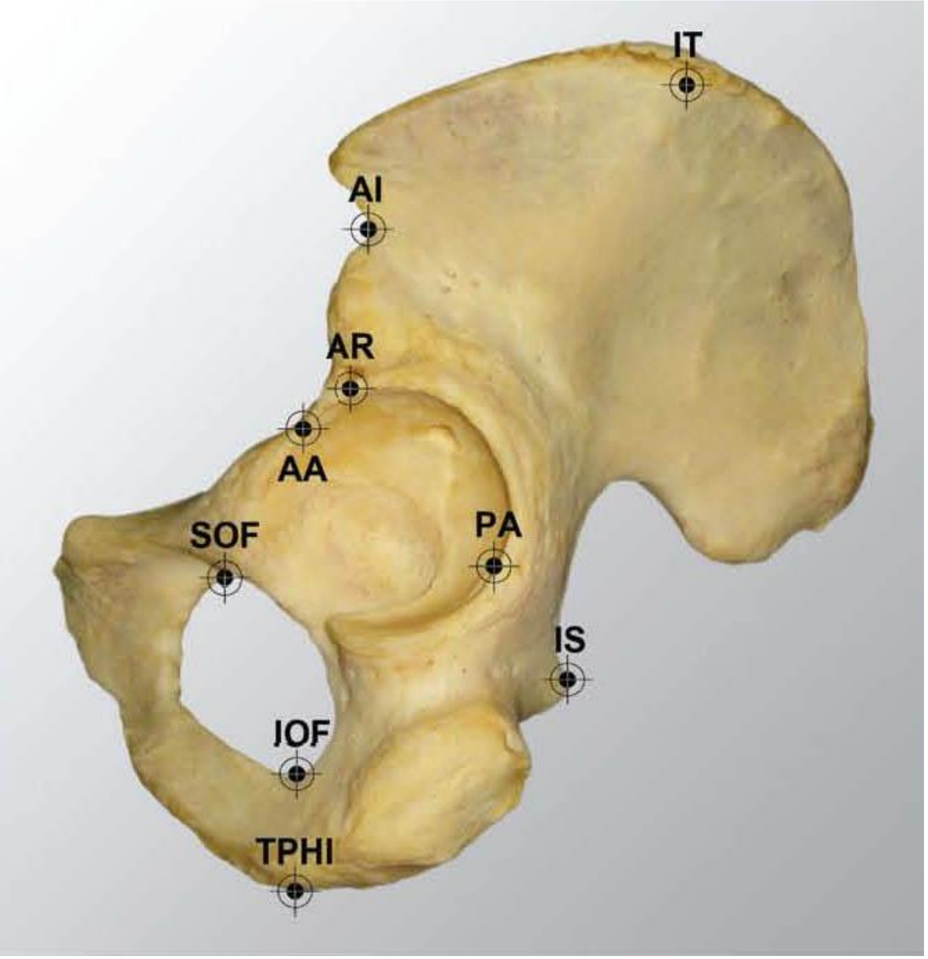


5,000 Interlandmark Distances (ILDs) vs. Measurements

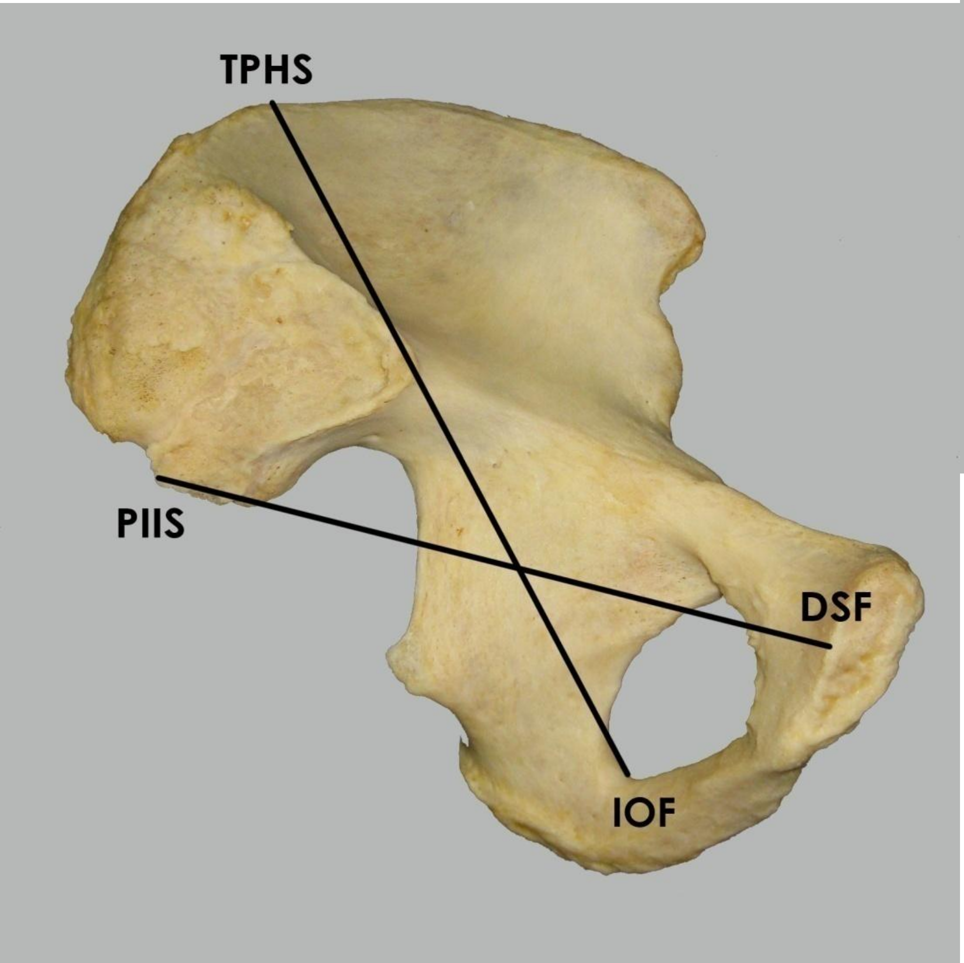


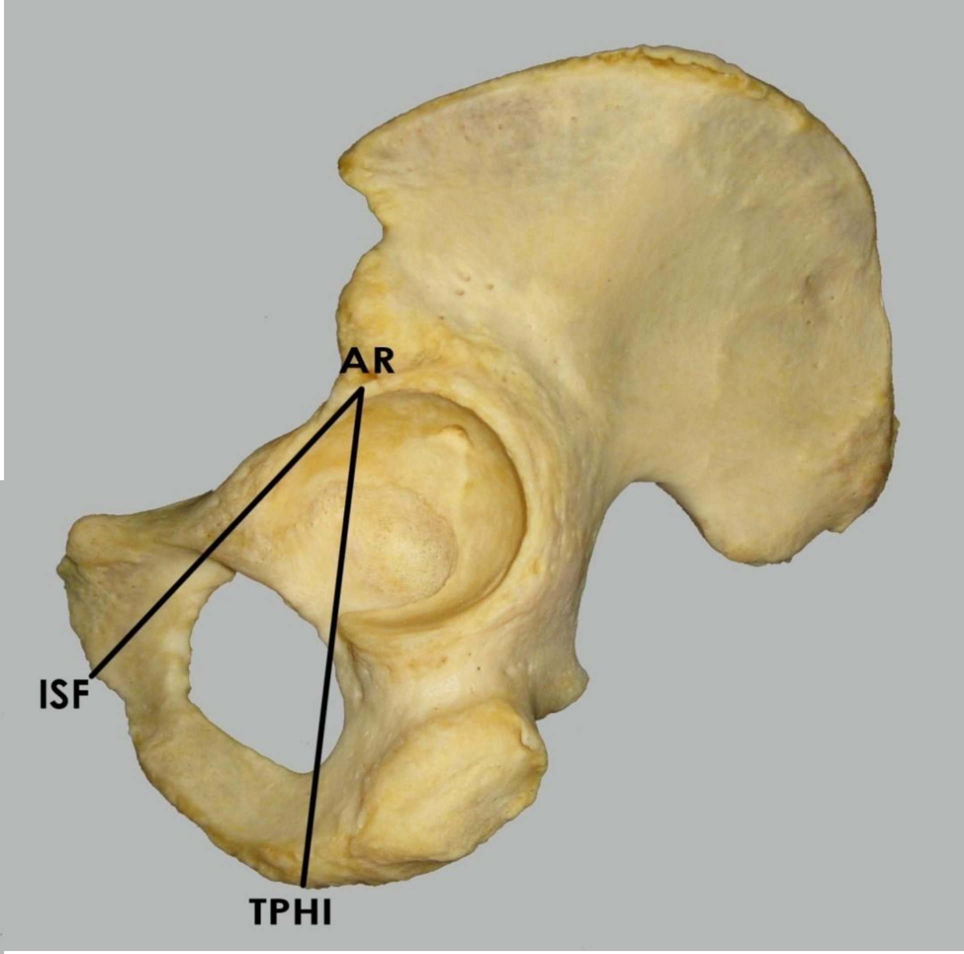
ILDs on the Innominate





Sex Classification Using Simple Measurements





Osteometrics in the *Standards*

Definitions of landmarks, measurement techniques, and instrumentation are based on Martin (1957, as translated in Moore-Jansen et al. 1994), with additional notes on landmarks from Bass (1987) and White and Folkens (1991). Further details concerning osteometry can be found in Bass (1987); Hrdlicka (1952); Montagu (1960); Olivier (1969), and White and Folkens (1991).

