Guide to using the CRANID programs CR6bIND: for linear and nearest neighbours discriminant analysis

by

Richard Wright
richwrig@tig.com.au
richard@cranid.com

28 Nov 12

This program and manual may be cited as:

"Wright, Richard (2012) *Guide to using the CRANID programs Cr6bInd: for linear and nearest neighbours discriminant analysis.* Retrieved [*insert your date*] from
[here insert the URL from which you downloaded the file]."

CONTENTS

SUMMARY ADVICE

HOW DO I GET A RESULT WITH THE PROGRAMS?

Unzip the file Cr6bInd.EXE into a directory of your own choosing (let us say C:\Cr6b). The directory can have any name of eight characters or less, though these notes assume it is C:\Cr6b. Do not unzip it into any existing CRANID directory, e.g. C:\CR5 or C:\CR6.

All operations will be carried out within this directory, including the entering of data into the form, running of the two programs and output of the results.

No further installation or setup is needed.

To get a result for an individual cranium you will take the following five steps:

1. Enter the 29 measurements into the spreadsheet file CR6Form.CSV.[1]

2. Save the spreadsheet file to the directory C:\Cr6b under its real archival name, eg. Shoalhaven.CSV.

3. Save this spreadsheet file a second time to the directory C:\Cr6b, this time with the name MYFILE.CSV (this is the file that the programs will analyse in step 4 below.

4. Find and run the program RUN!!!!.EXE by double clicking on it.

5. Interpret the results of this analysis in the output file, eg. OUT013.TXT.

In this guide you will find explanations for each of the above steps and definitions of the 29 measurements to be used.

---

[1] If you are using a system that is set to the English language, then you will not need to read the rest of this footnote. If not English, then before taking this step you must ensure your system’s Regional & Language Options customization is set so that the decimal symbol is a full stop, not a comma. Alternatively temporarily set your Regional and Language Options to a version of English. Both steps are taken via Start -> Settings -> Control Panel. I am grateful to Dr Adolfo Scatena of Argentina for drawing my attention to this matter.

INTRODUCTION

This package allows you to do a linear discriminant analysis and a nearest neighbour discriminant analysis with 29 measurements on an individual cranium. It assumes that the individual cranium is within the range of variation of modern *Homo sapiens*. The cranium will be classified after automated comparison with 74 samples that include 3,163 crania from around the world.[2]

Essentially the programs are evaluating morphological similarity and dissimilarity. From these morphological results, and because of the high correlation between cranial morphology and geographical origin, ancestry can be inferred.

These notes concentrate on the nuts and bolts of classifying a cranium by discriminant analysis. The notes assume that the user:

(1) understands something of the assumptions underlying craniometry and ancestry;[3]

(2) understands something about the workings of both linear discriminant analysis and nearest neighbour discriminant analysis.[4]

The programs that go with these notes do linear and nearest neighbour discriminant analyses that are sensitive to both size and shape. A cranium must be complete enough to acquire all the 29 measurements.

Extra analyses (eg. of shape alone, or of incomplete crania) are available by e-mailing me. These extra analyses are described below in the section headed "Notes on the full CRANID analysis."

---

[2] The CR6Ind.EXE predecessor of this package used the same crania, and the same analytical methods. This new version Cr6bInd.EXE is simpler to use and has improved, reformatted output.

[3] For a thorough review see Pietrusewsky, M. (2000) Metric analysis of skeletal remains: Methods and applications. In M.A. Katzenberg and S.R. Saunders (eds.) *Biological Anthropology of the Human Skeleton*, pp. 375-415. New York: Wiley-Liss.

[4] A readable general text is Huberty, C.J. & Olejnik, S. (2006) *Applied MANOVA and Discriminant Analysis*, 2nd ed. John Wiley & Sons. I have also written about these methods in a chapter devoted to CRANID5, but which contains principles relevant to CRANID6: “Detection of Likely Ancestry using CRANID”, in *Forensic Approaches to Death, Disaster and Abuse* edited by M. Oxenham, Australian Academic Press. 2008. pp.111-122.

COPYRIGHT AND CONDITIONS OF USE

I keep copyright on the notes and software. However you may, without breaching my copyright, use, copy and distribute the notes and software on the following conditions:

> in any report or publication you acknowledge the author of the program as Richard Wright with the email address *richwrig@tig.com.au*;
>
> if distributing the package you
>
>> make no charge;
>>
>> include the downloaded and zipped file Cr6bInd.EXE in its unaltered state;
>
> you use this free program for personal interest and research in the ordinary course of your employment.
>
> you do not use this free program to undertake a consultancy that requires the use of this program.[5]

I hope to be able to advise you on the use and results of this free program, but give no undertaking to do so. If you ask for advice please also attach a copy of any CSV data file that you think is problematical.

Users that get the software other than directly from me are advised to send their email addresses to me. I like to keep users up to date with any improvements to the program and revisions of the manual.

I have produced these notes and accompanying software to offer assistance to people wanting to enhance their identification of ancestry from crania. They are offered free and ‘as is’. I have tried to compile the notes and programs by taking every care. However inaccuracies may still occur. These could lead to the user being misinformed or getting wrong results. While regretting the occurrence of any inaccuracy, I cannot accept responsibility for any consequences that may result from that.

---

[5] I freely distribute the package for private use, and am happy to consider questions from the private user. In a commercial context, I shall be happy to negotiate a price for doing the necessary work myself or acting as a contracted consultant. One reason for my not freeing up the use of the software for commercial purposes is that the commercial user gets the money, but I am left with answering questions from both the commercial user and their client.

FILES THAT SHOULD BE IN THE DIRECTORY

Remember that all operations must take place within the directory into which you have unzipped the package Cr6bIND.EXE.

You should keep the original file Cr6bInd.EXE in case any of the supporting files become corrupted or accidentally deleted.

Your directory must always contain the following files:

C.DAT
Checking Table.xls
Cr6bInd.EXE
CR6CVO.DAT
CR6CVO.LBO
CR6CVO.LBV
CR6Form.csv
CR6GPNam.txt
CR6LDA.EXE
Cranid6.GPN
CRANID6.MNS
CRANID6.PWT
CRANID6.SDS
CRANID6b_Manual.pdf
CVACR.WTS
MEAN.DAT
MNNDMDFC.EXE
NNID1.EXE
NNID2.EXE
RUN!!!!.EXE
STDDEV.DAT
SUBTENSE.EXE
TestFile.csv

Never delete these listed files. If at any time a program will not run then try restoring the files by running the self-extracting file Cr6bInd.EXE again.

A tip - after considerable use, the directory gets cluttered with unwanted files. Get Windows to *View* the directory by *Details*. Set *Arrange Icons by Modified*. After saving elsewhere the results that you want to keep, this procedure will allow you safely to delete the detritus of earlier analyses.

FILLING IN AND SAVING THE SPREADSHEET

You will find the read-only file CR6Form.CSV in the directory.

To fill in the form, open the file CR6Form.CSV with a spreadsheet program such as Excel.[6]

You will be substituting your values for the question marks on the screen.

Leave the remainder of the form untouched – just replace the question marks.

(1) Enter name of your specimen in row 4 as replacement for the question mark. Enter values as replacements for the question marks	
(2) Save file (including these first four lines) as a comma separated value (CSV) text file called MYFILE.CSV	
(3) Repeat save as a backup, using name of your own choice, e.g. NeolithicCranium.CSV.	
?	
GOL	?
NOL	?
BNL	?
BBH	?
XCB	?
XFB	?
AUB	?
ASB	?
BPL	?
NPH	?
NLH	?
OBH	?
OBB	?
JUB	?
NLB	?
MAB	?
ZMB	?
SSS	?
FMB	?
NAS	?
EKB	?
DKB	?
WMH	?
FRC	?
FRS	?
PAC	?
PAS	?
OCC	?
OCS	?

Table 1. The empty form CR6Form.CSV. It is this form that you fill in with measurements taken on your unknown cranium.

1. Into Table 1, type in an identifying label for your specimen to replace the question mark in the fourth row.

---

[6] See footnote on page 3 regarding the decimal symbol.

2. Replace each of the 29 question marks (that follow a variable name) with your measured values. Later in these notes you will find definitions of each of these measurements. ***Note the idiosyncratic definitions that Howells' uses for FMB and ZMB (see pages 40-41 below).*** In this version of the program CRANID you must enter all 29 values; in other words, missing values are not catered for because the coefficients that analyse your cranium were calculated for all 29 measurements in the database. You can enter any number of decimal places, but the original database has measurements to the nearest mm. The same should suffice for your individual cranium.

3. Save your file with a real, archival name, eg. Shoalhaven.CSV, to an archival directory. This will not be the C:\Cr6b directory that has the program Cr6bInd.EXE.

4. Save the same file again, this time as MYFILE.CSV and to the directory C:\Cr6b. This is the file that will be used by the programs in Cr6bInd.

This file MYFILE.CSV is always over-written when subsequent analyses are done, which is why you also saved your entries archivally under (step 3 above).

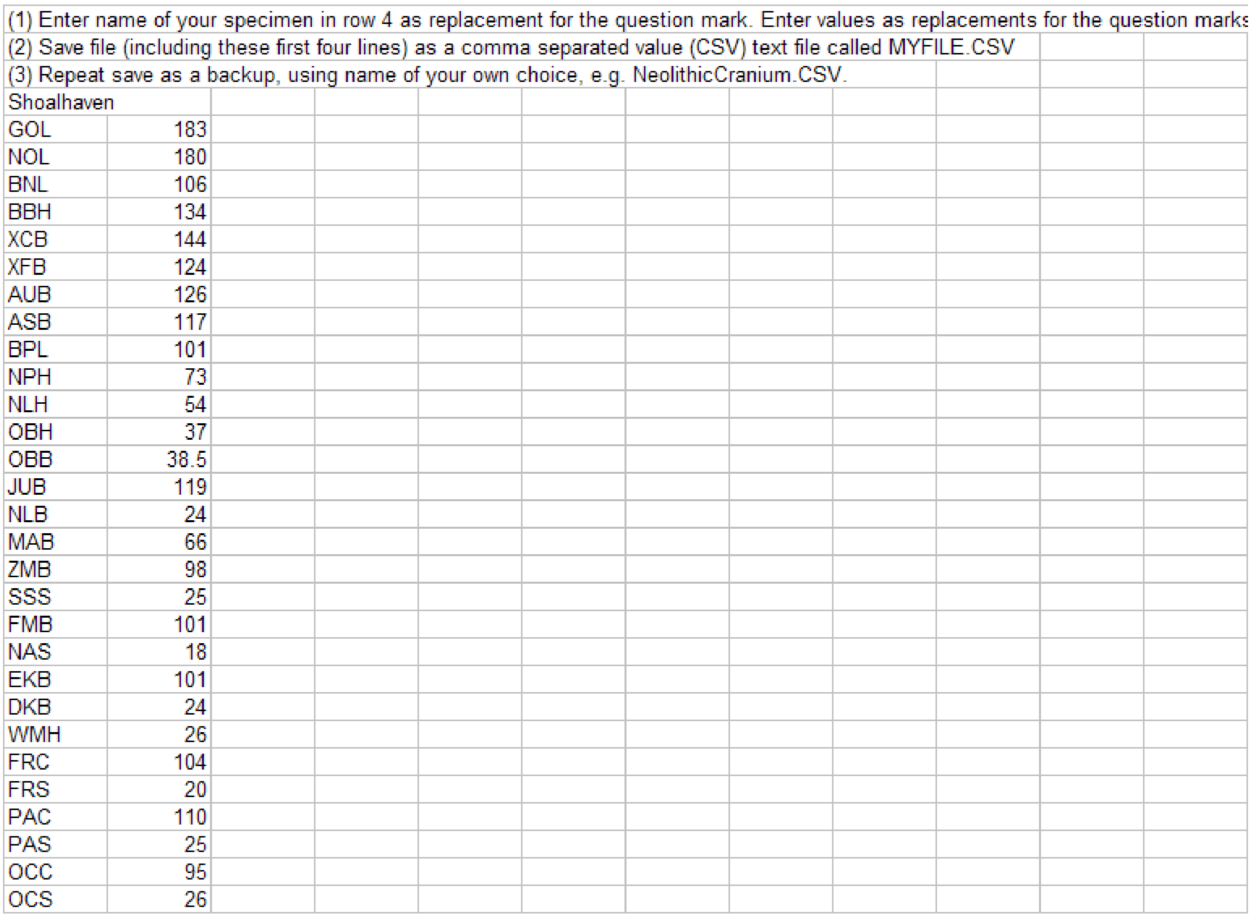
Table 2 below shows an example of a filled in form:

Note that with this version of CRANID you must have a cranium that is complete enough for you to get (or reliably estimate) all 29 measurements. No missing values are catered for. If you want to analyse an incomplete cranium then see section on CRANID services in **Charges for Additional Analysis** below.[7]

How do you analyse one of the files that is saved under its real archival name? Simply open it with the spreadsheet program and save it as MYFILE.CSV into the directory C:\Cr6b, thereby leaving the real archival file intact in its directory. Do not merely change the real file's name to MYFILE.CSV, or you may lose track of it later on.

---

[7] For an example of CRANID applied to an incomplete set of variables see: Hughes, S., Wright, R. & Barry, M. (2005). Virtual reconstruction and morphological analysis of the cranium of an ancient Egyptian mummy. Australasian *Physical & Engineering Sciences in Medicine*. 28(2), 122-127. This case also has interesting forensic implications.



(1) Enter name of your specimen in row 4 as replacement for the question mark. Enter values as replacements for the question marks	
(2) Save file (including these first four lines) as a comma separated value (CSV) text file called MYFILE.CSV	
(3) Repeat save as a backup, using name of your own choice, e.g. NeolithicCranium.CSV.	
Shoalhaven	
GOL	183
NOL	180
BNL	106
BBH	134
XCB	144
XFB	124
AUB	126
ASB	117
BPL	101
NPH	73
NLH	54
OBH	37
OBB	38.5
JUB	119
NLB	24
MAB	66
ZMB	98
SSS	25
FMB	101
NAS	18
EKB	101
DKB	24
WMH	26
FRC	104
FRS	20
PAC	110
PAS	25
OCC	95
OCS	26

Table 2. Example of a filled in form.

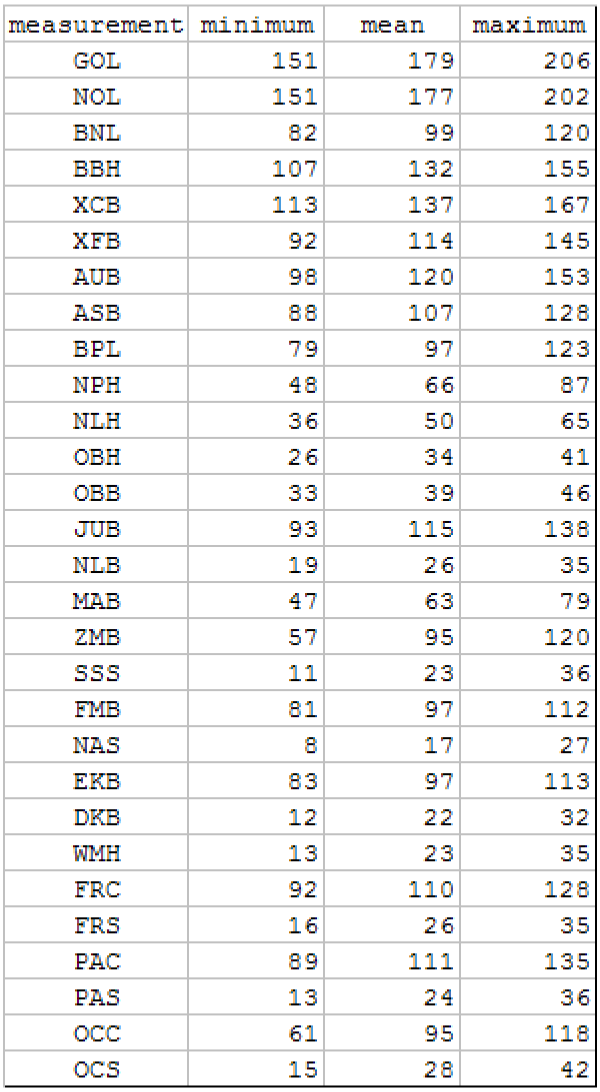
ERRORS IN MEASUREMENT – USING A TABLE OF MINIMA AND MAXIMA IN THE DATABASE OF 3,163 CRANIA

Table 3 will help you avoid errors of measurement. It shows the minimum, mean and maximum values for each variable in the database of 3,163 crania.