(Buikstra and Ubelaker 1994:69-70)

Maximum Femur Length (Martin)

Größte Länge des Femur (Caput-Condylenlänge, maximum length, longueur maximale): Abstand des höchsten Punktes des Caput vom tiefsten Punkt des Condylus medialis bzw. lateralis. Meßbrett. Das Femur wird mit seiner dorsalen Seite so auf das Meßbrett in dessen Längsrichtung gelegt, daß der Condylus medialis (bzw. lateralis) die kurze Querwand berührt. Indem man den Winkel an die höchste Erhebung des Femurkopfes anlegt, sucht man, unter seitlicher Verschiebung des Knochens, die größte Länge. (Martin and Knussmann 1988:216)

Maximum Femur Length (Martin)

[**Maximum Length of the femur**: Distance from the most superior point of the femur head to the most inferior point of the medial or lateral condyle. Osteometric board. **The femur is laid on the osteometric board with the dorsal side down, oriented so the medial (or lateral) condyle is in contact with the moveable upright. In order to orient the femur so the most superior point on the femur head is in contact with the endboard, the bone should be moved side to side to get the maximum length.]**

(Martin and Knussmann 1988:216)

Maximum Femur Length (DCP 1.0)

Maximum Length of the Femur: **The distance from the most superior point on the head of the femur to the most inferior point on the distal condyles** (Figure 36).

Instrument: osteometric board

Comment: Place the femur parallel to the long axis of the osteometric board and **resting on its posterior surface**. Press the medial condyle against the vertical endboard while applying the movable upright to the femoral head. **Raise the bone up and down and shift sideways until the maximum length is obtained** (Bass 1971:168; Martin and Saller 1957:561 #1; Olivier 1969:260; Trotter and Gleser 1952:473).

Maximum Femur Length (*Standards*)

Femur: Maximum Length: **distance from the most superior point on the head of the femur to the most inferior point on the distal condyles**. *Instrument*: **osteometric board**. *Comment*: **Place the medial condyle against the vertical endboard while applying the movable upright to the femoral head** (Figure 54).

Hrdlicka's (1920) Osteometric Standards

The *length maximum* of the femur is measured in the same way as the maximum length of other bones (see under Humerus). *Anthropometry* (Hrdlicka 1920:128)

[Humerus]

Notes.-The *length* is taken on the osteometric board. Apply head to the vertical, take hold of bone by left hand, apply block to distal extremity, and raising bone slightly, **move up and down as well as from side to side until maximum length is determined**.

Anthropometry (Hrdlicka 1920:126)

DCP 2.0

NIJ grant number 2013-DN-BX-K038 to Natalie Langley for new osteological observational standards

Measurement guide
https://fac.utk.edu/wp-content/uploads/2016/03/DCP20_webversion.pdf

Instructional video:
https://www.youtube.com/watch?v=BtkLFl3vim4

Example: Clavicle midshaft diameters
DCP 1.0: Anterior-Posterior and Superior-Inferior
DCP 2.0: Minimum and Maximum
- can be measured from CT scans

Lessons Learned from New Methods

2. New methods need verified data and software to perform them.

Example: 3D-ID (Ross and Slice 2009) 1.0

- classified using cranial landmarks and GM
- very cumbersome to input case information
- outliers in data affected analyses

Problem with software funded by grants?

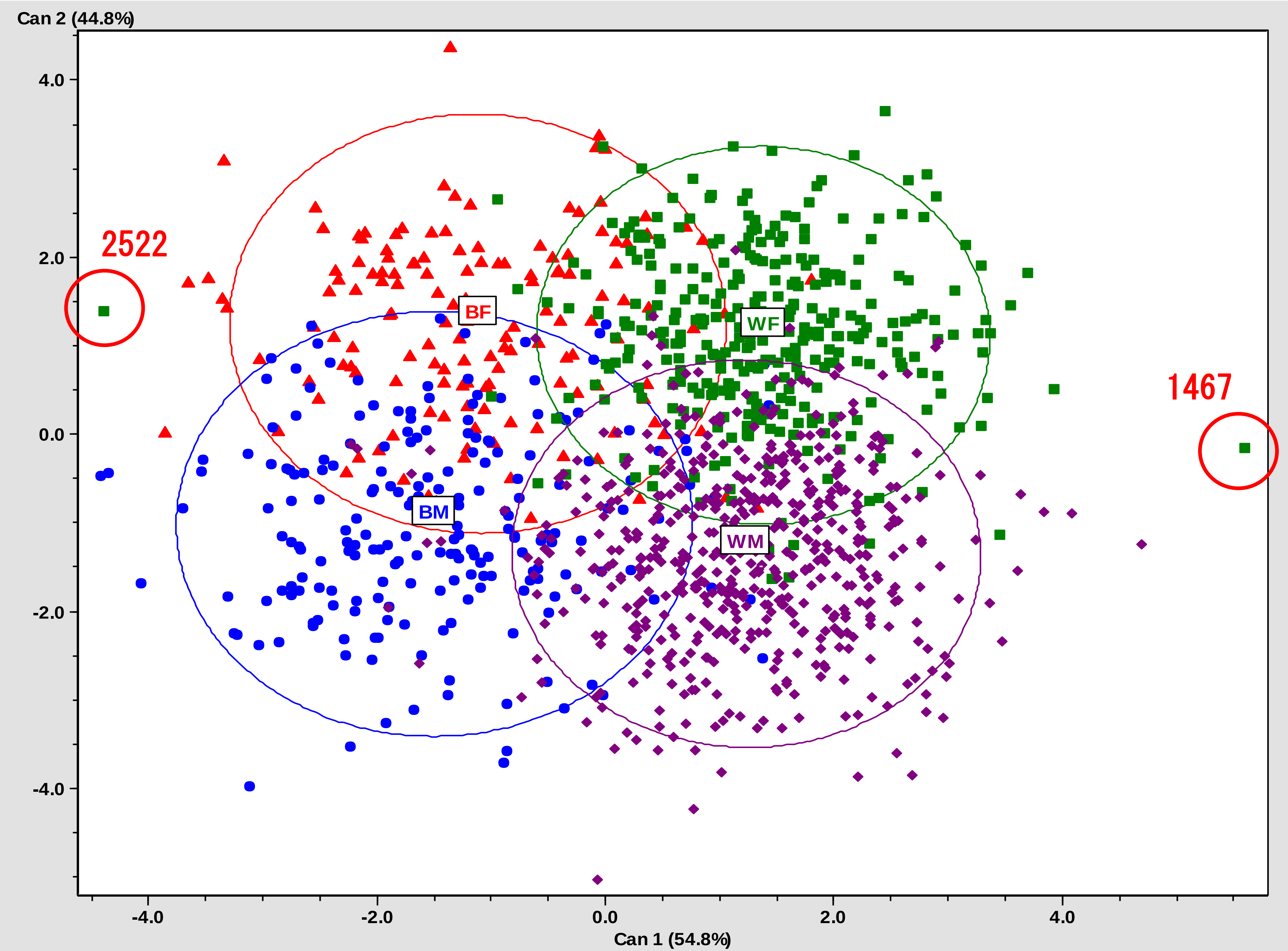
- software development continues until grant is over

Lessons Learned from New Methods

Fordisc uses cranial and postcranial measurements and linear discriminant function analysis (LDFA) to classify skeletal remains into sex and ancestry