It is of course possible that you have a cranium with a variable that truly lies outside the range of the CRANID database, but this is not likely. ***For a more rigorous approaches to error detection, see Appendix 1.***

RUNNING THE PROGRAMS

To run the program, find the file RUN!!!!.EXE with Windows Explorer, and double click on it.



measurement	minimum	mean	maximum
GOL	151	179	206
NOL	151	177	202
BNL	82	99	120
BBH	107	132	155
XCB	113	137	167
XFB	92	114	145
AUB	98	120	153
ASB	88	107	128
BPL	79	97	123
NPH	48	66	87
NLH	36	50	65
OBH	26	34	41
OBB	33	39	46
JUB	93	115	138
NLB	19	26	35
MAB	47	63	79
ZMB	57	95	120
SSS	11	23	36
FMB	81	97	112
NAS	8	17	27
EKB	83	97	113
DKB	12	22	32
WMH	13	23	35
FRC	92	110	128
FRS	16	26	35
PAC	89	111	135
PAS	13	24	36
OCC	61	95	118
OCS	15	28	42

Table 3. Some statistics of the CRANID6 database.

The program will automatically open the prepared data file MYFILE.CSV and analyse it.

The program will automatically write out the report file to the directory C:\Cr6b. When the analysis is finished (and it only takes a second or two) you will be told the name of the report file, which is numerically incremented each time the program is run. If the previous output file was OUT013.TXT your current result will be in OUT014.TXT.

So results of each cranium analysed will never be overwritten. Nevertheless it will be an advantage to rename your output file of results to signify the real name of the cranium you have analysed, e.g. rename OUT014.TXT to Report_Shoalhaven.TXT

There are warnings built into the program:[8]

1. If there is a blank cell or successive commas in MYFILE.CSV then the program cannot continue and tells you so. There will be no report file generated.

    To correct this problem, have another look at MYFILE.CSV in your spreadsheet program. You may be able to spot the mistake.

    If you see no mistake then run a text editor and open MYFILE.CSV. Have a look for successive commas.

    If the successive commas are within the text of the numbers then cut them to one comma

    If the successive commas come after all the numbers then delete them altogether.

    Save your file and try again.[9]

2. If any of the 29 values you enter is outside the minima and maxima of the database then you will see a warning in the report. If this happens, check what you entered and/or measure again. If the warning persists then there are two possible explanations:

    For your cranium the variable(s) is truly outside the range of the database of the 3,163 crania;

    You have misunderstood the definition of the measurement in question and done it wrongly.

It is to help you in identifying problematical values that I have given the table of minima and maxima in Table 3.

---

[8] I shall add more warnings in the light of comments I get back from users. So if you run into problems let me know. If possible I shall alleviate them in future versions. In the programming world this process is unkindly known as 'idiot-proofing'.

[9] If this problem of successive commas appears it is usually a sign that the user has unnecessarily formatted cells that are outside the area for data entry provided in CR6Form.CSV. Do not do anything fancy with CR6Form.CSV. Just replace the question marks.

TESTING THE PROGRAMS

You can at any time test whether the programs are working properly by opening the file TestFile.CSV and saving it as MyFile.CSV. The file TestFile.CSV is for the Shoalhaven cranium.

Then run the program, and check that you get the results shown below. If everything is in order, your results file will match what you see in this manual.

INTERPRETING THE RESULTS

What follows is an example of an output report file from Cr6bInd.EXE.

Use a text editor such as Notepad, or a word processor such as Microsoft Word to open the report file with the file extension .TXT.[10]

Header

```
Report file: OUT011.TXT
Linear and Nearest Neighbour Discriminant Analysis by Cr6b
distributable.
richwrig@tig.com.au
Name of Cranium: Shoalhaven
Date: 10/Aug/10   Time: 09:32
```

First check that the entry after *Name of Cranium* gives you the identifying label that you expect. If not, you have mistakenly used an earlier MYFILE.CSV belonging to another cranium. So you must open the required real archival .CSV file (eg. the file Shoalhaven.CSV of the earlier instructions) and re-save this backup file as MYFILE.CSV, overwriting the previous MYFILE.CSV. Then run the program RUN!!!!.EXE again.

The file OUT011.TXT will not be over-written by subsequent use of the software, but you may care to rename it with a real name such as *Shoalhaven_Cr6b_Report.TXT.*

---

[10] Your text editor or word processor may show lines that are broken, because they are too wide for the screen. If this happens then you can widen the window on the screen. If that does not work then ensure that the text is set to a font size of 10. Always use a Courier, or other monospaced font to retain the alignment of columns in the tables.

Goodness of fit of your cranium with the CRANID6 database.[11]

In the case of Shoalhaven, this part of the report is telling us that there are no problems with this cranium so far as the database goes.

```
IS Shoalhaven WELL CATERED FOR BY THE CRANID6 DATABASE?

******

Distance of Shoalhaven from its nearest neighbour:
    4.687

Very well catered for. Distance from nearest neighbour is
less than the mean for the database.

*****

Distance of Shoalhaven from the centroid:
    5.919

Acceptable.
```

However with other crania you may get warning signs that the cranium is not a good fit with the database.

The chief causes of this outcome are:

errors in measurement;

deformity of the cranium.

See more discussion of this topic below, in the section *MORE ON GOODNESS OF FIT* on page 19 and in Appendix 1.

Warning that one or more measurements is out of range.

You will be warned if one or more of your measurements is out of range of those in the database.

In the case of Shoalhaven, none of the measurements is out of range.

Output of results for Linear Discriminant Analysis (LDA)

```
LINEAR DISCRIMINANT ANALYSIS (LDA)

Order   Sample                 Probability
  1    Denmark Neol M            0.30657
  2    London Mdvl M             0.25040
  3    Norse Norway Mdvl M       0.18460
```

---

[11] For more on this topic see below on p.##

4	Berg Austria Mdvl M	0.10510
5	Italian post-Mdvl M	0.04910
6	San Cruz I Calif M	0.02958
7	Zalavar Hung. Mdvl M	0.02049
8	N. Japan Hokkaido M	0.01032
9	Peru Youyos M	0.00896
10	Denmark Neol F	0.00621
11	Egypt 26-30 Dyn M	0.00592
12	Arikara Dakota M	0.00557
13	Patagonian F	0.00551
14	Poundbury UK RB M	0.00233
15	London Mdvl F	0.00165
16	Italian post-Mdvl F	0.00142
17	Patagonian M	0.00100
18	Moriori Chat Is F	0.00091
19	Norse Norway Mdvl F	0.00075
20	Ainu Hokkaido M	0.00067
21	S. Japan Kyushu M	0.00059
22	Arikara Dakota F	0.00053
23	Hainan China M	0.00047
24	Berg Austria Mdvl F	0.00035
25	Moriori Chat Is M	0.00027
26	Zalavar Hung. Mdvl F	0.00016
27	San Cruz I Calif F	0.00008
28	Punjab M	0.00008
29	Ainu Hokkaido F	0.00007
30	Mokapu Hawaii M	0.00006
31	Maori New Zealand M	0.00005
32	Lachish W Asia M	0.00005
33	Philippines M	0.00004
34	Beduin W Asia MF	0.00003
35	Atayal Taiwan M	0.00003
36	Peru Youyos F	0.00002
37	Sydney Aboriginal M	0.00001
38	N. Japan Hokkaido F	0.00001
39	Poundbury UK RB F	0.00001
40	Lachish W Asia F	0.00001
41	Buriat Siberia M	0.00000
42	Egypt 26-30 Dyn F	0.00000
43	S. Japan Kyushu F	0.00000
44	Easter I. M	0.00000
45	Hainan China F	0.00000
46	Mokapu Hawaii F	0.00000
47	Anyang China M	0.00000
48	Eskimo Greenland M	0.00000
49	Andaman Is. M	0.00000
50	Buriat Siberia F	0.00000
51	India M	0.00000
52	Zulu S. Afr M	0.00000
53	Guam Latte Period F	0.00000
54	Tolai New Britain M	0.00000
55	Guam Latte Period M	0.00000
56	Atayal Taiwan F	0.00000
57	Dogon W. Afr M	0.00000
58	S Australia M	0.00000
59	Teita E. Afr M	0.00000
60	Tolai New Britain F	0.00000
61	Sydney Aboriginal F	0.00000
62	S Australia F	0.00000
63	Eskimo Greenland F	0.00000
64	Tasmania M	0.00000

```
65    Easter I. F              0.00000
66    Zulu S. Afr F            0.00000
67    India F                  0.00000
68    Punjab F                 0.00000
69    Tasmania F               0.00000
70    Teita E. Afr F           0.00000
71    Dogon W. Afr F           0.00000
72    Andaman Is. F            0.00000
73    Bushman Afr M            0.00000
74    Bushman Afr F            0.00000
```

Table 4. Output of LDA results, for *Shoalhaven*.

In the results of Table 4 you will find all the samples in the database ranked in descending order of probability. The probability is that the individual comes from the associated sample. You will normally want to report the first few samples, but all are given because it is sometimes useful to know what samples are most distant from an unknown cranium.

The probabilities are expressed to five decimal places, and technically never reach zero.

I have found it useful, in a summary report, to include only those samples that have a probability rounding up to 0.1. I have highlighted the relevant samples in blue.

Output of results for Nearest Neighbour Discriminant Analysis (NNDA)

The first part of the output for NNDA is a tallied list of the sample membership of the first 56 nearest neighbours:

```
NEAREST NEIGHBOUR DISCRIMINANT ANALYSIS (NNDA)

Actual nearest neighbour of Shoalhaven, is from: Italian_M

Sample number Sample name            Hits  Weighted score
   1     Norse Norway Mdvl M         6      345
  63     London Mdvl M               5      304
   3     Berg Austria Mdvl M         5      282
  22     Egypt 26-30 Dyn M           5      273
  72     Denmark Neol M              4      253
  68     Italian post-Mdvl M         5      251
  41     Moriori Chat Is F           4      248
  71     Denmark Neol F              1      186
  14     San Cruz I Calif M          3      186
  13     Arikara Dakota M            2      151
   2     Zalavar Hung. Mdvl M        2      119
  15     Peru Youyos M               2      115
  12     Moriori Chat Is M           2      111
  66     Patagonian F                1       90
  53     Ainu Hokkaido F             1       83
  31     Zalavar Hung. Mdvl F        1       70
   5     Dogon W. Afr M              1       67
```

```
25     Ainu Hokkaido M            1     66
64     London Mdvl F              1     65
17     S. Japan Kyushu M          1     63
43     San Cruz I Calif F         1     62
30     Norse Norway Mdvl F        1     58
44     Peru Youyos F              1     58

Classified by weighted score: Norse_Norway_M
```

Table 5. Tallied nearest neighbour results for *Shoalhaven*.

Table 5 shows a tally of the 56 nearest morphological neighbours to the cranium *Shoalhaven.*[12]

The first line of this section shows the sample membership of the particular cranium in the database that is most like Shoalhaven (a cranium in the Italian male sample). The last line shows how Shoalhaven is classified after the results are tallied; it is classified with the Norse male sample.

The column headed *Hits* shows how many crania from each sample were found among the 56 nearest neighbours.

There is a logical inadequacy in simply relying on a tally of hits. In relying on hits we are overlooking the fact that samples in the database are of unequal size (see Table 12). Therefore samples that are large in number will tend to have, all other things being equal, more individuals among the nearest neighbours. So a second statistic of *Weighted score* is computed and forms the basis for ordering the samples and working out the classification. The weighted score is proportional to the number of crania in a sample.

As with the report for LDA, it is often helpful to have a cut off point for summary reporting. I have found that a weighted score of 300 is a useful cut off point.

The second part of the NNDA report is a simple list of the 56 nearest neighbours. We have already seen that the actual nearest neighbour of Shoalhaven is an Italian male. The complete list, in increasing order of distance from Shoalhaven, is:

---

[12] The inclusion of 56 nearest neighbours is a convenient but arbitrary choice derived from the square root of the number of crania in the database. There is no a priori way of determining the ideal number of nearest neighbours. If you want fewer then you can retabulate them by taking them top downwards from the results presented in the second part of the NNDA output.

NEAREST 56 NEIGHBOURS TO Shoalhaven, IN INCREASING ORDER OF DISTANCE

Individual	Sample number	Sample name	Order	Distance
2918	68	Italian post-Mdvl M	1	4.687
106	2	Zalavar Hung. Mdvl M	2	5.200
607	12	Moriori Chat Is M	3	5.327
2747	63	London Mdvl M	4	5.361
23	1	Norse Norway Mdvl M	5	5.381
1022	22	Egypt 26-30 Dyn M	6	5.425
2750	63	London Mdvl M	7	5.448
2396	53	Ainu Hokkaido F	8	5.466
147	3	Berg Austria Mdvl M	9	5.497
856	17	S. Japan Kyushu M	10	5.502
36	1	Norse Norway Mdvl M	11	5.516
44	1	Norse Norway Mdvl M	12	5.579
3111	72	Denmark Neol M	13	5.593
156	3	Berg Austria Mdvl M	14	5.685
741	15	Peru Youyos M	15	5.731
2731	63	London Mdvl M	16	5.756
144	3	Berg Austria Mdvl M	17	5.800
1155	25	Ainu Hokkaido M	18	5.836
628	13	Arikara Dakota M	19	5.900
1915	41	Moriori Chat Is F	20	5.994
695	14	San Cruz I Calif M	21	6.003
2017	43	San Cruz I Calif F	22	6.019
1943	41	Moriori Chat Is F	23	6.070
3110	72	Denmark Neol M	24	6.084
2939	68	Italian post-Mdvl M	25	6.101
3082	72	Denmark Neol M	26	6.110
93	2	Zalavar Hung. Mdvl M	27	6.128
2064	44	Peru Youyos F	28	6.131
124	3	Berg Austria Mdvl M	29	6.149
2934	68	Italian post-Mdvl M	30	6.161
1405	30	Norse Norway Mdvl F	31	6.179
610	13	Arikara Dakota M	32	6.188
13	1	Norse Norway Mdvl M	33	6.213
2730	63	London Mdvl M	34	6.215
1041	22	Egypt 26-30 Dyn M	35	6.225
2851	66	Patagonian F	36	6.247
1026	22	Egypt 26-30 Dyn M	37	6.258
230	5	Dogon W. Afr M	38	6.271
689	14	San Cruz I Calif M	39	6.279
2788	64	London Mdvl F	40	6.289
1457	31	Zalavar Hung. Mdvl F	41	6.298
2714	63	London Mdvl M	42	6.300
1054	22	Egypt 26-30 Dyn M	43	6.307
3063	71	Denmark Neol F	44	6.317
725	15	Peru Youyos M	45	6.328
2954	68	Italian post-Mdvl M	46	6.358
1040	22	Egypt 26-30 Dyn M	47	6.362
3112	72	Denmark Neol M	48	6.368
1911	41	Moriori Chat Is F	49	6.378
5	1	Norse Norway Mdvl M	50	6.379
553	12	Moriori Chat Is M	51	6.381
137	3	Berg Austria Mdvl M	52	6.384
38	1	Norse Norway Mdvl M	53	6.387
1948	41	Moriori Chat Is F	54	6.389
697	14	San Cruz I Calif M	55	6.393
2955	68	Italian post-Mdvl M	56	6.401

Table 6. Part 2 of nearest neighbour results for *Shoalhaven*.

Table 6 is the output that is tallied for presentation in Table 5. It gives the user a clear, and cautionary, idea of the overlap in cranial form that exists within samples of our species. Identification is very much an example of the principle 'on the balance of probabilities'.

CUT OFF POINTS FOR REPORTING

I have suggested cut off points for summary reporting. For LDA I suggested probabilities that round up to 1.0 and for NNDA weighted scores that are equal or above 300.

For some purposes it is entirely inappropriate to set any cut off point. For example, it may be required to show that a cranium cannot be, *beyond reasonable doubt,* a member of a particular set of samples, e.g. cannot, beyond reasonable doubt, be of Australian Aboriginal ancestry.

Such was the case with Shoalhaven, which was at some time in the past labelled as *Australian Aboriginal.* We have already seen that the samples within the cut off zones are of samples that are geographically from the Europe/Mediterranean area. This result for Shoalhaven comes from both LDA and NNDA.

Given the research objective, we might also want to know how close any Australian Aboriginal samples are to Shoalhaven. If we use the suggested cut offs then this information is lost.

In the case of Shoalhaven we see two relevant points in the extended lists:

> in LDA there is no Australian Aboriginal sample until we reach sample number 37 in the list, and the probability of Shoalhaven being a member of this sample is around 1:100,000.
>
> we also see that in the results for NNDA, there is not a single Australian cranium represented in the list of 56.

In brief, and given the purposes of certain analyses, reporting the samples that an unknown cranium *least* resembles may be as useful as reporting those that it *most* resembles.

MORE ON GOODNESS OF FIT

Remember there will always be a nearest neighbour to an unknown, however far away this nearest neighbour might be in 29 dimensional space.

Enter the 29 measurements of the cranium of an orangutan and you will get a result. However the results will be peppered with warnings of the lack of fit between the orangutan and the database.

Therefore it is useful to look at the two statistics of

(1) distance of unknown from its nearest neighbour in the space of 29 dimensions;

(2) distance of unknown from the centroid of the 29 dimensions.

This topic was broached above. It needs more explanation.

The first point is technical. The units of distance reported for NNDA are derived from the 29 dimensional space of the unstandardised canonical variate scores. They do not have a link with any other reality.

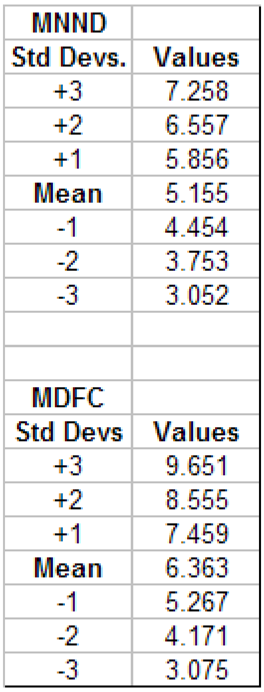
To this end, the database of 3,163 crania has been looked at to compute *Mean Nearest Neighbour Distance* (MNND) and the associated standard deviation. Also computed is the *Mean Distance From Centroid* (MDFC), again with its associated standard deviation.

Table 8 will help you work out how well the database caters for an unknown cranium. It will do this by answering the two questions of:

whether the unknown is exceptionally far from its nearest neighbour

whether the unknown is, as it were, in 'outer space' when we compute its distance from the centroid.

This table is wired into the software, and does not normally need to be consulted if the output report does not detect a problem with the unknown cranium being catered for by the database.



MNND	
Std Devs.	**Values**
+3	7.258
+2	6.557
+1	5.856
Mean	5.155
-1	4.454
-2	3.753
-3	3.052
MDFC	
Std Devs	**Values**
+3	9.651
+2	8.555
+1	7.459
Mean	6.363
-1	5.267
-2	4.171
-3	3.075

Table 8. Values for MNND and MDFC in the database of 3,163 crania in the CRANID6 database. This table is used to work out whether an unknown cranium is well or poorly catered for by the database.

For the cranium *Shoalhaven*, the distance from its nearest neighbour is 4.687, which places it slightly under the mean value of 5.155 for the MNND of the CRANID6 database. So it is slightly closer to its nearest neighbour than the average, so this cranium clearly had no problem in finding a nearest neighbour.

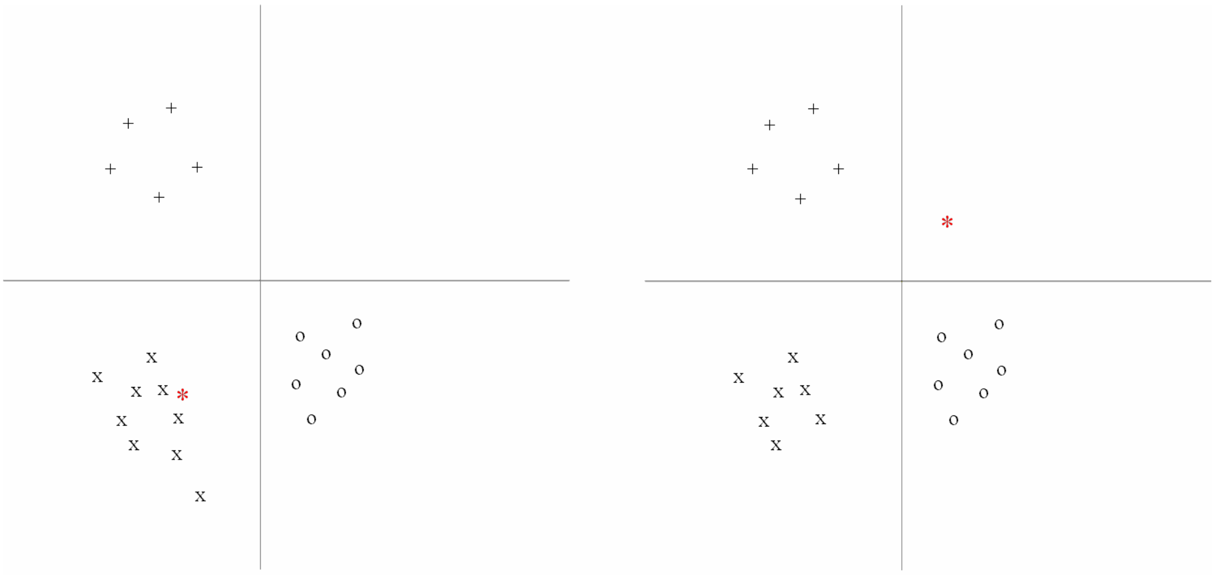
The distance from centroid of *Shoalhaven* is 5.919, which places it closer to the centroid than the mean value of 6.363 for the MDFC of the CRANID6 database. In other words, Shoalhaven is far from being an outlier compared with other crania in the database.

In summary, the cranium *Shoalhaven* is well catered for by the CRANID6 database.

The principles being used are illustrated in Figure 1.

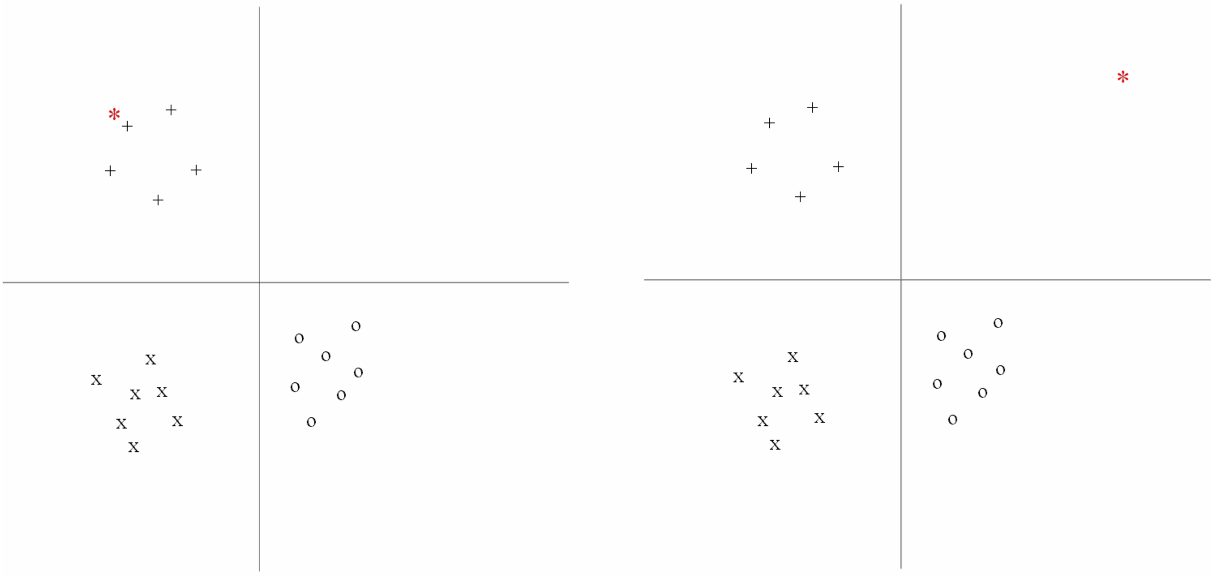
If a cranium yields an MNND that is below average then it is closer than average to its nearest neighbour. If the MNND is above average then it is further than average from its nearest neighbour.

If a cranium yields an MDFC that is below average then it is rather closer to the centroid than the average cranium. If the MNND is above average then it is further than the average cranium from the centroid.



Low on MNND, low on MDFC.
Particularly well catered for by data.

High on MNND, low on MDFC.
Poorly catered for by data.



Low on MNND, high on MDFC.
Well catered for by data.

High on MNND, high on MDFC.
Particularly poorly catered for by data.

Figure 1. Simulating the meaning of MNND and MDFC in two dimensional space.

Using the CRANID6 database, the results for MNND and MDFC are computed in the 29 dimensional space of the canonical variates. We cannot imagine the patterns of the 3,163 crania in this hyperspace. Yet the meaning of high and low values of MNND and MDFC can be simulated by idealised locations of an individual cranium (the red asterisk) in relation to the distribution of all the crania.

Returning to consider the actual database, any cranium that has values of greater than +2 standard deviations above the average, whether for MNND or MDFC, must be considered as poorly catered for by the database. Above +3 standard deviations, and it must be considered very poorly catered for. The classifications reported must be treated with suspicion. There is a lack of fit between the cranium and the database.

Any lack of fit to the database may be due to one or more of several unrelated reasons, including:

> the data of the unknown cranium are wrongly measured or wrongly entered into the file MyFile.CSV; common causes of mismeasurement are misunderstanding of a measurement itself (go back to the definitions) and misreading of the calipers;[13]
>
> the geographical area from which the cranium came is poorly represented in the database;
>
> the cranium is morphologically atypical of its group, for example because of unusual growth or artificial deformation;
>
> the cranium is not deformed, but is an extreme member of its group - that group being itself at the extreme of the distribution of the samples in 29 dimensional space;[14]
>
> the person is of mixed ancestry.

Here is an example of the warning that results if we make a mistake with Shoalhaven, erroneously entering a value for BNL of 120 mm instead of the actual 106 mm.

---

[13] In my experience, a common source of error is misreading of the calipers. It is not so often that the value itself is misread, but the wrong part of the caliper is used to read the number. I suggest that you mark the correct part of the caliper with something that draws your attention to it - such as white-out.

[14] We cannot see this scenario in 29 dimensional space. If, however, we look at the 2D scattergram of Figure 9 we can imagine a cranium that is a 'member' of the fringe London Medieval group, but is well beyond the plotted value. In such a position it finds it difficult to locate a nearest neighbour and is much further from the centroid than even the sample mean.

Cr6b reports:

```
IS Shoalhaven WELL CATERED FOR BY THE CRANID6 DATABASE?

******

Distance of Shoalhaven from its nearest neighbour:
    7.517

Poorly catered for. Distance from nearest neighbour is
between three and four standard deviations from mean
for database.

You should consider whether there are mistaken measurements
entered or whether the cranium is deformed.

*****

Distance of Shoalhaven from the centroid:
    9.874

Poorly catered for. Distance from centroid is greater than
three standard deviations from centroid.

EXTREME CAUTION NECESSARY given difficulty cranium also had
finding a nearest neighbour.
```

Such a warning does not necessarily mean that the ancestry subsequently reported will be misleading.

For example, the summary LDA results for Shoalhaven, with the mistaken value for BNL, are:

```
LINEAR DISCRIMINANT ANALYSIS (LDA)

Order   Sample              Probability
  1   London_Mdvl M          0.34636
  2   Denmark Neol M         0.28876
  3   Norse Norway Mdvl M    0.13005
  4   Italian post-Mdvl M    0.11081
  5   N. Japan Hokkaido M    0.06975
```

So even with the mistake we would conclude that Shoalhaven had an ancestry that was geographically Europe/Mediterranean.

Note that even the imagined BNL of 120 is not out of range, hence the more subtle interrogation of the results that this section discusses.

However one is risking false results by ignoring the warnings without investigation. One should only proceed after going back to first base. Do you understand the measurements? Have you read the calipers correctly? Have you entered the correct measurements into the CRANID form? Have you checked your measurements by using the methods given in Appendix 1?

If none of the above is at fault, then (and only then) consider the question of whether the cranium is an extreme member of its group - that group being itself at the extreme of the distribution of the samples in 29 dimensional space. One test of this explanation is that the list of results show others near the top of the list that are geographically relevant to the identification of the unknown. For example, if the unknown is identified with a particular European/Mediterranean sample - ask yourself whether the next few in the list are also European/Mediterranean samples.

VALIDITY OF RESULTS

Linear discriminant analysis (LDA) makes assumptions about the distribution of the underlying data from which the samples have been drawn. These assumptions particularly affect the estimation of probabilities. The two major key assumptions for linear discriminant analysis are that the variables are multivariately normal and that the dispersion and covariance matrices for the groups are equal.

These issues are widely discussed on the web. Search for keywords "discriminant analysis" and "assumptions". There are no reasons that I know of for suspecting that the CRANID data deviates from the assumptions so much that the *general* magnitudes of the probabilities are regularly invalid.[15]

You can avoid the need to make these assumptions by using *Nearest Neighbour Discriminant Analysis*, which is included in the report on each cranium. For crania on which I have done parallel analyses (linear discriminant and nearest neighbour analyses) I have found reasonable agreement between the two methods, such that either method normally gives a reasonable prediction of the other.

---

[15] Linear discriminant analysis is the method used by FORDISC. http://web.utk.edu/~anthrop/FACpubs.html

AN EXAMPLE OF CRANID ANALYSIS APPLIED TO A CRANIUM BEING CONSIDERED FOR REPATRIATION

In my experience, most cases being considered for repatriation produce CRANID results that are consistent with the ancestry imputed by their documentation.

Shoalhaven is an interesting exception, the context of which I will briefly outline. I will also present the results to show differences that can arise between the results of linear and nearest neighbour discriminant analysis.

Briefly, a cranium from the Shoalhaven district, south of Sydney, was being considered for repatriation. It was housed with Aboriginal skeletal remains and had a label tied to it attributing it to a named person, followed by the writing "(Aboriginal)".

A more recent note in the box, written by an anthropologist, suggested the person was of mixed ancestry.

As we have seen, the CRANID results give no support whatever to either entirely Aboriginal ancestry or mixed ancestry.

The results for *Shoalhaven* have been already presented in Tables 4-6. We saw that *Shoalhaven* is entirely compatible with samples of European origin.

Note that in Table 4 Australian Aboriginal samples appear far down in the list of 74 samples – in places 37, 58, 61, 62, 64 and 69.

Also, Table 5 shows that not one of the 56 nearest neighbours is an Australian Aboriginal.

So morphological analysis gives no ground for concluding that this *Shoalhaven* cranium has any Australian Aboriginal ancestry.

Subsequent historical research proved this morphological conclusion to be correct. The person named in copperplate writing on the cranium was not Aboriginal. He was a brewer from Yorkshire, shipped out to Australia as a convict. He committed suicide in 1842.

For *Shoalhaven* there are some differences between the results for linear and nearest neighbour discriminant analysis, though none of them suggest Aboriginal ancestry. Let us look at the results from LDA again, this time considering only the major probabilities:

LDA - size and shape

Order	Sample	Probability
1	Denmark Neol M	0.30657
2	London Mdvl M	0.25040
3	Norse Norway Mdvl M	0.18460
4	Berg Austria Mdvl M	0.10510

Table 9. Closest samples for linear discriminant analysis of Shoalhaven.

In Table 9, the sample with the highest probability is from the Danish Neolithic, but the other results are entirely European and male.

Let us also look at the results from NNDA again, this time considering only the major weighted scores:

NNDA - size and shape

Actual nearest neighbour is an Italian post Mdvl M

	Sample name	Hits	Weighted score
1	Norse Norway Mdvl M	6	345
63	London Mdvl M	5	304
3	Berg Austria Mdvl M	5	282
22	Egypt 26-30 Dyn M	5	273
72	Denmark Neol M	4	253

Table 10. Major nearest neighbour results for Shoalhaven.