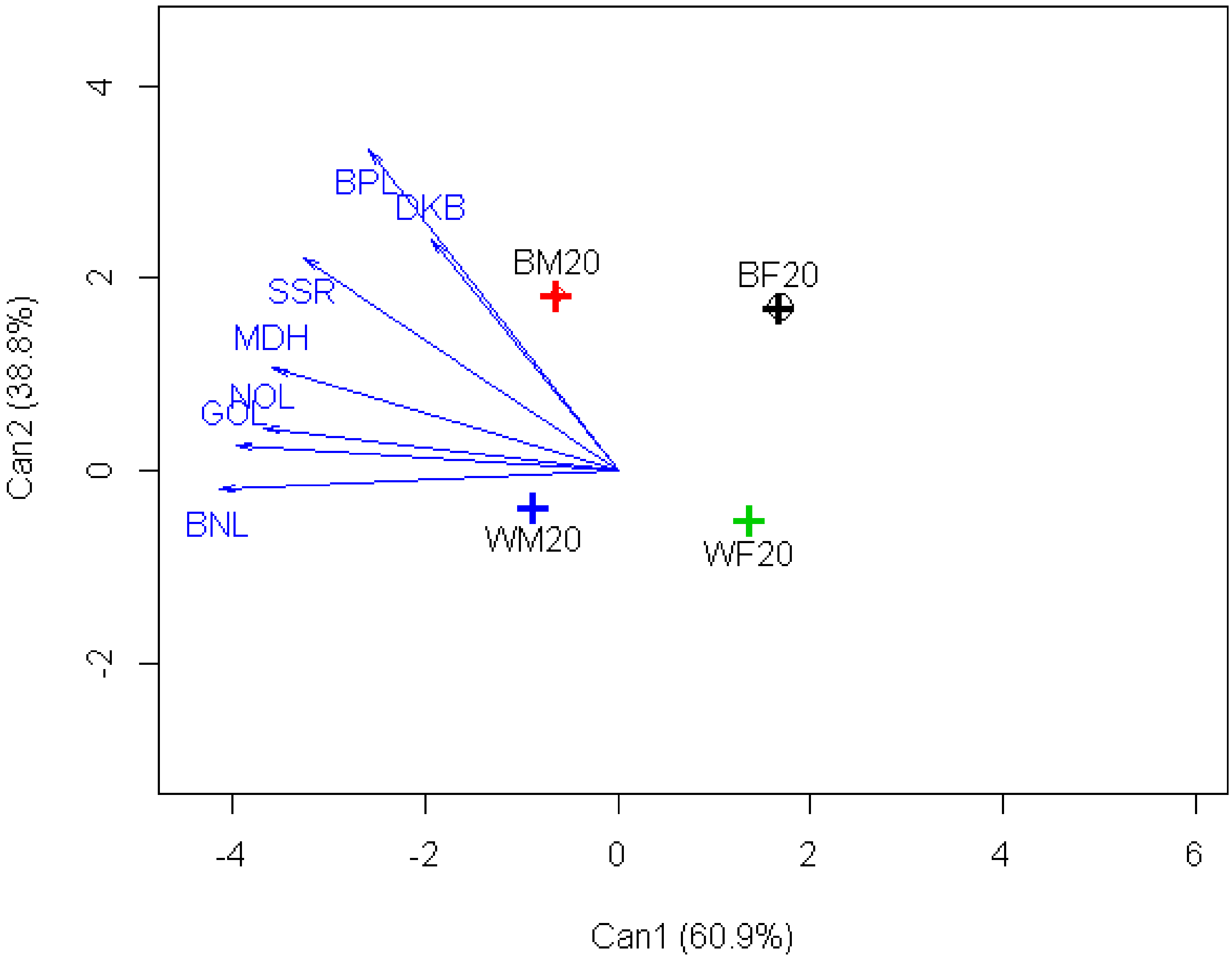
LDFA generally performs very well

-but there are many more classification methods, including Machine Learning

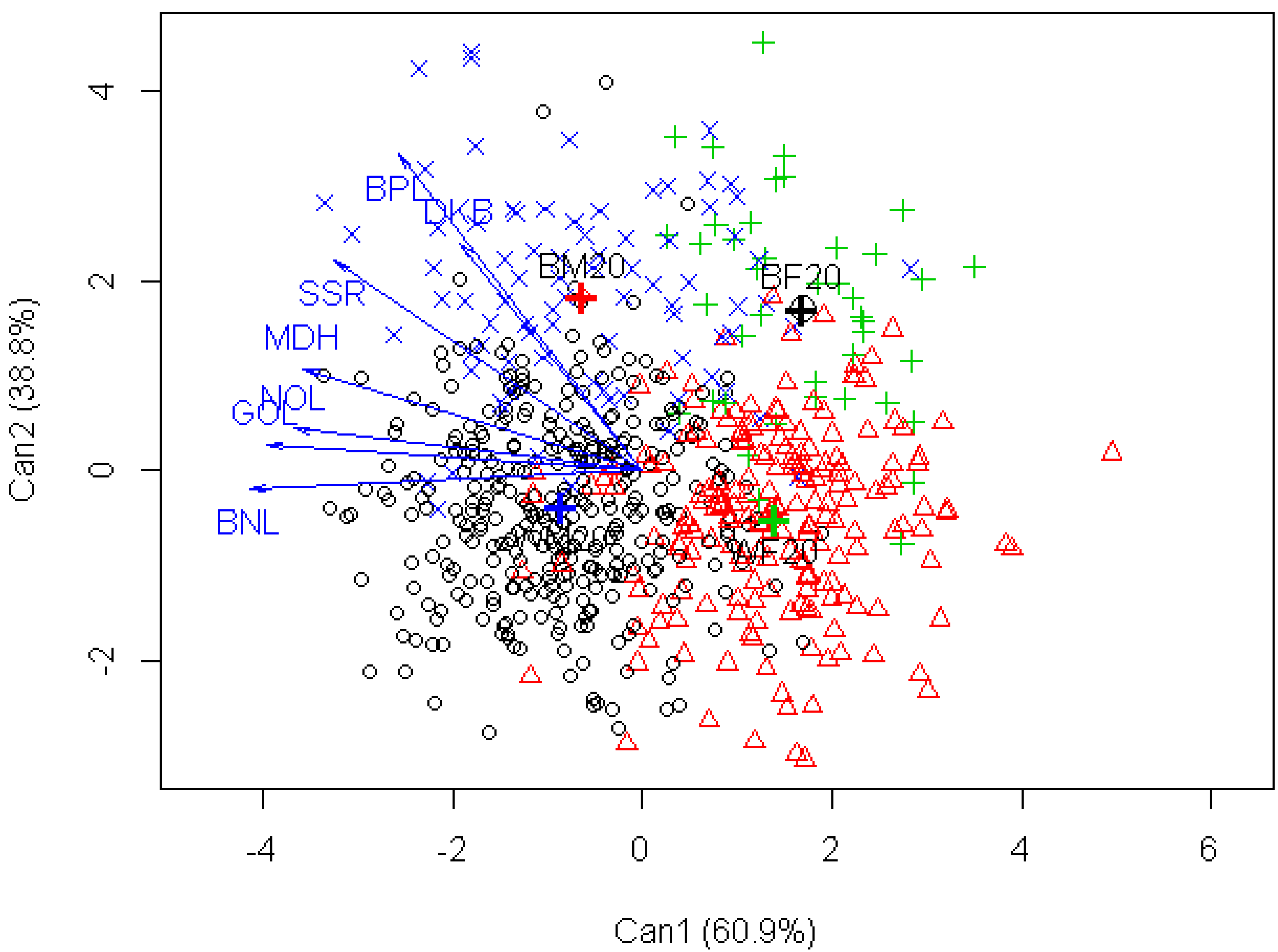


Can 2 (44.8%)
4.0
2.0
0.0
-2.0
-4.0
-4.0
-2.0
0.0
2.0
4.0
Can 1 (54.8%)
BF
WF
BM
WM
2522
1467