In the results of Table 10 we see a pretty even spread of weights over the five samples, but with the Danish Neolithic relegated to fifth place.

What do we make of these differences between linear and nearest neighbour discriminant analysis? Using a broad-brush approach (which is what one should do with such analyses) we can be confident that Shoalhaven is very close to European males in morphology.

We have evidently been looking at a case where the assumptions needed by LDA are not quite met, hence the differences between the two analyses. How do these types of difference come about? To answer the question we would have to look at the distribution of the crania in 29 dimensions, which we obviously cannot do. However the following chart of hypothetical data shows the problem in two dimensional space.

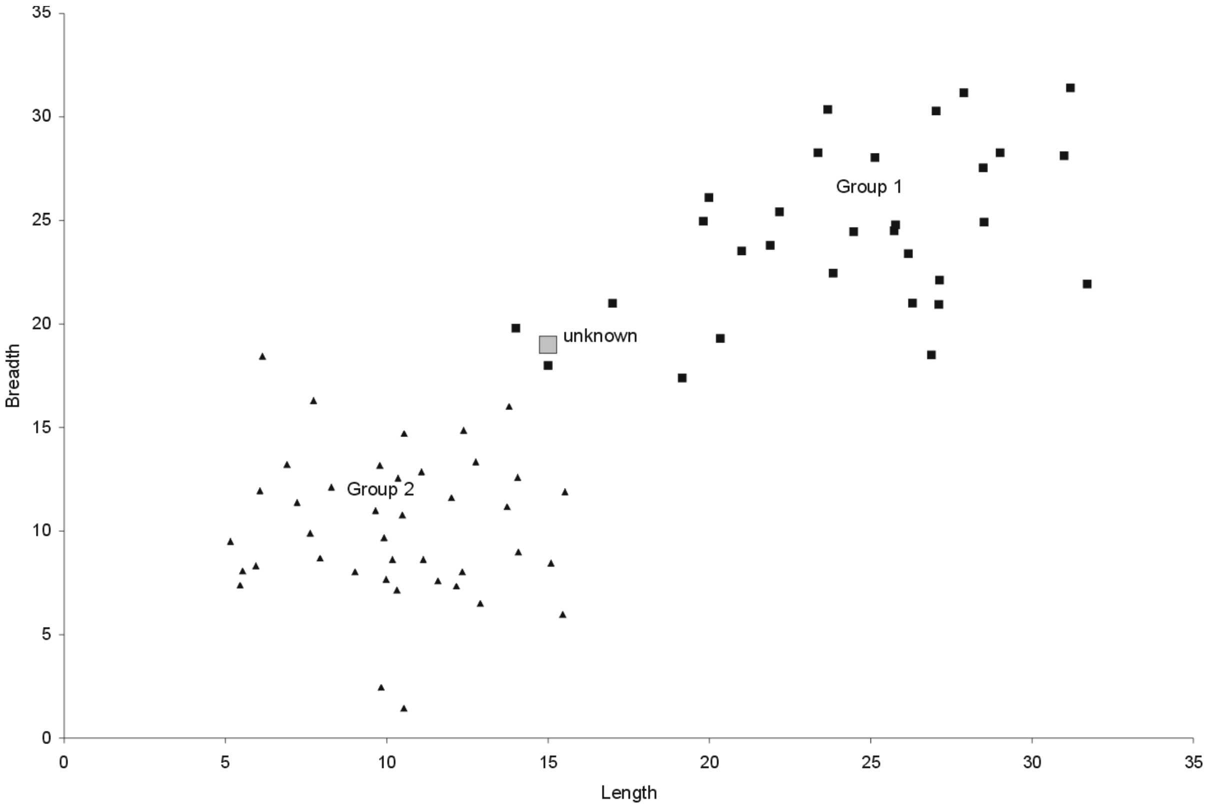


Figure 2. Hypothetical example, in two dimensions, to show why there can be differences between the results of linear and nearest neighbours discriminant analysis in multidimensional space.

Figure 2 shows two groups. We see that Group 2 is relatively coherent and circular, whereas Group 1 has a long tail stretching toward Group 2.

> Linear discriminant analysis of the data of the two measurements classifies the unknown with Group 2 because the unknown is closer to the centroid of Group 2. The method takes no account of the tail of Group 1. Indeed it assumes that both groups are drawn from populations that are circular in their distribution.
>
> Nearest neighbour discriminant analysis shows that the two nearest neighbours of the unknown are members of Group 1. The method takes account of the tail of Group 1.

In brief, linear discriminant analysis is what is known as a parametric method (it makes assumptions about the data from which the database is constructed). Nearest neighbour discriminant analysis is a non-parametric method. It does not make the same assumptions about the data.

AN EXAMPLE OF CRANID ANALYSIS APPLIED TO AN ANCIENT CRANIUM

In the previous discussion of *Shoalhaven*, we have seen that you will have to reconcile the somewhat differing results of the methods outlined. This reconciliation proved easy with *Shoalhaven*, given the objectives of the analysis.

Sometimes reconciliation will not be easy and will be open to much judgment.

The following hints will help you interpret differences between results that might otherwise be difficult to explain. Note that these hints include results from the distributable package Cr6bInd.EXE and also additional results that you would have received from me if you had asked for a complete CRANID analysis. In other words the following results include the analysis by size and shape in the distributable package, and by shape alone. The latter is not in the distributable package.

The following example is a full CRANID analysis of the famous individual UC-101, a Late Pleistocene cranium of *Homo sapiens* from the Upper Cave of Beijing.

The cut offs have been slackened to an LDA probability of 0.01 and a weighted score of 200.

METHOD I - LDA size and shape *(done by the distributable CRANID program)*

Order	Sample	Probability
1	Norse Norway Mdvl M	0.61
2	Zalavar Hung. Mdvl M	0.20
3	Buriat Siberia M	0.08
4	S Australia M	0.04
5	Italian post-Mdvl M	0.02
6	Patagonian M	0.02
7	Berg Austria Mdvl M	0.01

METHOD II - LDA shape only *(done by the non-distributable CRANID program)*

Order	Sample	Probability
1	Norse Norway Mdvl M	0.53
2	Zalavar Hung. Mdvl M	0.33
3	S Australia_M	0.04
4	Berg Austria Mdvl M	0.02
5	Italian post-Mdvl M	0.02
6	Tasmania M	0.01

METHOD III - NNDA size and shape *(done by the distributable CRANID program)*

Order	Group	Hits	Score
1	S Australia M	6	365
2	Buriat Siberia M	6	345
3	Zalavar Hung. Mdvl M	5	298
4	Norse Norway Mdvl M	4	230
5	Tasmania M	3	211
6	Italian post-Mdvl M	4	201

METHOD IV - NNDA shape only *(done by the non-distributable CRANID program)*

Order	Group	Hits	Score
1	Berg Austria Mdvl M	5	282
2	S Australia M	4	243
3	Zalavar Hung. Mdvl M	4	239
4	Arikara Dakota_F	2	234
5	Norse Norway Mdvl M	4	230

Table 11. CRANID analysis of UC101, also using methods of shape not included in Cr6bInd.EXE. Results are abbreviated.

These results for UC-101 are too inconsistent to suggest a likeness to any particular CRANID sample.

We might nevertheless conclude, given the differences between the results for LDA and NNDA on shape only, that the cranium is affected by the statistical assumptions underlying LDA.

In summary, UC101 seems to occupy a place in multivariate space that is somewhat distant from any sample in the database, and a space that is intermediate between samples that are not thought to be closely related to each other.

At this point we should consider the distance of UC101 from its nearest neighbour and from the centroid. These statistics are available for size and shape:

```
IS UC101.CSV WELL CATERED FOR BY THE CRANID6 DATABASE?

******

Distance of UC101.CSV from its nearest neighbour:
    7.040

Not well catered for. Distance from nearest neighbour
is between two and three standard deviations from mean
for database.

*****

Distance of UC101.CSV from the centroid:
    8.872

Acceptable.
```

Being somewhat high in relation to both MNND and MDFC, UC101 is rather poorly catered for by the database. See Figure 1 for an explanation of what this means. UC101 conforms best with the model on the bottom right of Figure 1, though its position is less extreme than the case that this model illustrates.

The measurements for this cranium have been checked and rechecked. There is no reason to suspect that errors have been made or that the cranium is deformed.

It is interesting to note, however, that the cranium has been said to be Mongoloid in shape[16], a conclusion already negated in pre-CRANID days.[17] A Mongoloid affinity was also negated by an earlier version of CRANID[18], and recently by primitive LDA using the so-called Giles-Elliot coefficients.[19]

A full CRANID analysis now shows again that UC-101 is ruled out from being morphologically Mongoloid.

Of course it might be argued for UC-101 that a major (if undetermined) problem with its appearing not to be East Asian in morphology lies with the database of 3,163 crania, and/or the 29 variables used, and/or the algorithms of LDA and NNDA.

If it is so argued then I would challenge sceptics to produce a database, variables and algorithms that place UC-101 morphologically closer to East Asians than to other groups from around the world. Without such a demonstration I must regard the supposition of unsuitability as mere special pleading.

---

[16] Thorne, A.G. and Wolpoff, M.H. 1992. The multiregional evolution of humans. *Scientific American,* 266(4): 28-33.

[17] Kamminga, J. and Wright, R.V.S. 1988. The Upper Cave at Zhoukoudian and the origins of the Mongoloids. *Journal of Human Evolution,* 17:739-767.

[18] Wright, R.V.S. 1992. Correlation between cranial form and geography in *Homo sapiens*: CRANID - a computer program for forensic and other applications, *Archaeology in Oceania* 27:105-112.

[19] Wright, R.V.S. 1995. The Zhoukoudian Upper Cave skull and multiregionalism. *Journal of Human Evolution,* 29: 181-183.

VALIDATION BY CLASSIFICATION SUCCESS RATES

One method of testing the usefulness of linear discriminant analysis is to look at the *classificatory success rate*. What is this rate? When I originally computed the coefficients from the raw data I was, as it were, generating a set of rules that maximally separated all the 74 samples from each other. Despite this maximal separation there remains considerable overlap between individuals. One way of measuring this overlap (or lack of classificatory success) is to apply the coefficients again to the raw data of each of the 3,163 crania and see how often the crania are classified with their correct sample.

This method of resubstitution produces a classificatory success rate of 68.2% for Cr6b LDA. That is excellent when we consider that if there was no association between cranial form and the 74 geographical samples we would, on average, get a classificatory success rate by chance alone of roughly 74 / 3,163 x 100 = 2.3%.[20]

The classificatory success rate of 68.2% is even better than it looks at face value. It is a severe test. It is severe because, for example, a Romano-British male from Poundbury is wrongly classified if it is classified as a Poundbury female. This error is counted as equally mistaken as classification of a Poundbury male as a Patagonian female.

If we select and merge those samples that would, in traditional anthropology, have been classified as Caucasoid, Mongoloid and Negroid then the classification success rate is increased to 89%.

A different perspective on validity comes from putting all statistical measures aside and looking at charts. It turns out that there is visual sense in the plot of the 74 samples in the space of the first two canonical variates (discriminant functions). The scattergram (Figure 10) and dendrogram (Figure 11) at the end of these notes are computed from scores, where the scores are averages of both sexes of each group.

The method of craniometric analysis is purely morphological, and blind to the actual geographical proximities of the 74 samples. So the post hoc geographical labels on Figures 10 and 11 have been superimposed on results derived from the cranial

---

[20] 'Roughly', because the samples are not equal in size.

morphology alone. In other words, a sample known to be European in origin was free, when analysed by discriminant analysis, to position itself among the samples from the Pacific.

Encouragingly, the scattergram of the first two canonical variates resembles a map of the world, if turned anti-clockwise through 90 degrees. There is clearly a high correlation between cranial morphology and geography. The dendrogram, which uses all 29 canonical variates, also validates even better the usefulness of the database in predicting geographical origin.

The sceptical Williams *et al* have published a paper in which they suggest that "skeletal specimens or samples cannot be accurately classified by geography".[21] This conclusion is drawn from a single study of Nubian crania in which they use only 11 Howells' type measures when 21 are available in computer program Fordisc! However the faults in classification are generalised by the authors and said to (1) lie with Fordisc itself (as though 11 variables are as good as the 21 that Fordisc allows) and (2) with the concept of there being a relation between cranial form and ancestry. In my opinion the faults they identify might lie with their restricted data. This inadequate experiment is not sufficient to overthrow the correlation between geography and cranial form demonstrated in numerous studies.

The inadequacy of the paper by Williams *et al* is trenchantly dealt with by Ousley *et al*.[22]

HISTORY OF THE DATABASE

The database was started by W.W. Howells[23]. Many people, including myself, have used that database for forensic and other identifications based on cranial form.[24] My

---

[21] Frank L'Engle Williams, Robert L. Belcher, and George J. Armelagos "Forensic Misclassification of ancient Nubian crania: implications for assumptions about human variation." *Current Anthropology*, 46(2) 340-346. 2005.

[22] Ousley, S., Jantz, R. and Freid, D. "Understanding race and human variation: why forensic anthropologists are good at identifying race." *American Journal of Physical Anthropology*, 139:68-76 (2009).

[23] Howells, W.W. 1973. *Cranial Variation in Man: a Study by Multivariate Analysis of Patterns of Difference Among Recent Human Populations.* Papers of the Peabody Museum, Harvard University, Vol.67.

[24] Wright, R.V.S. 1992. Correlation between cranial form and geography in *Homo sapiens*: CRANID - a computer program for forensic and other applications, *Archaeology in Oceania* 27:105-112. In this

analyses have tended to use a subset of 29 measurements and it is these that are used in the programs of Cr6bInd.[25]

Howells did a massive job of collection and generously distributed his database on disk to interested persons in the 1980s, and later through the internet.[26]

Understandably there are geographical gaps in his database. There are no samples included from the British Isles, no samples from the Middle East (except for a Hellenistic period sample from Egypt) and no samples from the Indian sub-continent.

My version of Howells's database was first expanded in 1998, thanks to a sample of Beduin measurements supplied by Marta Lahr. Then Robert Kruszynski, of London’s Natural History Museum, collected data from Poundbury (Romano-British 4th century, Dorset, England), Lachish (Iron Age Palestine) and the Indian Subcontinent. I measured samples from Spitalfields Market (London Medieval)[27] and Patagonia[28]. These samples comprised the data for CR5Ind2.EXE – the immediate predecessor of the current Cr6 series.

For the CR6 database I have added:

1. Crania of Aborigines from the Sydney area, measured by myself; courtesy of the La Perouse Local Aboriginal Land Council, the Metropolitan Local Aboriginal Land Council, and D. Donlon (Curator of the Shellshear Museum at the University of Sydney). The addition of a sample from coastal New South Wales has proved valuable in problematical cases for repatriation.

2. Crania from the area of the Punjab. The subcontinent is not catered for at all in Howells' database. One sample was included in CR5Ind2. Now there is a second set from the subcontinent. The crania were measured by P. Raghavan, in collaboration with D. Bulbeck.

---

paper I set out various theoretical justifications for this sort of morphological work, and distinguish it from the older school of impressionistic cranial typology.

[25] There are considerable differences between what I did in my 1992 paper and what the program CR6LDA.EXE now does. In 1992 I used only crania from the Howells's database, analysed principal component scores, and used nearest neighbour analysis as the method of evaluating similarities. In CR6LDA.EXE I am using an expanded database, analyzing canonical variate scores, and using linear discriminant analysis as the method of evaluating similarities.

[26] Howells, W.W. 1996. Howells’ craniometric data on the Internet. *American Journal of Physical Anthropology*, 101, 441-442.

[27] Courtesy Museum of London.

[28] Courtesy Museo Etnográfico, Buenos Aires.

3. Recent crania from southern Italy. These were measured by R. Kruszynski of the Natural History Museum. A European sample from the area of the Mediterranean has not been included in earlier versions of CRANID.

3. Neolithic crania from Denmark. These were measured by myself, courtesy of P. Bennike of the Panum Institute, Copenhagen. Earlier versions of CRANID have never had a European/Mediterranean sample with considerable time depth.

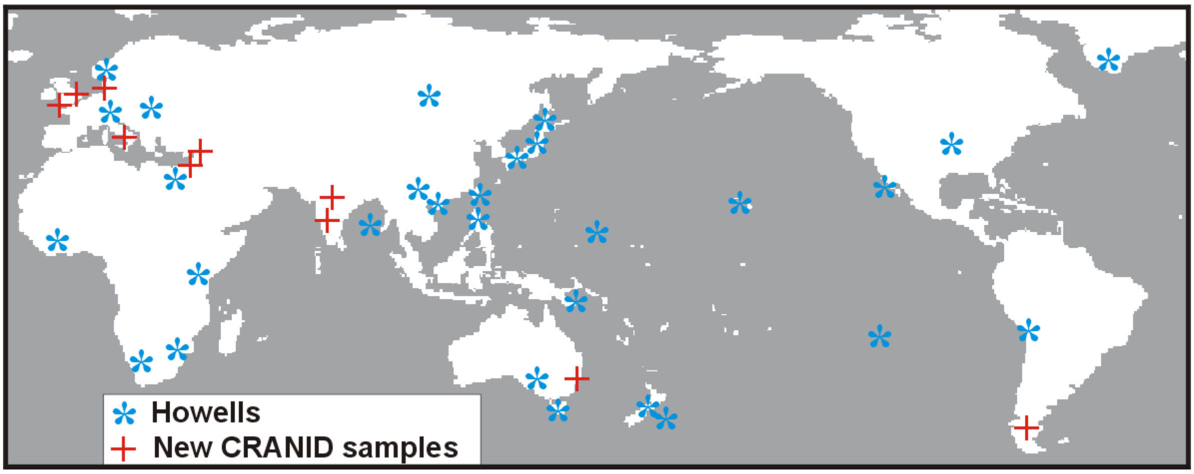


Figure 3. The geographical origins of samples in the CRANID6 database.

You will note that males and females are included in the same database. I see no justification for separating the sexes in a discriminant analysis.[29] Male and female of a single geographical group are simply treated as two independent samples. Analysis by LDA can thus be cautiously used as a sexing device in addition to its primary role of suggesting ancestry.

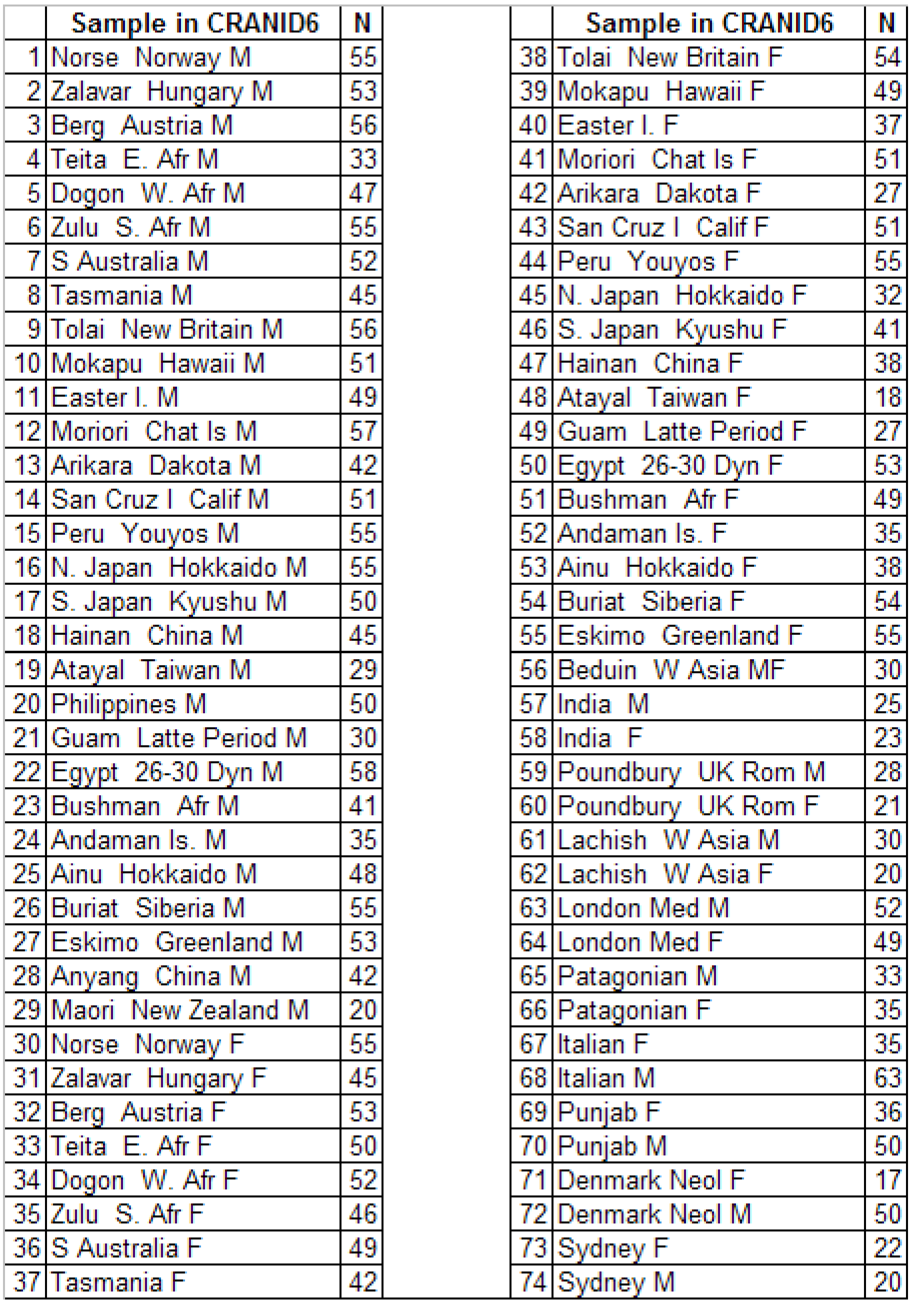
Some users have asked for details of the number of crania in each sample. It is important to realise that CRANID's version of linear discriminant analysis is not influenced by the number of cases in a sample. All samples are treated as though equal in number of individuals when probabilities of sample membership are computed. Nevertheless I am including sample sizes for the sake of general interest and because larger samples are more likely to be representative of the population from which they are drawn.

---

[29] I suspect that separating females and males in discriminant analysis had to do with the small amount of memory in early computers.

LANDMARKS AND MEASUREMENTS

Definitions of all the landmarks and measurements used in Cr6bInd.EXE are taken from Howells, W.W. 1973. *Cranial Variation in Man: A Study by Multivariate Analysis of Patterns of Difference among Recent Human Populations.* Papers of the Peabody Museum 67. Peabody Museum, Harvard University, Cambridge, Mass.



	Sample in CRANID6	N		Sample in CRANID6	N
1	Norse Norway M	55	38	Tolai New Britain F	54
2	Zalavar Hungary M	53	39	Mokapu Hawaii F	49
3	Berg Austria M	56	40	Easter I. F	37
4	Teita E. Afr M	33	41	Moriori Chat Is F	51
5	Dogon W. Afr M	47	42	Arikara Dakota F	27
6	Zulu S. Afr M	55	43	San Cruz I Calif F	51
7	S Australia M	52	44	Peru Youyos F	55
8	Tasmania M	45	45	N. Japan Hokkaido F	32
9	Tolai New Britain M	56	46	S. Japan Kyushu F	41
10	Mokapu Hawaii M	51	47	Hainan China F	38
11	Easter I. M	49	48	Atayal Taiwan F	18
12	Moriori Chat Is M	57	49	Guam Latte Period F	27
13	Arikara Dakota M	42	50	Egypt 26-30 Dyn F	53
14	San Cruz I Calif M	51	51	Bushman Afr F	49
15	Peru Youyos M	55	52	Andaman Is. F	35
16	N. Japan Hokkaido M	55	53	Ainu Hokkaido F	38
17	S. Japan Kyushu M	50	54	Buriat Siberia F	54
18	Hainan China M	45	55	Eskimo Greenland F	55
19	Atayal Taiwan M	29	56	Beduin W Asia MF	30
20	Philippines M	50	57	India M	25
21	Guam Latte Period M	30	58	India F	23
22	Egypt 26-30 Dyn M	58	59	Poundbury UK Rom M	28
23	Bushman Afr M	41	60	Poundbury UK Rom F	21
24	Andaman Is. M	35	61	Lachish W Asia M	30
25	Ainu Hokkaido M	48	62	Lachish W Asia F	20
26	Buriat Siberia M	55	63	London Med M	52
27	Eskimo Greenland M	53	64	London Med F	49
28	Anyang China M	42	65	Patagonian M	33
29	Maori New Zealand M	20	66	Patagonian F	35
30	Norse Norway F	55	67	Italian F	35
31	Zalavar Hungary F	45	68	Italian M	63
32	Berg Austria F	53	69	Punjab F	36
33	Teita E. Afr F	50	70	Punjab M	50
34	Dogon W. Afr F	52	71	Denmark Neol F	17
35	Zulu S. Afr F	46	72	Denmark Neol M	50
36	S Australia F	49	73	Sydney F	22
37	Tasmania F	42	74	Sydney M	20

Table 12. Composition of CRANID6 database.

I provide in this manual a set of measuring instructions and illustrations.[30] I urge all users to note the following:

1. All landmarks and definitions in CRANID follow those published by W.W. Howells (1973) *Cranial Variation in Man*. This choice, from a set of competing definitional systems, allows CRANID to incorporate Howells's database.

2. Users unfamiliar with cranial landmarks and measurements will find it an advantage to read the original definitions, cautions and discussions in W.W. Howells. His relevant pages are 163-183.[31]

3. The limited definitions and photos in the CRANID manual are intended as a guide for users already familiar with craniometry. The definitions and the photos should be studied in conjunction with each other. Neither is a sufficient guide on its own.

Landmarks.

This section is about landmarks and origins for measurements. The section after deals with the measurements themselves. However it is worth noting in passing that measurements are of three kinds - those that use:

1. cranial landmarks, e.g. biasterionic breadth (ASB);

2. instrumentally defined positions, e.g. maximum cranial breadth (XCB);

3. both cranial landmarks and instrumentally defined positions, e.g. maximum cranial length (GOL).

The following list of landmarks is taken from W.W. Howells.[32] The locations are illustrated in one or more of the nine *Defining Illustrations* prepared for this manual. In the notes that follow, these illustrations are referred to as 'DI'.

---

[30] Much of what follows, on the presentation of definitions, was prepared in cooperation with Ambika Flavel for inclusion in the Standard Operating Procedures of Inforce.

[31] This book has been out of print for some years. However I have seen an OCR transcript posted at http://www.cleber.com.br/howells.html

[32] These notes occasionally refer to Buikstra and Ubelaker (1994). This citation is of *Standards for Data Collection from Human Skeletal Remains,* Arkansas Archeological Survey Research Series No. 44.

Asterion as DI 2, 3, 6.

"The common meeting point of the temporal, parietal, and occipital bones, on either side."

Auriculare au DI 2, 3, 6

Not treated as a landmark by Howells. As Buikstra and Ubelaker (1994) state, it is defined as "a point on the lateral aspect of the root of the zygomatic process at the deepest incurvature, wherever it may be."

Basion ba DI 3, 4.

"On the anterior border of the foramen magnum, in the midline, at the position pointed to by the apex of the triangular surface at the base of either condyle, i.e., the average position from the crests bordering this area."

Bregma br DI 1, 2, 4, 9.

"The posterior border of the frontal bone in the median plane."

Dacryon dk DI 1, 8.

"The apex of the lacrimal fossa, as it impinges on the frontal bone.

Ectoconchion ek DI 1, 8.

"The intersection of the most anterior surface of the lateral border of the orbit and a line bisecting the orbit along its long axis.

Frontomalare anterior fm:a DI 1, 7.

"The most anterior point on the fronto-malar suture. It may be found with the side of a pencil lead held in the transverse plane."

Note: in my experience this landmark is the one most commonly misunderstood. Being the most anterior point on the suture, it differs from an alternative, and (from the point of view of the Howells's database) erroneous landmark, which is taken as the external point on the suture. If the external, as opposed to the anterior, points are used, then there will be errors in the measurements FMB and NAS - sometimes quite serious errors.

fm:a is therefore medial to the fmt of Buikstra and Ubelaker. Their fmt is not compatible with the Howells's database.

Ectoconchion ek DI 1, 8.

"The intersection of the most anterior surface of the lateral border of the orbit and a line bisecting the orbit along its long axis."

Note that in the illustration of Buikstra and Ubelaker (1994)the ectoconchion is wrongly placed, though the verbal definition is correct. In their illustrations they have confused the ectoconchion with the frontomalare orbitale

Glabella [a region] g DI 2, 4.

Not treated as a landmark by Howells. It is the most anterior point on the frontal bone, in the midline and above the nasion.

Jugalia ju DI 1, 2, 5.

"The deepest points in the curvature between the frontal and temporal processes of the malars."

Lambda la DI 2, 9.

"The apex of the occipital bone at its junction with the parietals, in the midline.

Nasion na DI 1, 2, 4, 7, 9.

"The intersection of the fronto-nsasal suture and the median plane."

Opisthion os DI 3, 9.

"The inferior edge of the posterior border of the foramen magnum in the midline."

Prosthion pr DI 1, 2, 3, 4, 7.

"The most anteriorly prominent point, in the midline, on the alveolar border, above the septum between the central incisors."

Subspinale ss DI 1, 2.

"The deepest point seen in the profile below the anterior nasal spine."

Zygomaxillare anterior zm:a DI 5.

"The intersection of the zygomaxillary suture and the limit of the attachment of the masseter muscle, on the facial surface."

Measurements

This section deals with the required 29 CRANID measurements, all of which are derived from Howells.

Users familiar with Buikstra and Ubelaker (1994) will find that only some of their measurements are used in CRANID (via Howells, of course). Some measurements will be unfamiliar to those familiar with Buikstra and Ubelaker.

Note in particular that there is a lack of congruence between Buikstra and Ubelaker's fmt-fmt and Howells's FMB.

The following abbreviated definitions for all 29 of these CRANID variables are taken from pages 5-7 in Howells, W. W., 1989. *Skull Shapes and the Map: Craniometric Analyses in the Dispersion of Modern Homo*. Peabody Museum of Archaeology and Ethnology, Vol. 79. Harvard University Press: Cambridge. The same appear in his work of 1973.

DI (Defining Illustration) refers to the illustrations that follow in this guide.[33]

1. **GOL Glabello-occipital length, DI 4**
   "Greatest length, from the glabellar region, in the median sagittal plane."

2. **NOL Nasio-occipital length, DI 4**
   "Greatest cranial length in the median sagittal plane, measured from nasion."

3. **BNL Basion-nasion length, DI 4**
   "Direct length between basion and nasion."

4. **BBH Basion-bregma height, DI 4**
   "Distance from basion to bregma, as defined."

5. **XCB Maximum cranial breadth, DI 5**
   "The maximum cranial breadth perpendicular to the median sagittal plane (above the supramastoid crests)."

6. **XFB Maximum frontal breadth, DI 5**
   "The maximum breadth at the coronal suture, perpendicular to the medial plane."

---

[33] I thank Amanda Reddick for the photos that I have annotated - DIs 1-3 and 5-8. DIs 4 and 9 are adapted from a photo posted by Tim McCormack at http://commons.wikimedia.org/wiki/File:Skull_-_midsaggital_section_P.2005.jpg.

7. **AUB Biauricular breadth, DI 6**
"The least exterior breadth across the roots of the zygomatic processes, wherever found."

8. **ASB Biasterionic breadth, DI 6**
"Direct measurement from one asterion to the other."

9. **BPL Basion-prosthion length, DI 4**
"The facial length from basion to prosthion, as defined."

10. **NPH Nasion-prosthion height, DI 7**
"Upper facial height from nasion to prosthion, as defined."

11. **NLH Nasal height, DI 7**
"The average height from nasion to the lowest point on the border of the nasal aperture on either side."

12. **OBH Orbit height, left, DI 8**
"The height between the upper and lower borders of the left orbit, perpendicular to the long axis of the orbit and bisecting it."