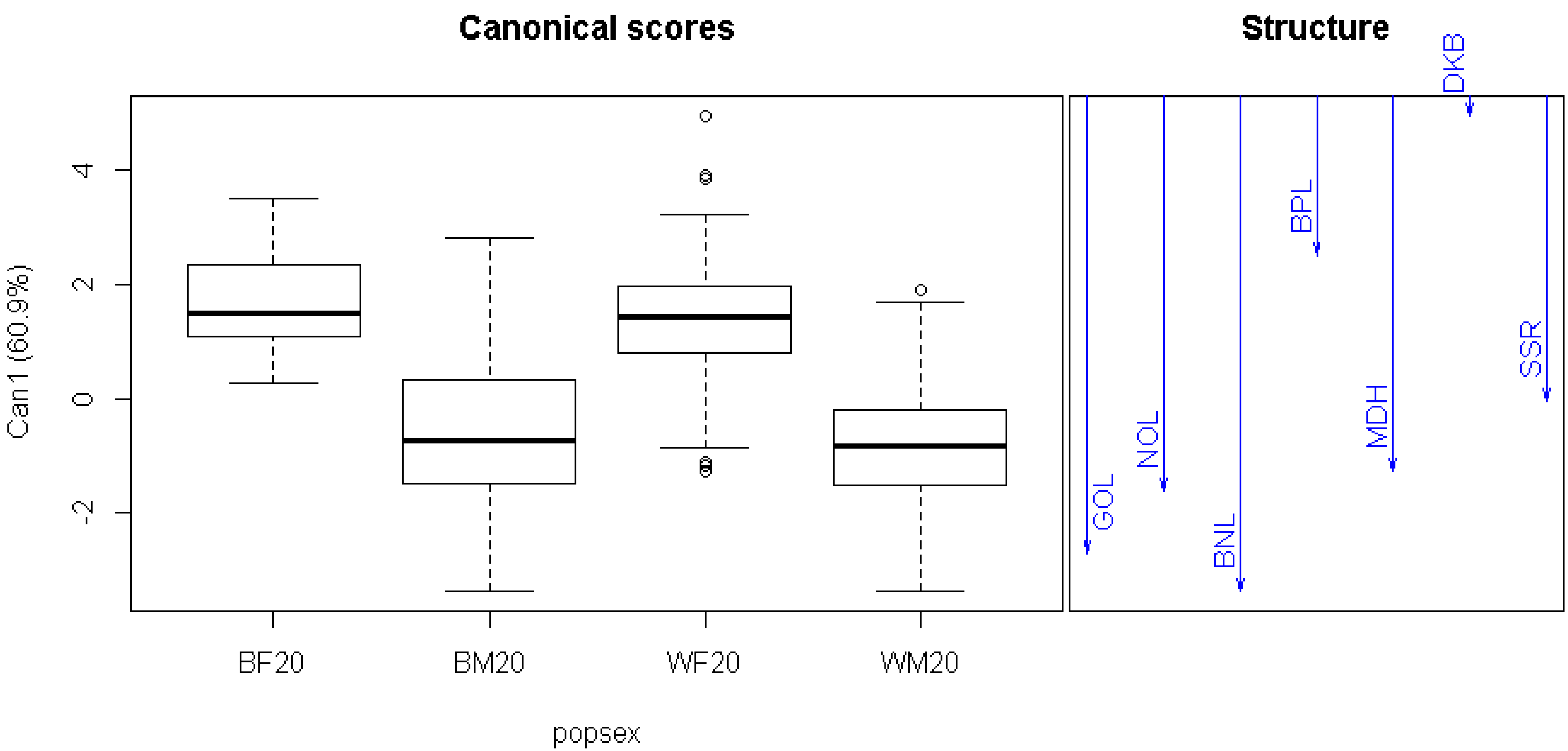
Biplots of Canonical Vectors



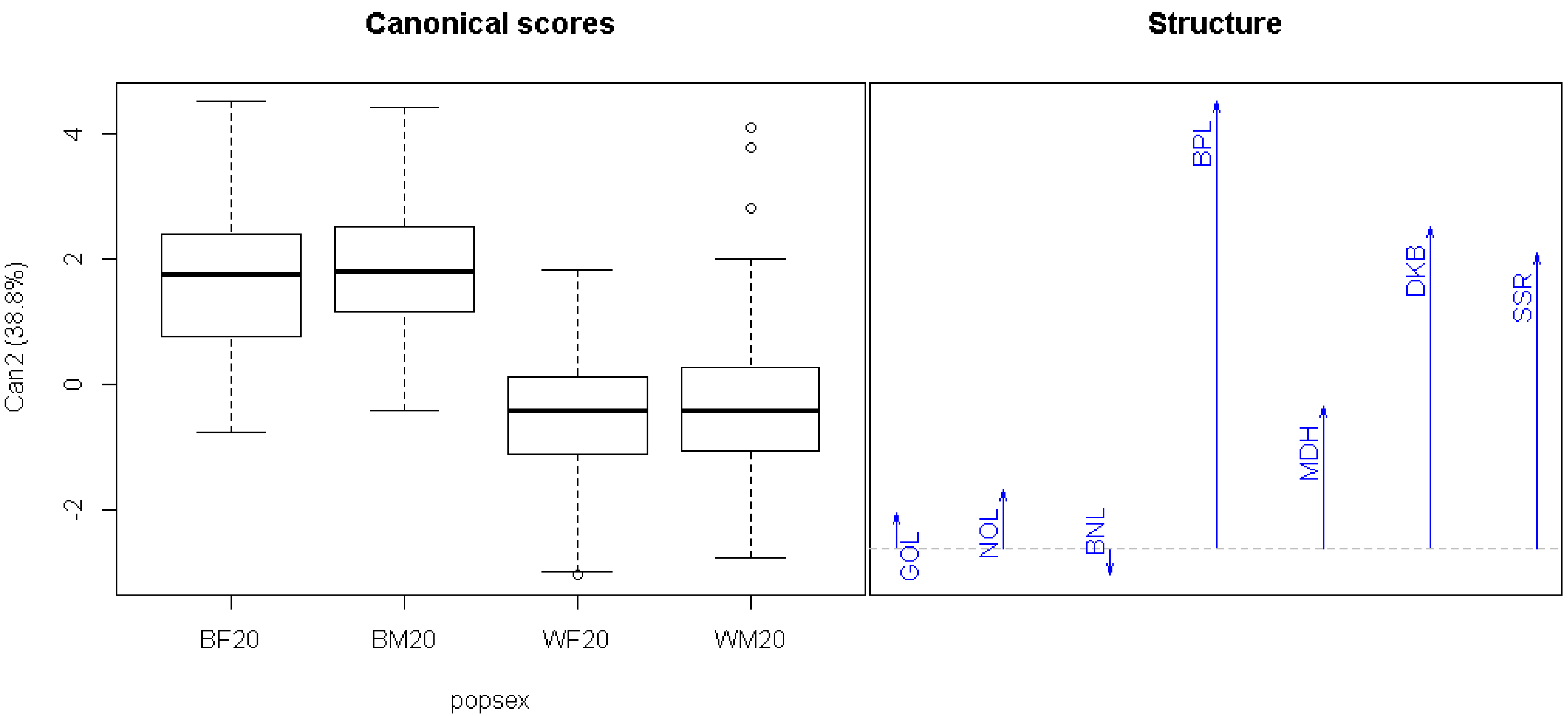
Biplots of Canonical Vectors



Structural Coefficients for CV 1



Structural Coefficients for CV 2



Individual *F*-typicality probability from Morrison (2005) and Konigsberg et al (2009) (PROBLEM)

$$F_{t,N-1-t} = \frac{(N-1-t) \times N \times d^2}{t\left((N-1)^2 - N \times d^2\right)}$$

where *N* is the total number in the group,
t is the number of measurements (traits)
d^2 is the Mahalanobis distance

Individual *F*-typicality probability in FD3 build 304+: FINALLY correct using Hawkins (1981): Jackknifed D^2

$$T^2 = (n_i - 1)(\mathbf{X}_{ij} - \mathbf{X}^*_{i\cdot})'\mathbf{S}^{*-1}(\mathbf{X}_{ij} - \mathbf{X}^*_{i\cdot})/n_i$$

follows a Hotellings T^2 distribution; that is, $F_{ij} = (v - p + 1)T^2/(vp)$ follows an F distribution with p and $v - p + 1$ degrees of freedom.

where n_i is the total number in the group,
X_{ij} is the individual removed,
S^{*-1} is the pooled VCVM without the individual
X_i^*. is the group mean without the individual
$v = N - g - 1$; p = number of measurements

F-Typicality Probability for each member of a reference group is:

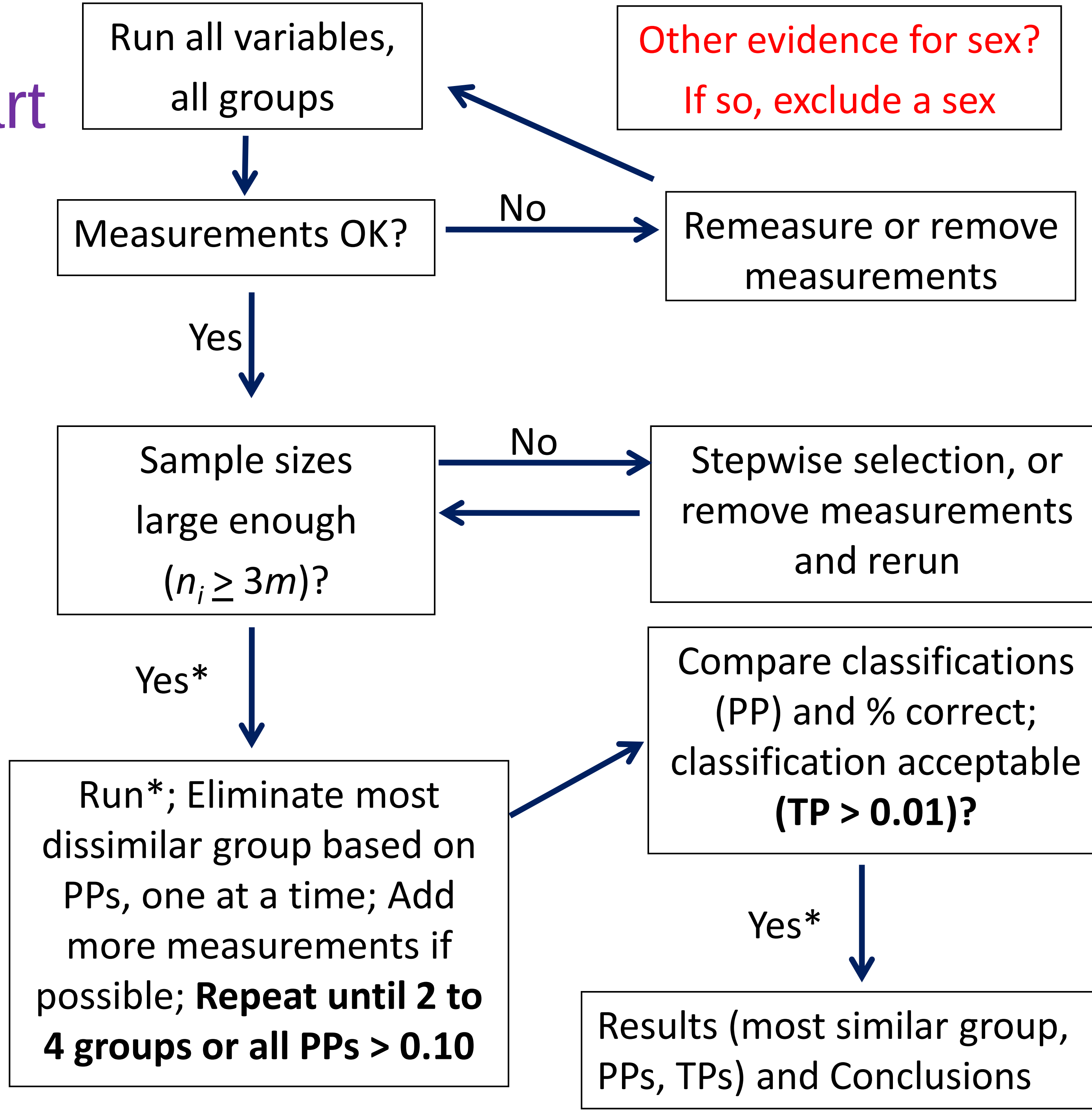
$$T^2 = \frac{(n_i - 1)D_{-j}^{\,2}}{n_i}$$