13. **OBB Orbit breadth, left, DI 8**
"Breadth from ectoconchion to dacryon, as defined, approximating the longitudinal axis which bisects the orbit into equal upper and lower parts."

14. **JUB Bijugal breadth, DI 5**
"The external breadth across the malars at the jugalia, i.e., at the deepest points in the curvature between the frontal and temporal process of the malars."

15. **NLB Nasal breadth, DI 7**
"The distance between the anterior edges of the nasal aperture at its widest extent."

16. **MAB Palate breadth, external, DI 6**
"The greatest breadth across the alveolar borders, wherever found, perpendicular to the median plane."

17. **ZMB Bimaxillary breadth, DI 5**
[Be careful with this one.] "The breadth across the maxillae, from one zygomaxillare [anterior] to the other."

18. SSS Zygomaxillary subtense
"The projection or subtense from subspinale to the bimaxillary width [ZMB]."

19. FMB Bifrontal breadth, DI 7
[Be careful with this one.] The breadth across the frontal bone between frontomalare anterior on each side, i.e., the most anterior point on the fronto-malar suture. [*Buikstra & Ubelaker (1994) measure between the most laterally positioned points on the fronto-malar suture; this distance is usually some 6 mm longer than Howells's FMB.*]

20. NAS Nasio-frontal subtense
"The subtense from nasion to the bifrontal breadth."

21. EKB Biorbital breadth, DI 8
"The breadth across the orbits from ectoconchion to ectoconchion." *The position of the ectoconchion, and thus their Biorbital Breadth, is incorrectly placed in the illustration of Buikstra and Ubelaker (1994).*

22. DKB Interorbital breadth, DI 8
"The breadth across the nasal space from dacryon to dacryon."

23. WMH Cheek height, DI 7
"The minimum distance, in any direction, from the lower border of the orbit to the lower margin of the maxilla, mesial to the masseter attachment, on the left side."

24. FRC Nasion-bregma chord (Frontal chord), DI 9
"The frontal chord, or direct distance from nasion to bregma, taken in the midplane and at the external surface."

25. FRS Nasion-bregma subtense (Frontal subtense), DI 9
"The maximum subtense, at the highest point on the convexity of the frontal bone in the midplane, to the nasion-bregma chord."

26. PAC Bregma-lambda chord (Parietal chord), DI 9
"The external parietal chord, or direct distance from bregma to lambda, taken in the midplane and at the external surface."

27. PAS Bregma-lambda subtense (Parietal subtense), DI 9

"The maximum subtense, at the highest point on the convexity of the parietal bones in the midplane, to the bregma-lambda chord."

28. OCC Lambda-opisthion chord (Occipital chord), DI 9

"The external occipital chord, or direct distance from lambda to opisthion, taken in the midplane and at the external surface."

29. OCS Lambda-opisthion subtense (Occipital subtense), DI 9

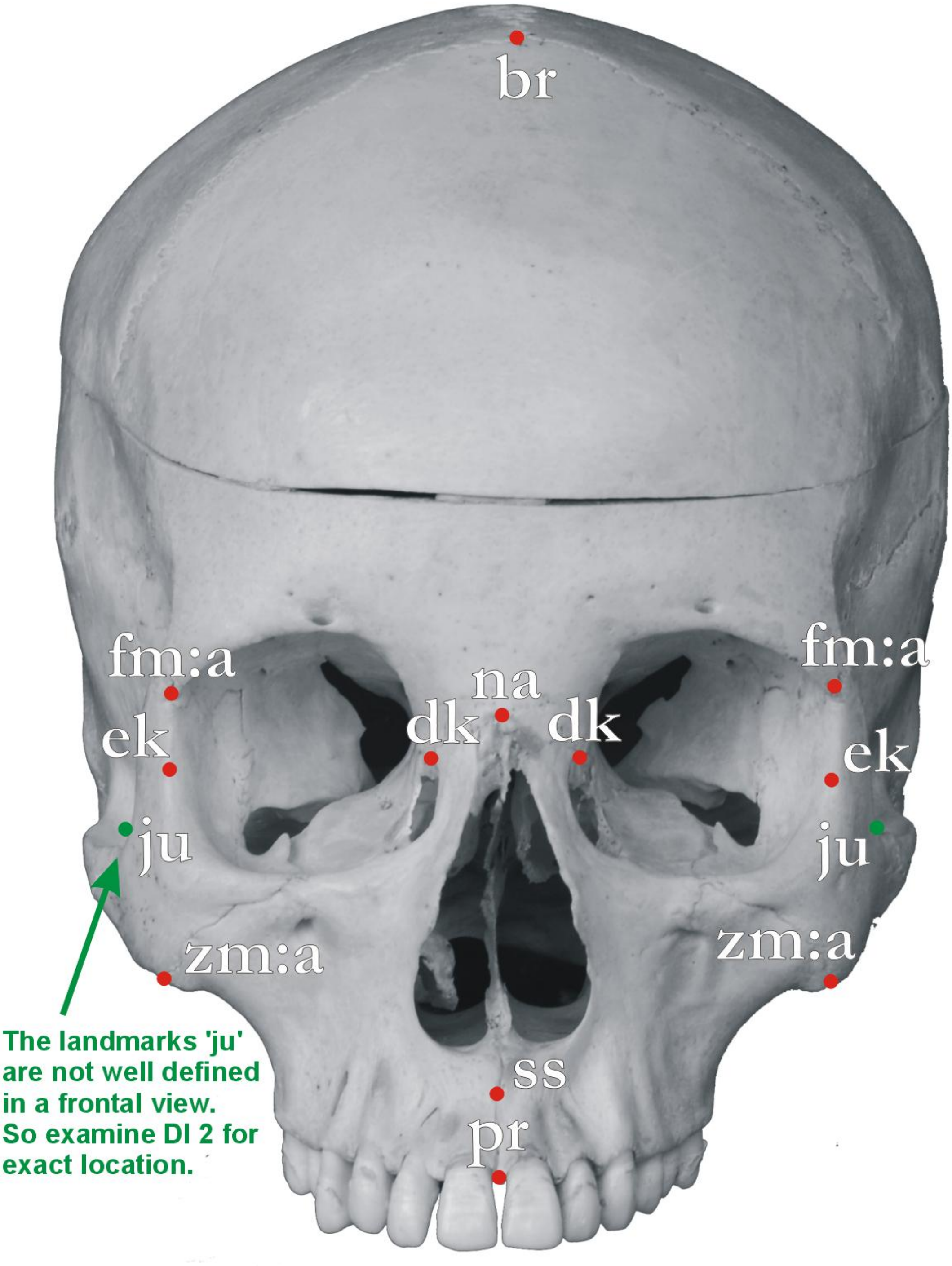
"The maximum subtense, at the most prominent point on the basic contour of the occipital bone in the midplane."

Note on Occipital subtense (OCS)

Howells's words 'basic contour' are related to his fuller description in "Cranial Variation in Man" page 182.

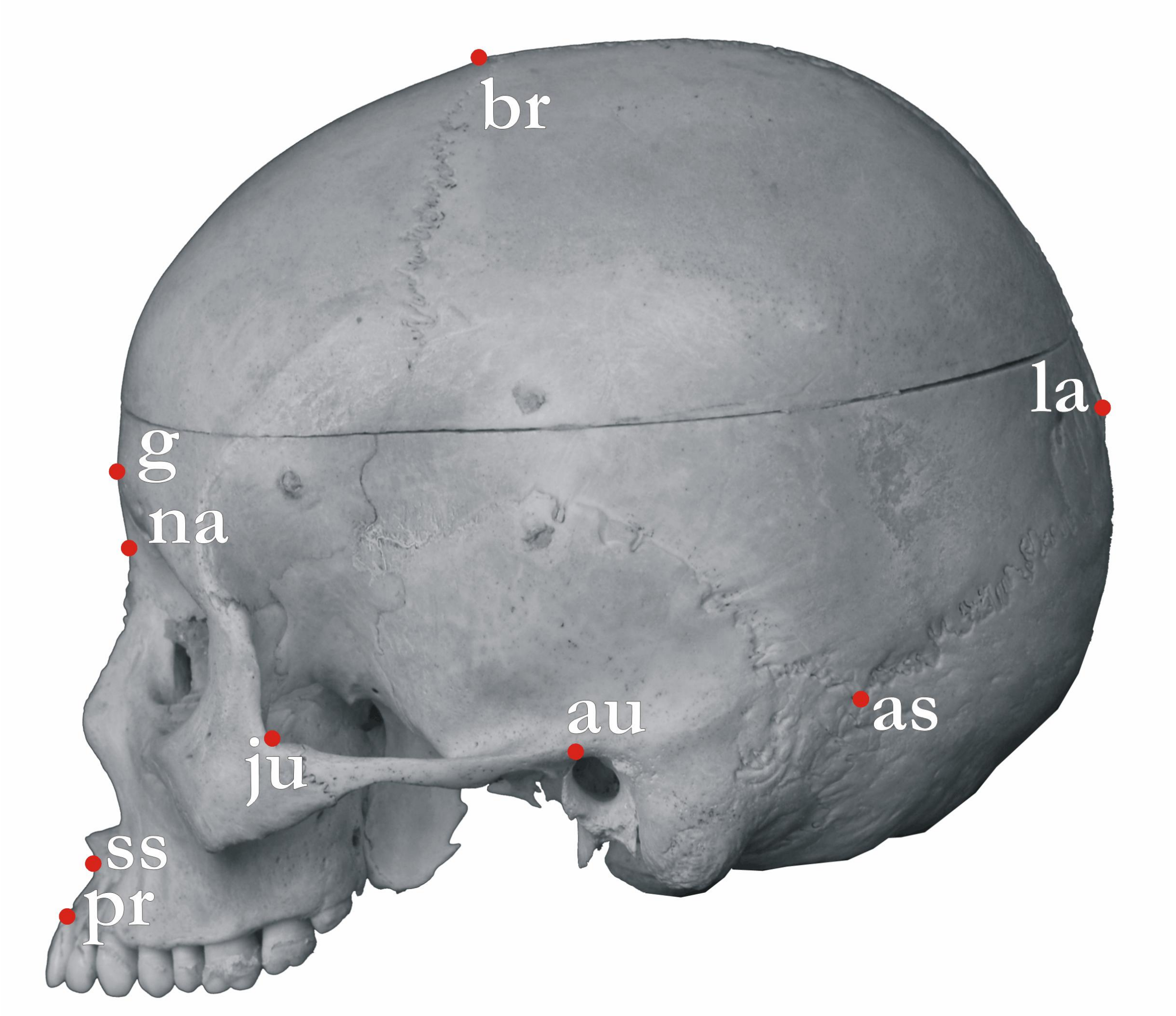
A note in the fuller description reads:

> "If there is a moderately well-developed but rounded nuchal crest forming part of the general contour and profile of the bone, this should be included if the subtense falls here. If, however, a central elevation or the inion stands out prominently, this should be discounted by placing the point in the notch directly above the midline downward curve of inion and crest. This is most likely to be at the level of the actual apex of curvature of the bone itself, and level with the highest point of the curved lines or the torus on either side. In any case the proper point for the subtense - the apex of curvature - is likely to be above the inion, either in this depression or higher."
>
> "This is a deliberate but perhaps not entirely satisfactory step to keep a local development here from giving a highly dislocated rendering of the actual angulation of the occipital bone. Often, however, the maximum subtense reading lies well above even a prominent inion, and the problem arises in only a few populations anyhow."

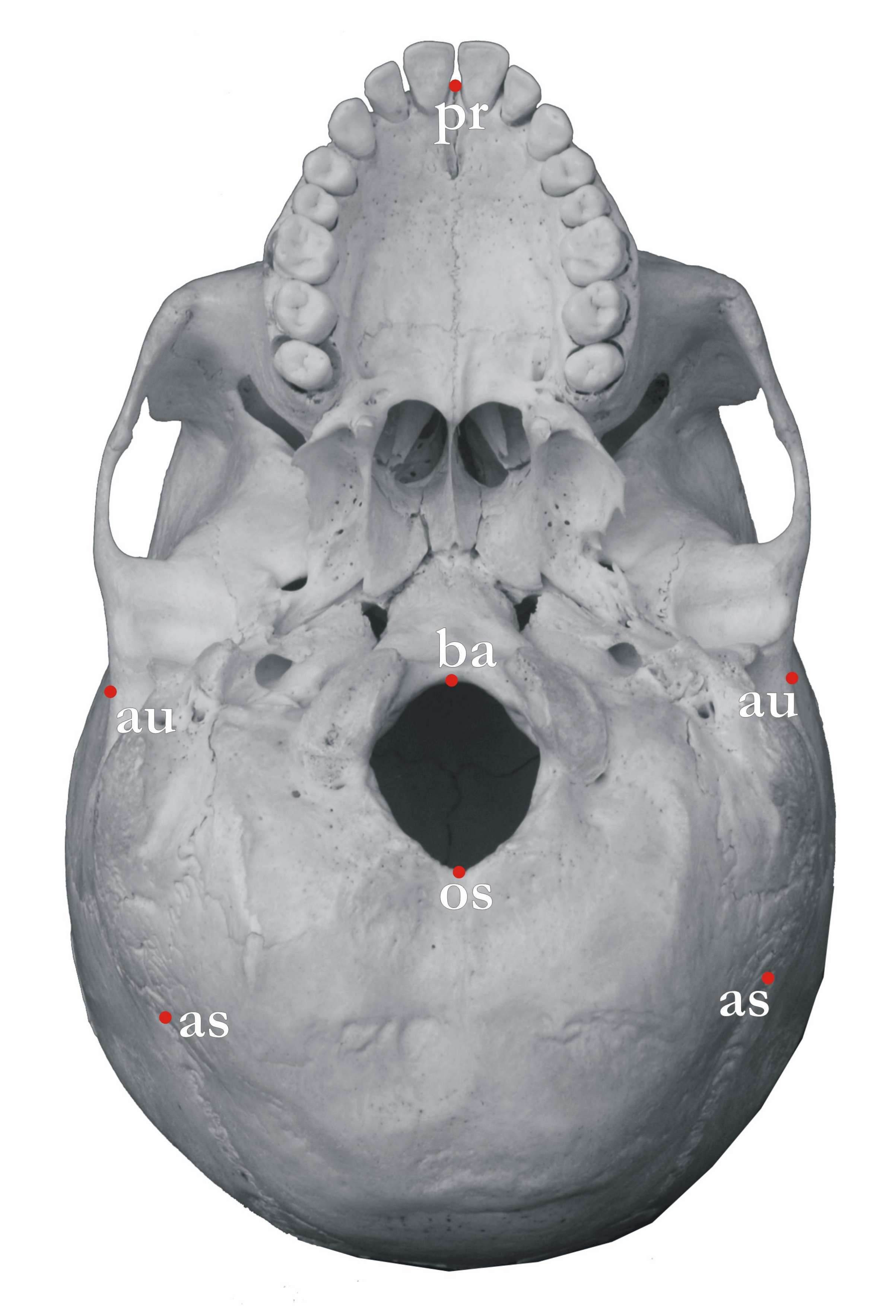


The landmarks 'ju' are not well defined in a frontal view. So examine DI 2 for exact location.