follows a Hotellings T^2 distribution; that is, $F_{ij} = (v - p + 1)T^2/(vp)$ follows an F distribution with p and $v - p + 1$ degrees of freedom.

where n_i is the total number in the group,
$D_{-j}^{\,2}$ is the jackknifed distance
$v = N - g - 1$; p = number of measurements

DFA Flowchart 1.07

(April 21, 2015)



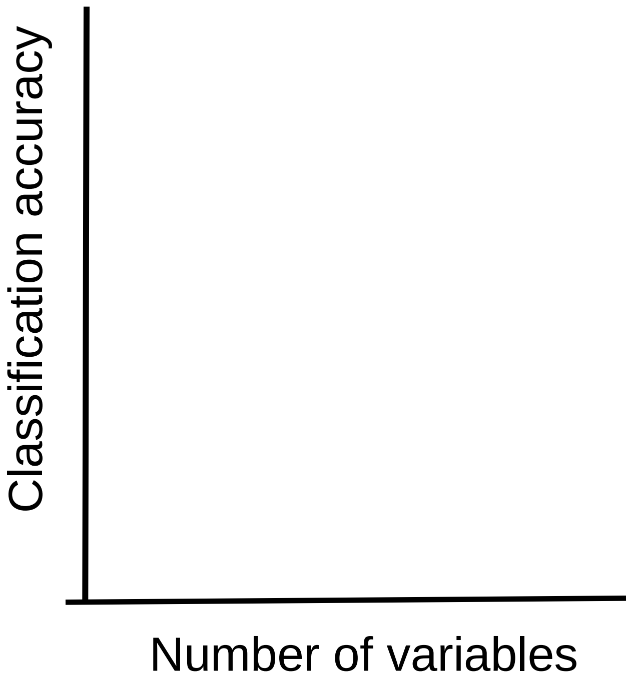
* Remove outliers from reference groups and check for unequal VCVMs

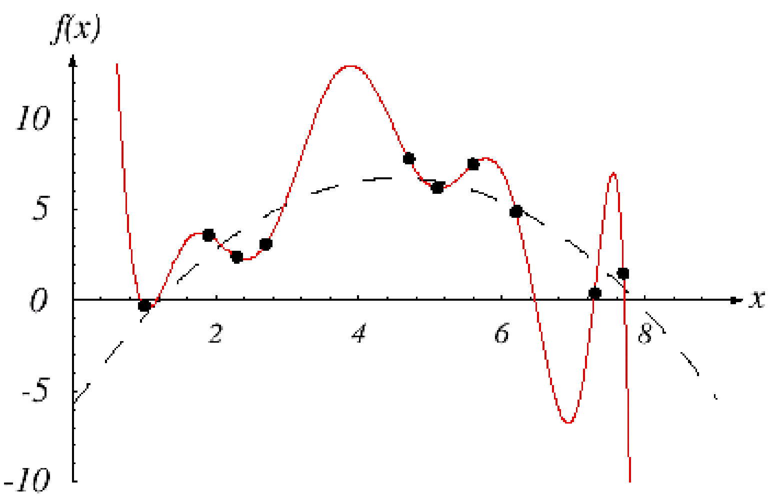
Applying DFA:
The Curse of Dimensionality

While classification accuracy usually increases with more measurements,

- **beyond a certain point, the inclusion of more measurements leads to worse performance**

- computational complexity grows
- the problem of overfitting arises





Machine Learning (ML)

"It's tough to make predictions, especially about the future!"

(Berra)

ML stresses empirical distributions rather than theoretical

- it is data-centric, with extensive resampling (bootstrap)

- its goal is prediction accuracy rather than understanding

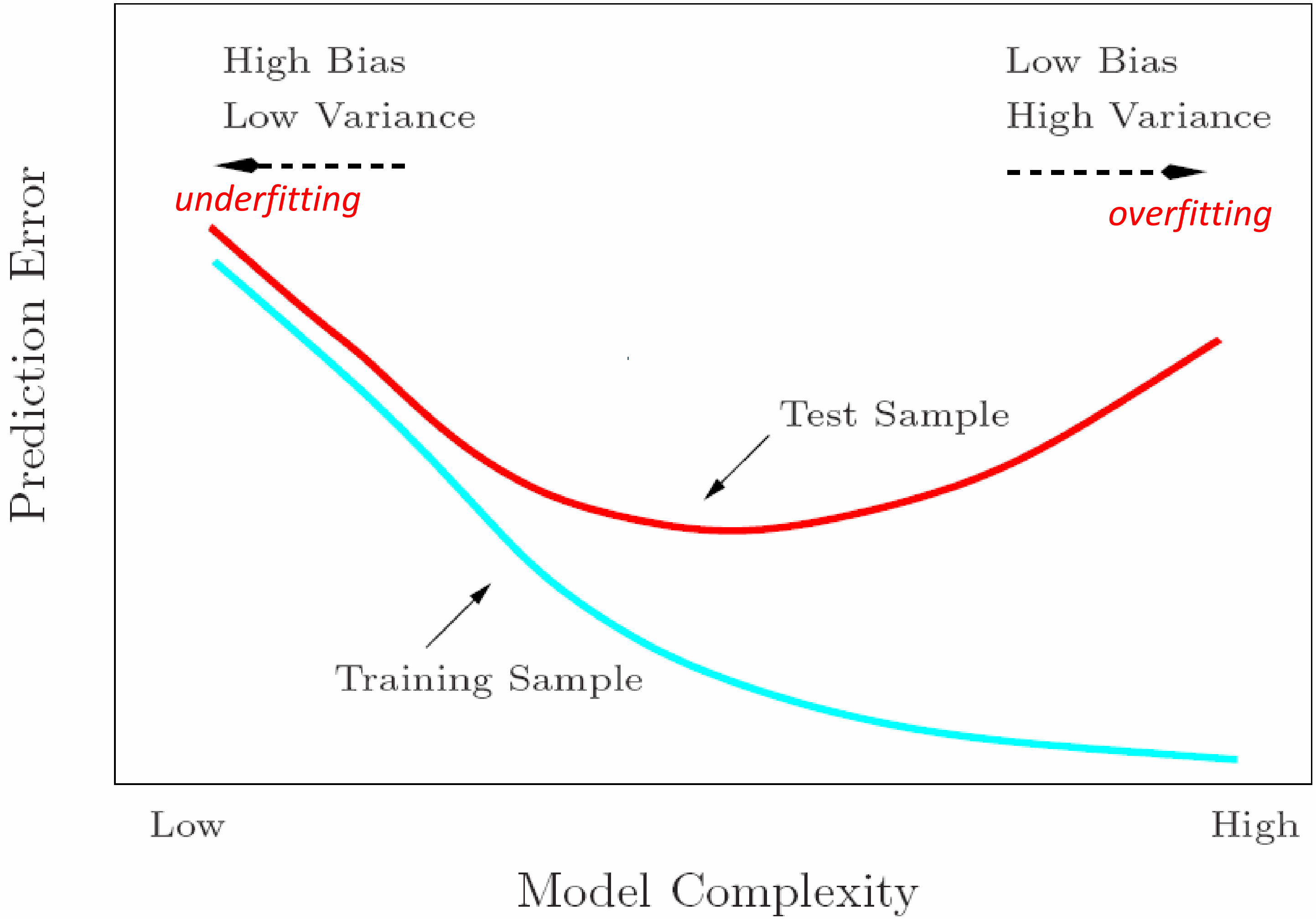
ML stresses independent training and testing samples

- training: for model estimation and fitting

- testing: to estimate prediction accuracy (in the future)

Averaged models make better predictions

The Bias-Variance Tradeoff



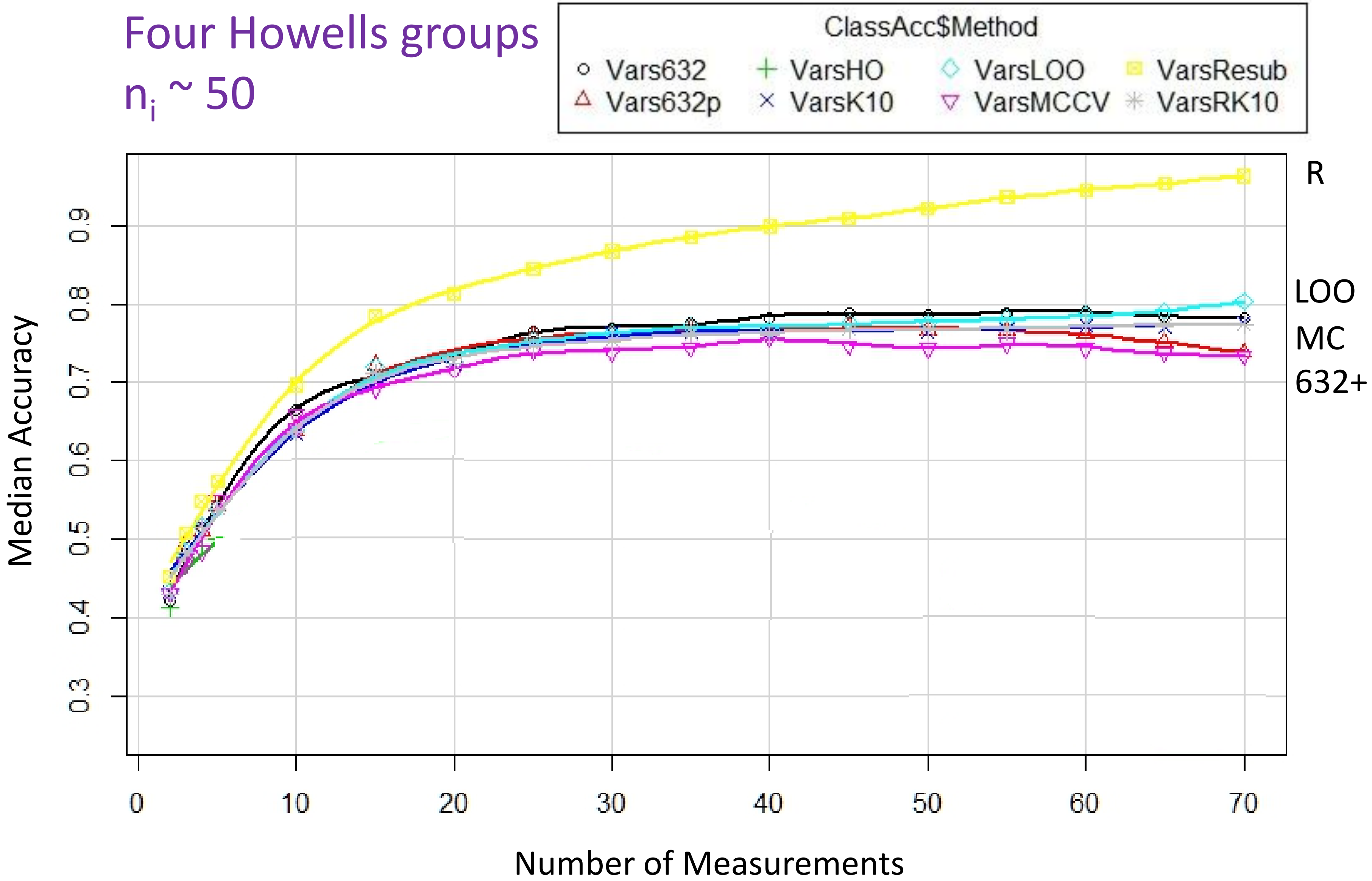
Plus:
unequal prior probabilities
unequal misclassification costs

From: *The Elements of Statistical Learning*, 2nd ed. (2009), by Hastie, Tibshirani, and Friedman.