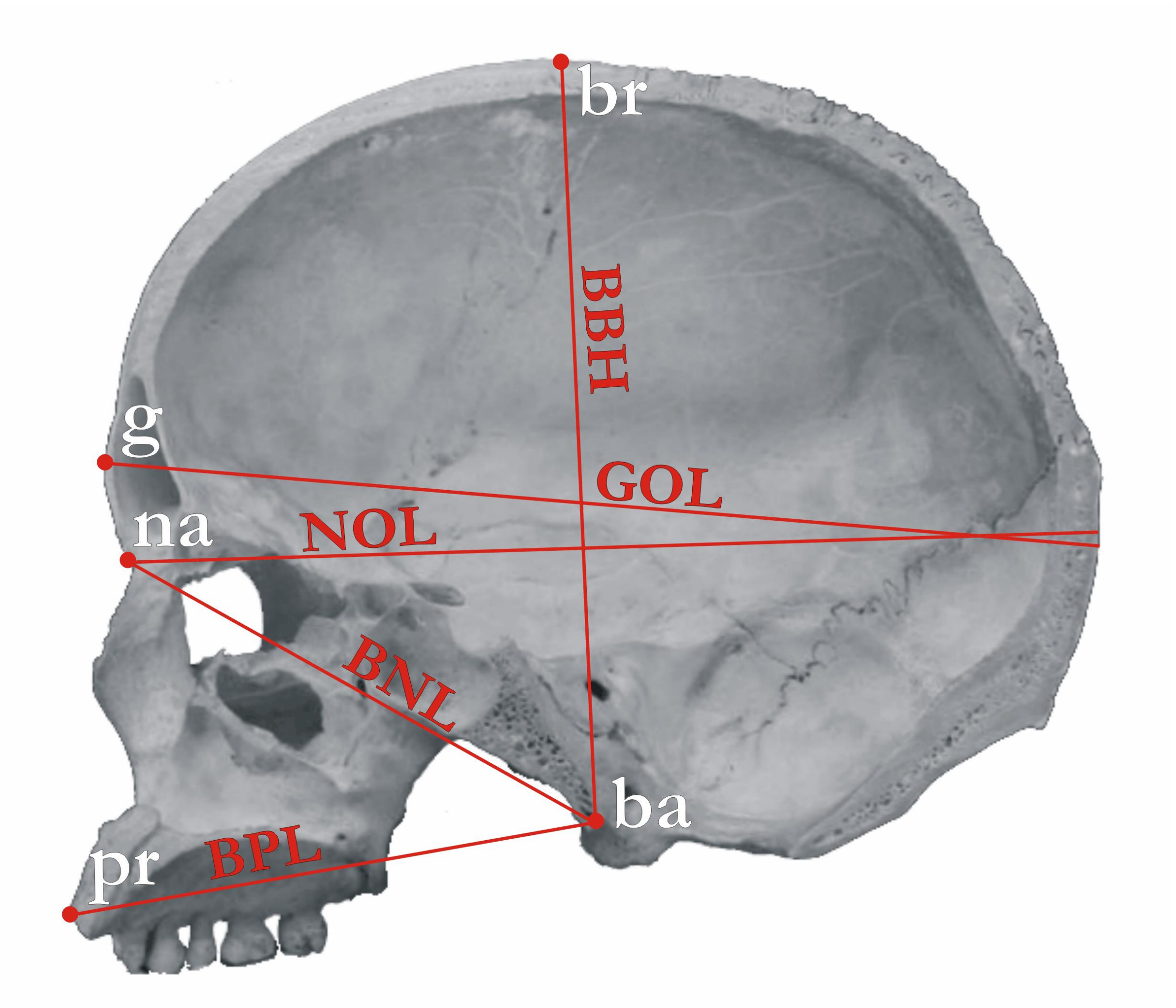
Defining illustration 1.



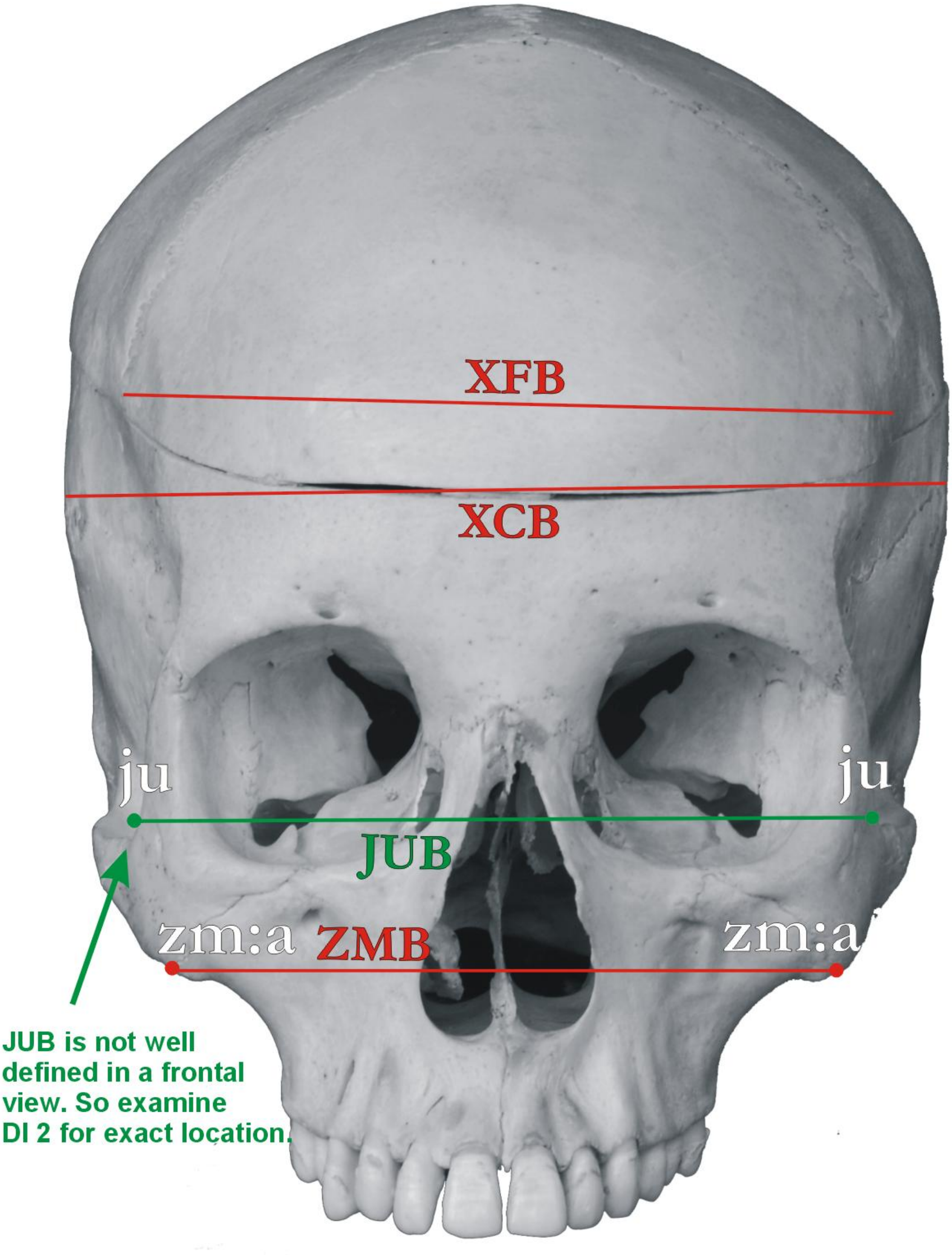
Defining illustration 2.



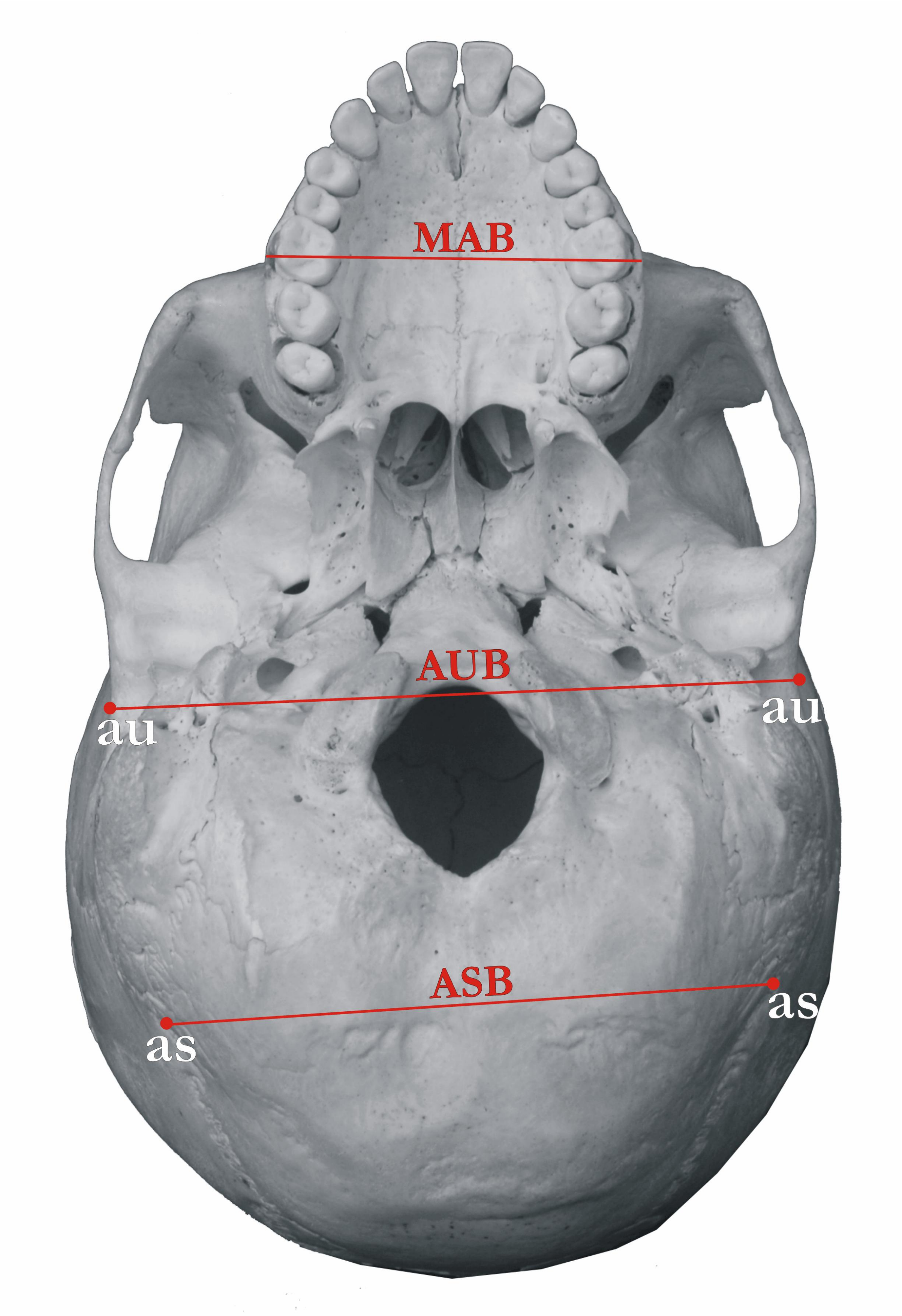
Defining illustration 3.



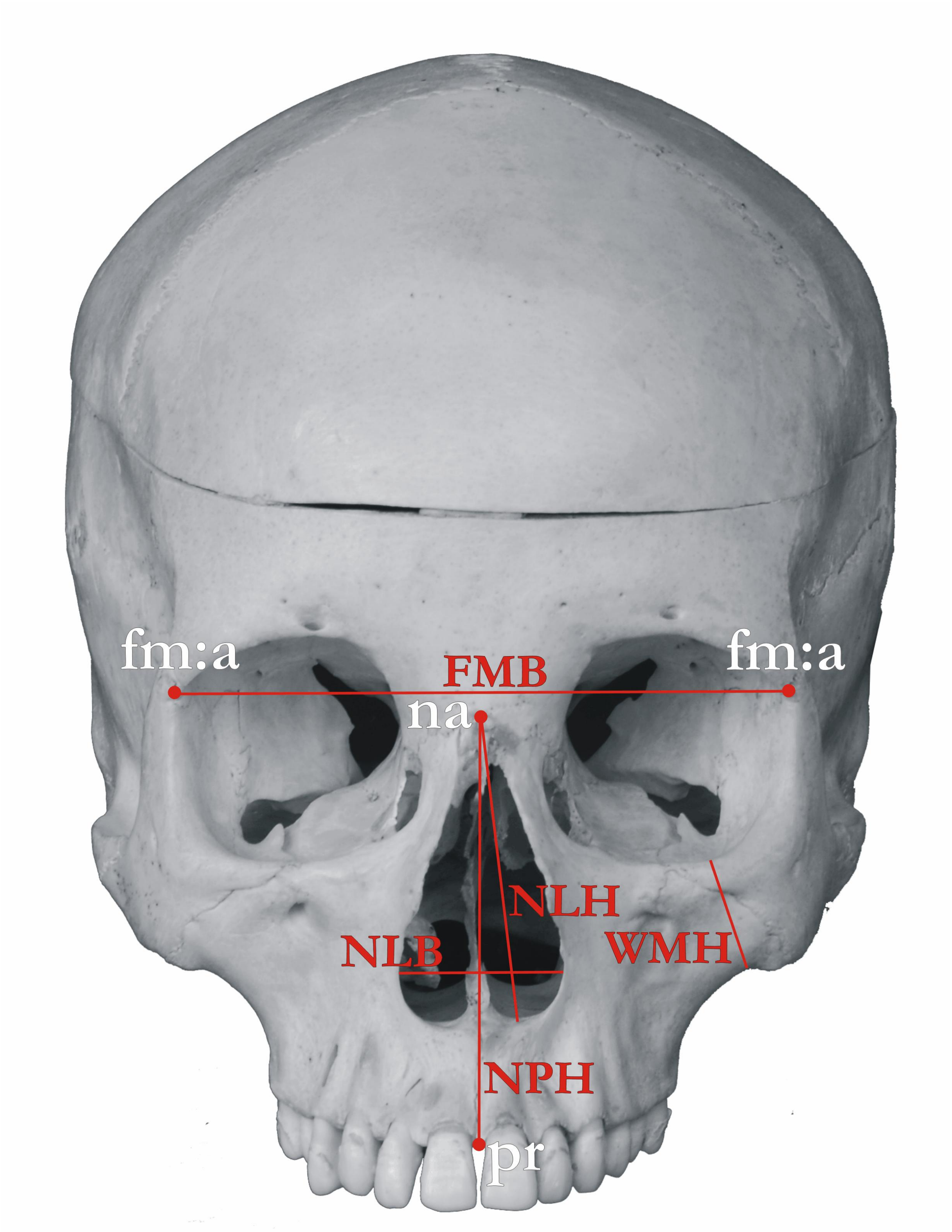
Defining illustration 4.



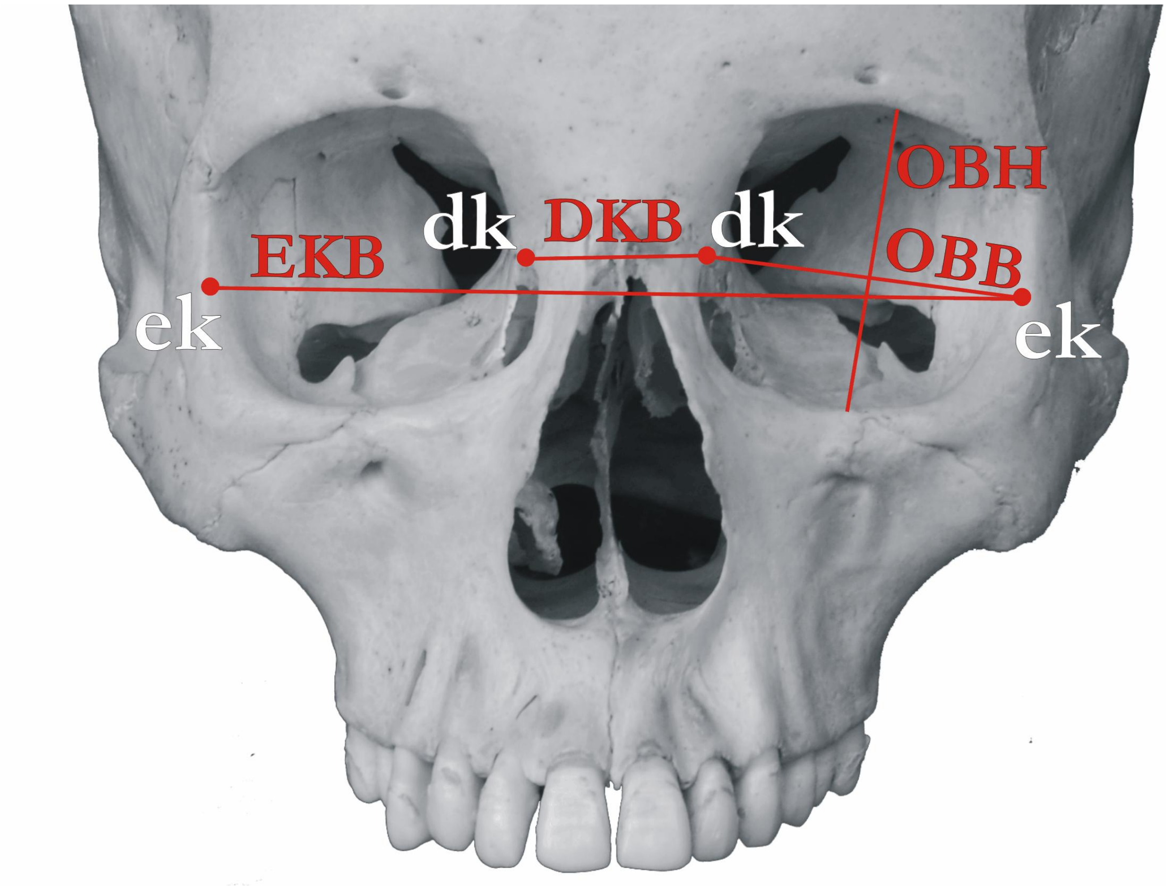
Defining illustration 5.