Estimating Accuracy: Methods

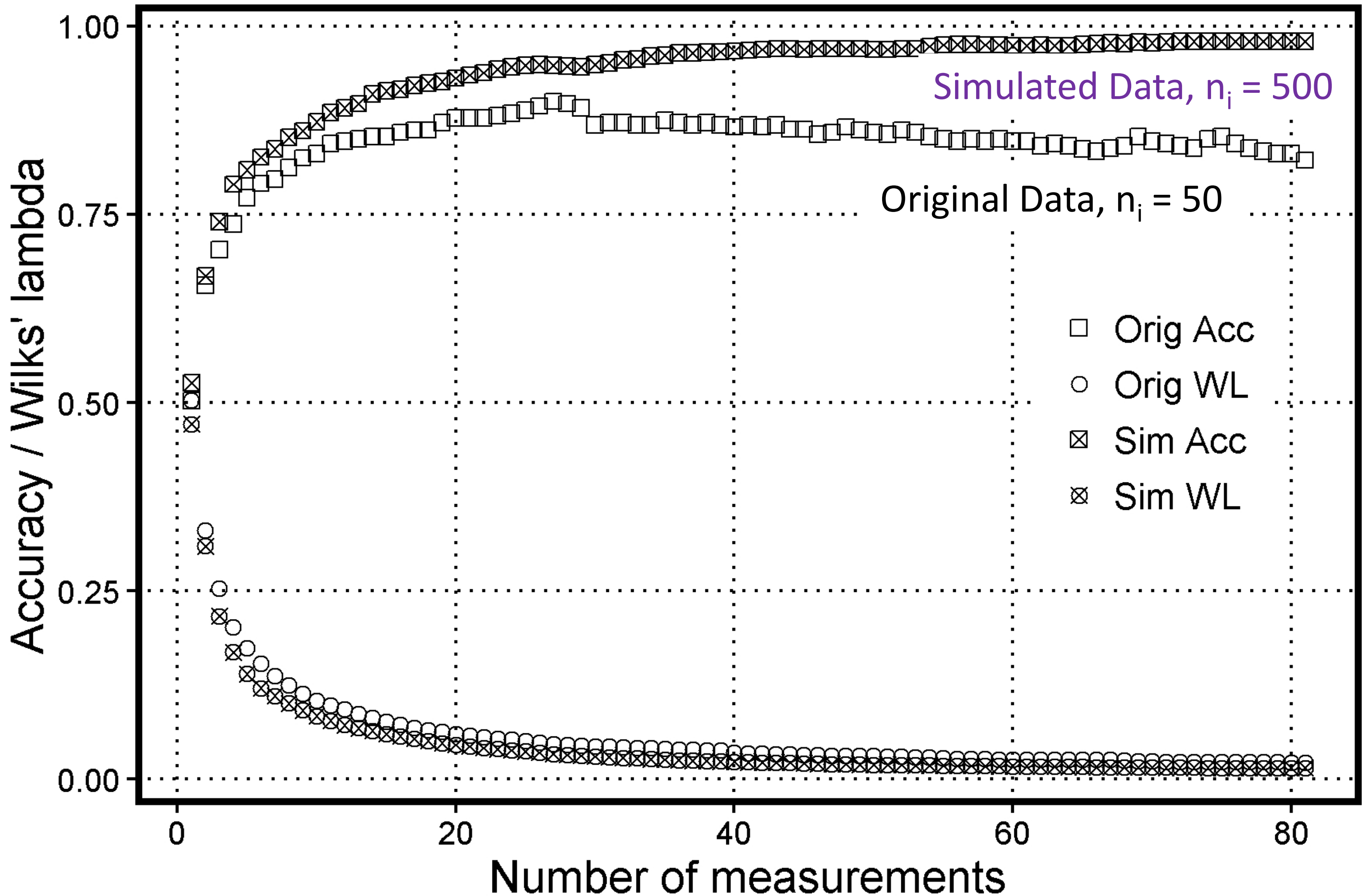
Four Howells groups

n_i ~ 50

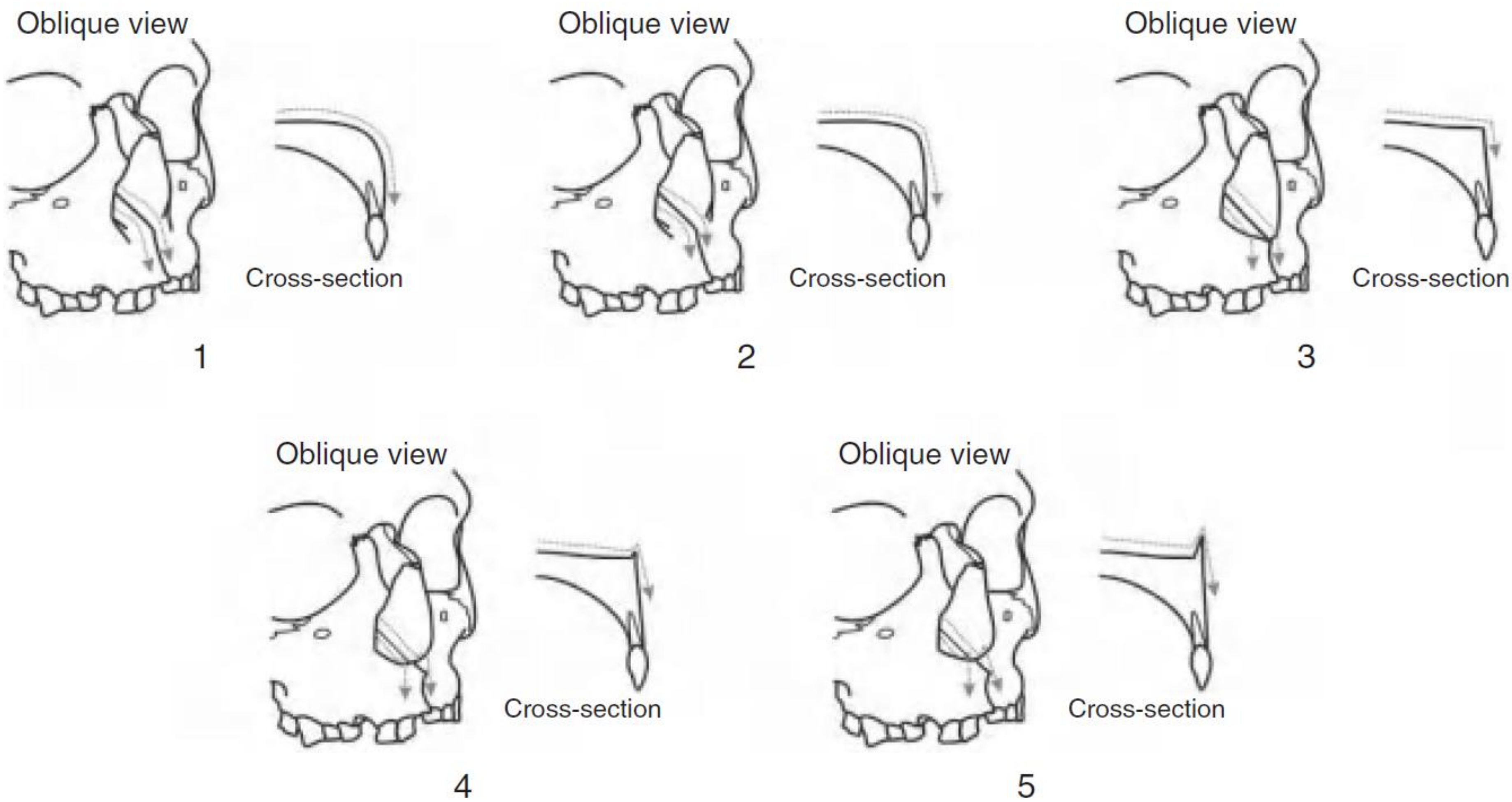


Accuracy Using Simulated Data

(Ousley 2016)



Macromorphoscopic Traits



Definition. The most inferior portion of the nasal aperture, just lateral to the anterior nasal spine, which, when combined with the lateral alae, constitutes the transition from the nasal floor to the vertical portion of the maxillae, superior to the anterior dentition.

Decision Trees: A Simple Visual Method for Classification

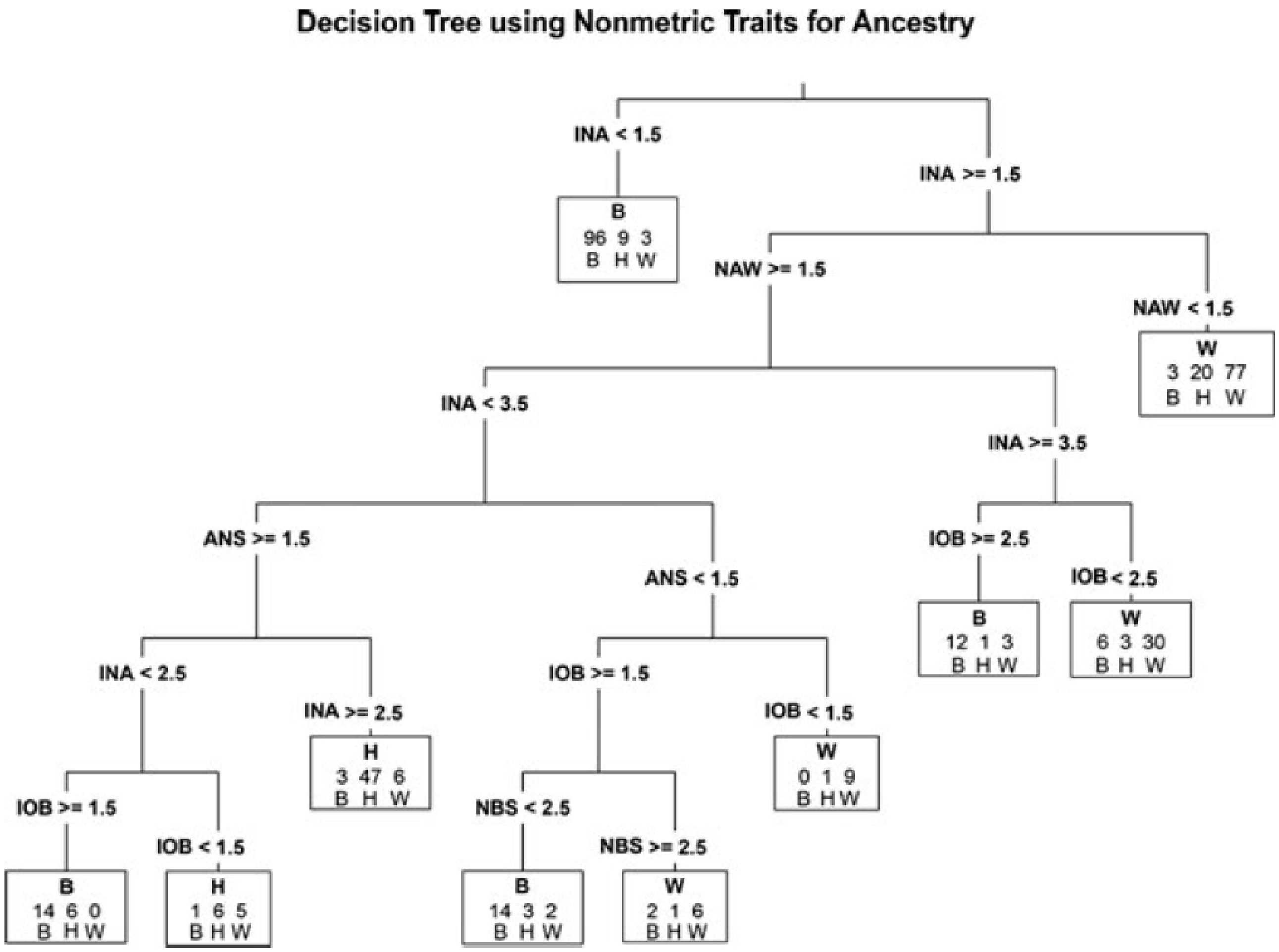


FIG. 3—*Decision tree using all five nonmetric traits to classify into ancestry. The total sample is divided into bifurcating branches according to the listed criteria. The ends of the branches are known as leaves and are shown as boxes with the number of individuals classified in the training sample into that leaf from each group, (B) Black, (H) Hispanic, and (W) White. Bold text above numbers represents the group into which an unknown is classified.*

Estimating Ancestry using Dental Morphology

Table 20.2. *Dental traits observed with associated breakpoints and scoring type*

Trait	Code	Absent	Present	Scoring type
Winging	WING	3	1–2	Categorical
UI1 Labial curvature	UI1LC	0–1	2–4	Ordinal
UI1 Shoveling	UI1SS	0–1	2–7	Ordinal
UI2 Shoveling	UI2SS	0	1–6	Ordinal
UC Shovel	UCSS	0–1	2–7	Ordinal
UI1 Double shovel	UI1DS	0	1–6	Ordinal
UI2 Double shovel	UI2DS	0–1	2–7	Ordinal
UP1 Accessory cusps	UP1AC	0	1–3	Categorical
UM1 Metacone	UM1MC	0–4	5–6	Ordinal
UM1 Hypocone	UM1HC	0–4	5–6	Ordinal
UM1 Cusp 5	UM1C5	0	1–5	Ordinal
LI1 Shoveling	LI1SS	0	1–7	Ordinal
LI2 Shoveling	LI2SS	0	1–7	Ordinal
LP1 Lingual complexity	LP1LC	0–1	2–9	Mixed ordinal/categorical
LP2 Lingual complexity	LP2LC	0–1	2–9	Mixed ordinal/categorical
LM1 Groove pattern	LM1GP	0–1	2–6	Ordinal
LM2 Groove pattern	LM2GP	0	1	Categorical

Lessons Learned from ML Methods

ML can produce very high classification accuracies

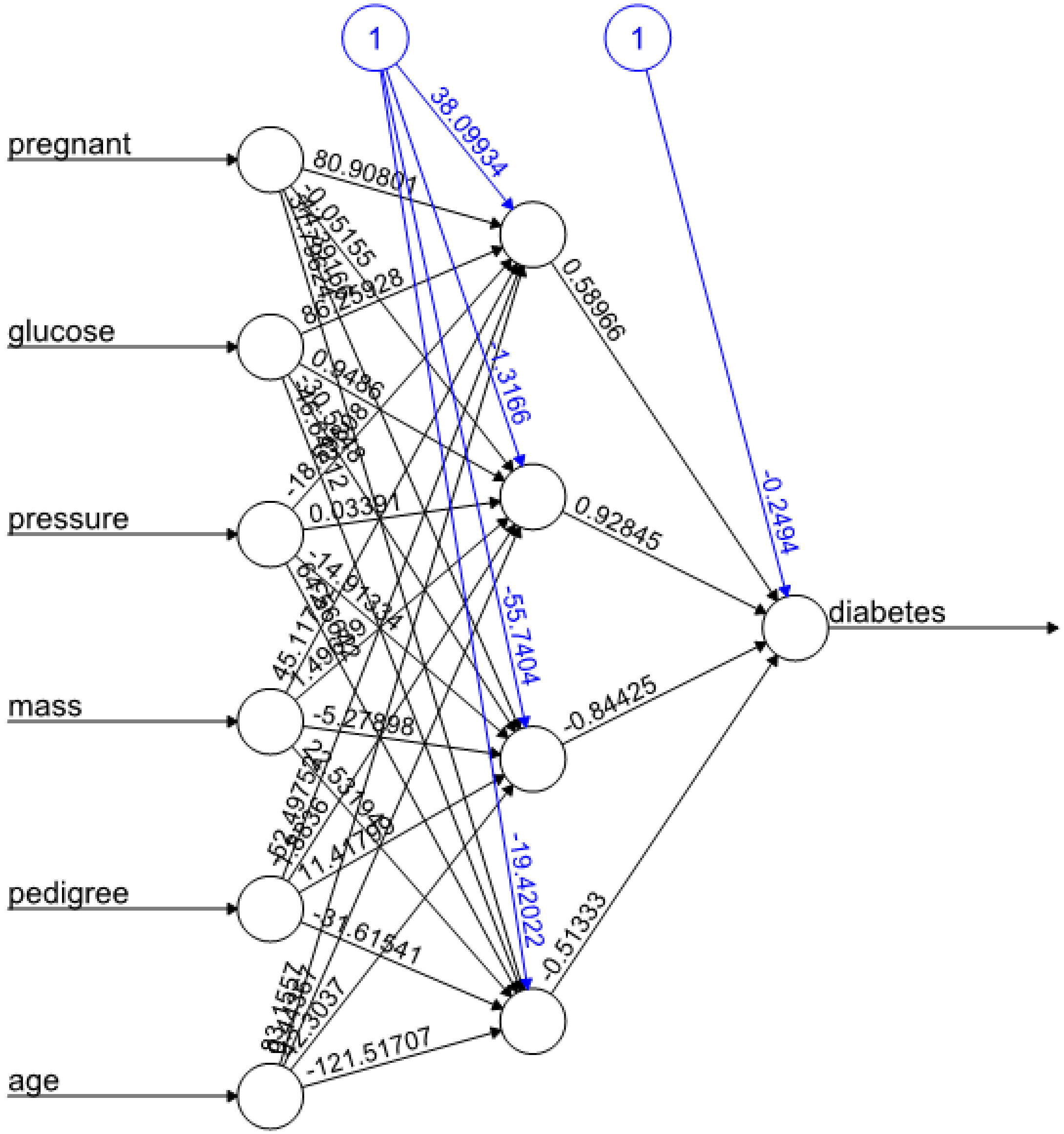
- but some methods only classify
- actual process can be opaque: Neural Networks

Some ML methods improved using simulated data

- larger samples help them perform better
- much more research needed

ML methods do not help us understand the data

ML: Neural Network Results



Error: 181.242328 Steps: 8448

Lessons Learned from New Methods

3. Practitioners are not the greatest scientists

- they look at one case at a time

- they are usually under time pressure

- faster "blink" methods favored over more accurate ones (*"complicated and expensive instruments"* like calipers are avoided)

They want a simple solution

They are more likely to ask for new useful features

They are less likely to find analytical bugs

Lessons Learned from New Methods

4. Academics are not the greatest practitioners

It is hard to create methods for people without experience

Teachers are more comfortable with familiar methods

Easier methods are more replicable.

Institutional momentum

Generational lag

“Schools” of analysis

They are LESS likely to ask for new useful features

They are MORE likely to find analytical bugs

Lessons Learned

5. We need new methods for academic evaluation

- forensic work is generally underappreciated
- computer programs are not traditional publications
- excellent help files are buried in programs
- a blog may have a far greater positive effects than a book

Different approaches to funding may be necessary

- making source code available, enabling further development and improvements
- having tiered software, some features free, others $
- probably sustainable with nominal subscription costs

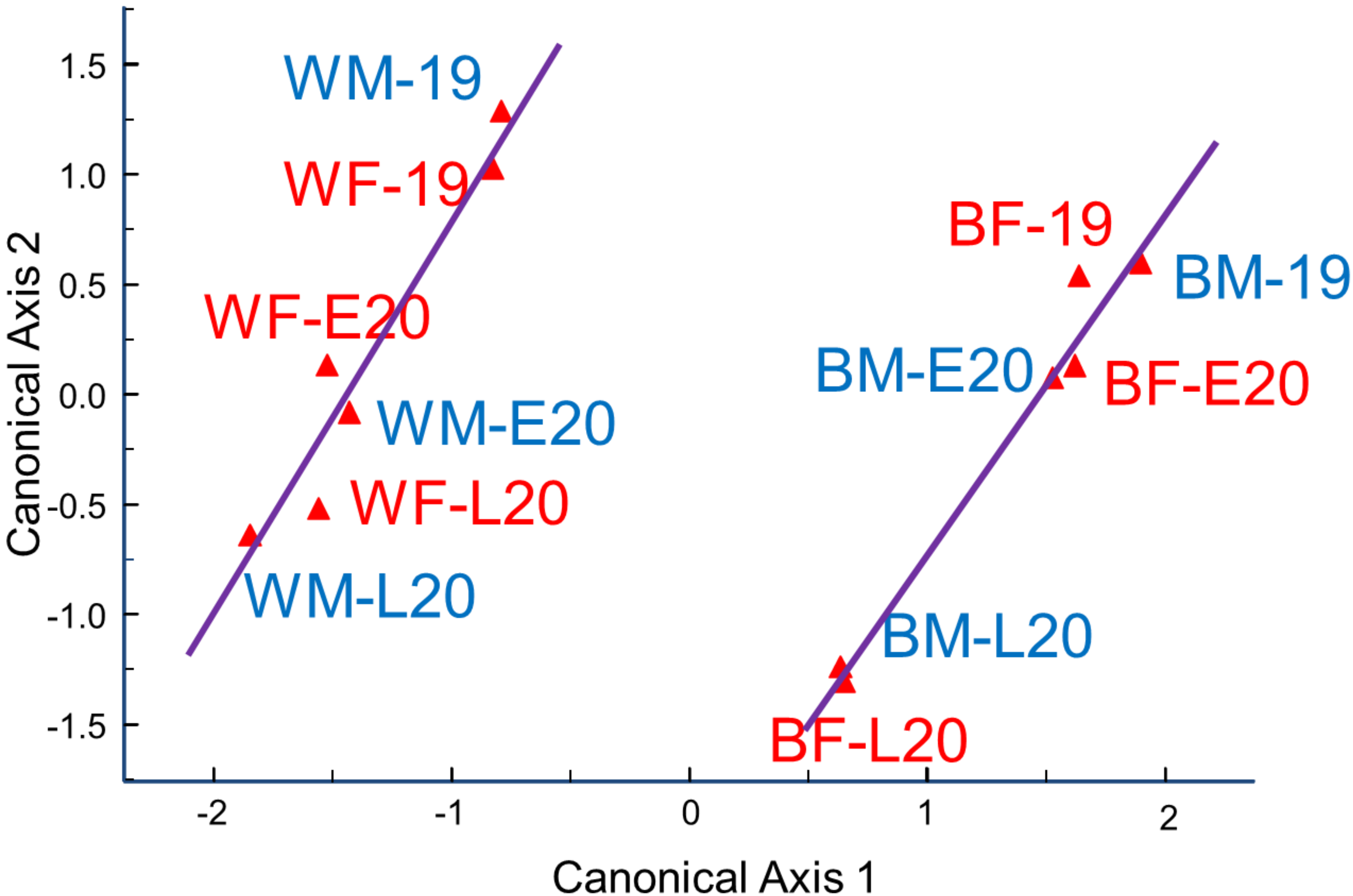
Lessons Learned

6. “It takes a village”: We need practitioners, academics, and independent scientists to work together – AND -
We need to lure more computer scientists, data scientists, and programmers to anthropology

Collaborations
Cross-disciplinary academic positions

We should ideally produce applications that are:
- Microsoft Windows-, Apple-, and Android-compatible
- Cloud enabled (Amazon Web Services/Google Cloud)

Secular Craniometric Changes in Black and White Americans



Thank You!