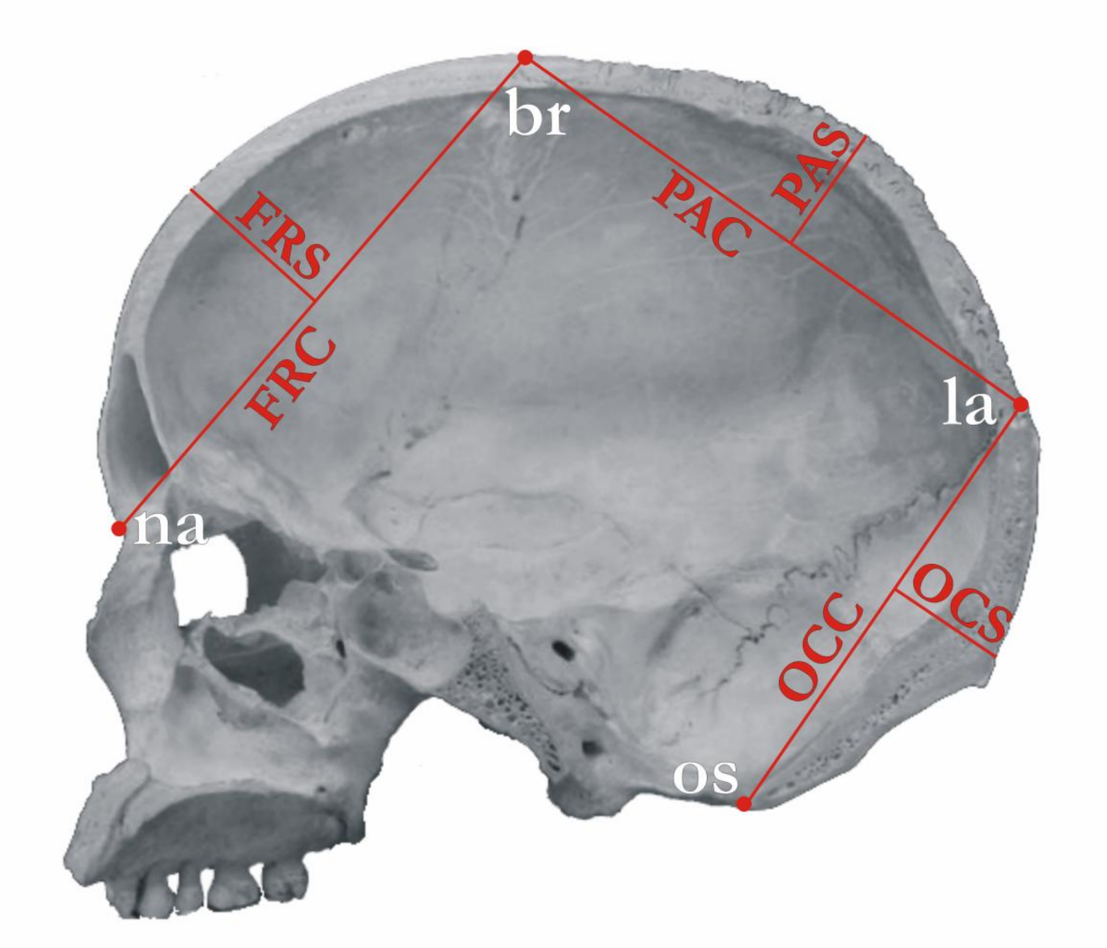
Defining illustration 6.



Defining illustration 7.



Defining illustration 8.



Defining illustration 9.

SUBSTITUTES FOR SPECIAL CALIPERS

Three types of caliper are used to obtain the 29 cranial measurements for CRANID:

sliding

spreading

coordinate

A set of these anthropometric devices will cost around $US1,500 and delivery may be some weeks away after ordering.

So the question arises whether some less expensive and more rapidly acquired substitutes are available, which will measure accurately to the nearest 1 mm that is acceptable for entry into CRANID.

Here are some suggested substitutes:

SLIDING CALIPERS

Your local hardware store is likely to carry plastic sliding calipers. These have the advantage of not damaging the bone to which they are applied; for this reason, plastic calipers are required by some osteological institutions.

Buy yourself several, since the points will wear out on abrasive bones. Buy the same model (see below, under coordinate calipers for the reason).

You may care to verify the accuracy of these plastic calipers by checking them against an engineer’s rule.

SPREADING CALIPERS

The standard anthropometric spreading calipers have a pivoted rule attached. I have never seen these in hardware stores.

Recently, digital spreading calipers have appeared on the market at astonishingly low prices. Search on the web for *8″ outside calipers*. Do not expect to pay more than US$40. Do not buy 6″ calipers, since they will not cope with cranial length (GOL).

Figure 4. Digital spreading calipers.

Finally, there are uncalibrated spreading calipers – (or *outside calipers*). These are to be found in large hardware stores.

The greatest diameter you will have to measure is GOL, which will be more than catered for by standard sized spreading calipers with a spread of 250 mm.

Try to get a set whose arms can be locked by a central nut, when the points are in place on the cranial landmarks.

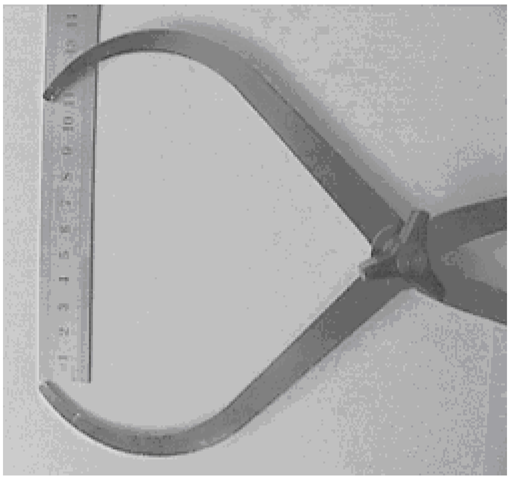


Figure 5. Improvised spreading calipers.

You will also need to get yourself a steel engineer’s rule. The measured diameters are read by moving the locked spreading calipers to this rule, as shown in the following picture.

COORDINATE CALIPERS - PHYSICAL SUBSTITUTION
Craniometric coordinate calipers are particularly damaging to bone, with their sharp steel points and general clumsiness of use. My feeling is that they should not be used. In other words, the plastic substitute described here should be required by curators of crania.

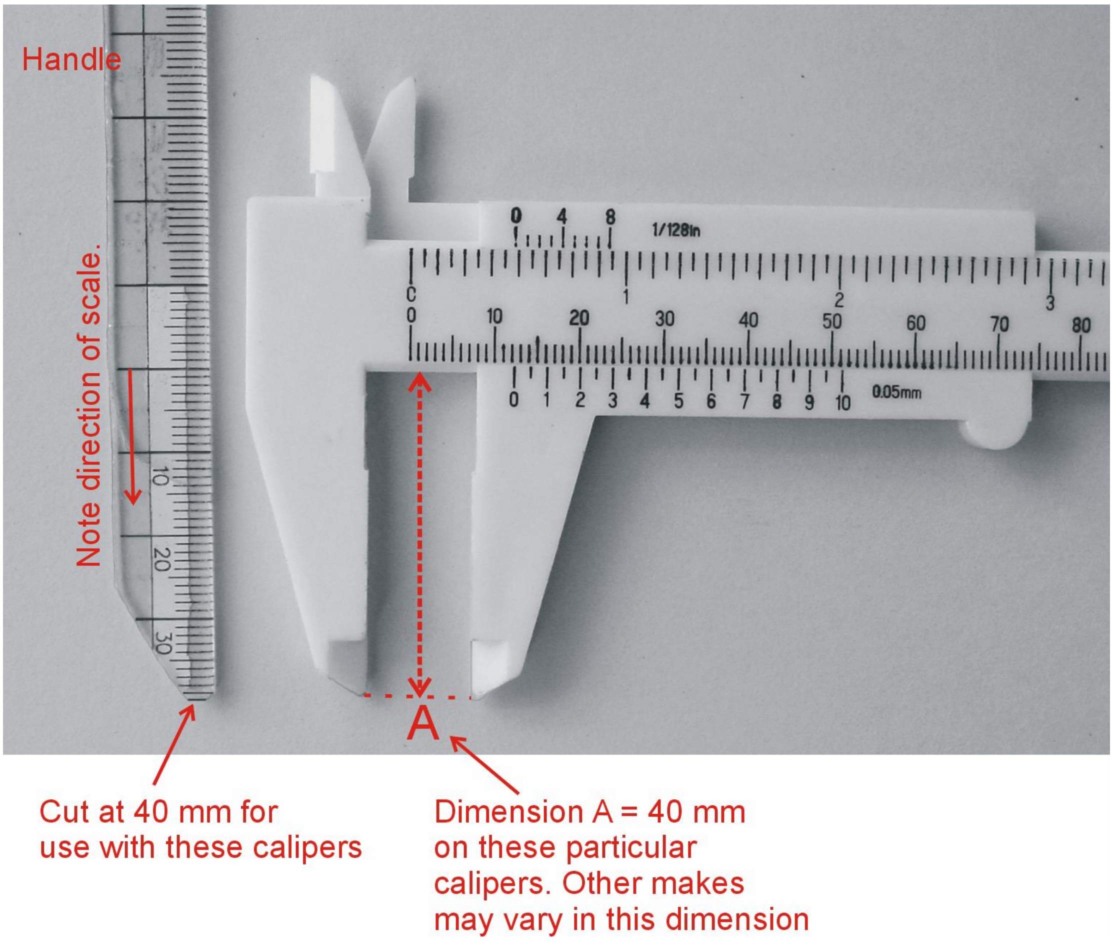


Figure 6. Preparing the ruler for making plastic coordinate calipers. Note that the scale on the ruler part runs backwards from the tip.

Substitution here is a little complicated. You will need:

plastic sliding calipers (as described above) plastic set square or rule with a scaled edge

hacksaw

file

With Figure 6 as your guide, first work out precisely how far the jaws on the sliding calipers project from the bar, i.e. the 'A' dimension in the picture.

I have found that 40 mm is a common projection with plastic sliding calipers, but you must check each brand and adjust the length of the subsequent cut accordingly.

With a hacksaw, cut off a length of the rule - retaining the zero end (as a handle) and cutting along the scale at a point that is just beyond the amount that the jaws project. If you do not have a narrow rule you will need also to cut with a hacksaw along the length of the rule, to produce a gadget with a form that is similar to the result in the above illustration.

Finish this gadget with a file, shaping it so that the tip is exactly the length of the projection, and with an angled end.

You may find it convenient to attach some sort of handle at the zero end of the cut out piece.

The calipers are used as shown in the following illustration, for the measurement of subtenses. The example shown is the measuring of FRS - the frontal subtense.

> With the sliding calipers in place on the nasion and bregma, slide the gadget along the bar of the sliding calipers until it is in contact with the place of maximum projection of the frontal bone.
>
> The gadget must, of course, be perpendicular to the bar of the sliding calipers.
>
> Read the subtense off the gadget where it intersects with the bar of the sliding calipers (25 mm in the illustration).

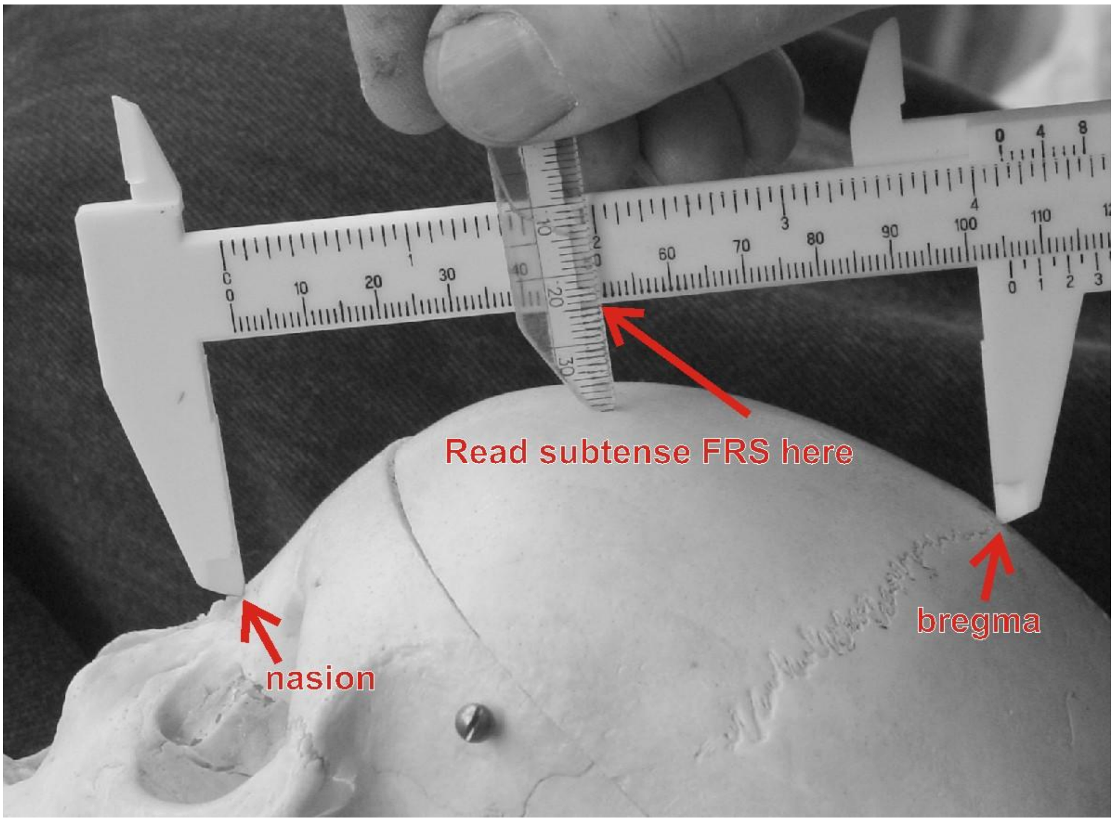


Figure 7. Using the plastic coordinate calipers.

In this example Frontal Chord (FRC) reads 101 mm and Frontal Subtense (FRS) 25 mm.

The direction of the scale in coordinate calipers can seem counterintuitive, unless you realise that the more the curve of the bone projects toward the sliding calipers the greater the value that will be read on the rule. Think of it this way - if there were no upward projection of the cranium beyond the straight line connecting the two points of the sliding calipers then the value of the subtense would be zero. The more the cranium projects upwards, the greater the value that will be read on the rule.

The points on the jaws of the plastic sliding calipers will become blunt through wear, particularly on bones that are impregnated with abrasive soil. So the gadget becomes increasingly unreliable because the projection becomes less. The only solution is to replace the sliding calipers with another set of the same model (or same projection). With careful use, a plastic sliding caliper can be used on some 30 crania before it wears out. The rule part will not wear out.

COORDINATE CALIPERS - SUBSTITUTION BY TRIGONOMETRY

Dave Bulbeck has kindly pointed out to me that trigonometry is available for computing a subtense.

The solution is not included in simple trigonometry teaching, but I have included a utility program SUBTENSE.EXE for computing the subtense from the entry of the chord (A) and the two sides of the triangle between the ends of the chord and the apex (B & C).

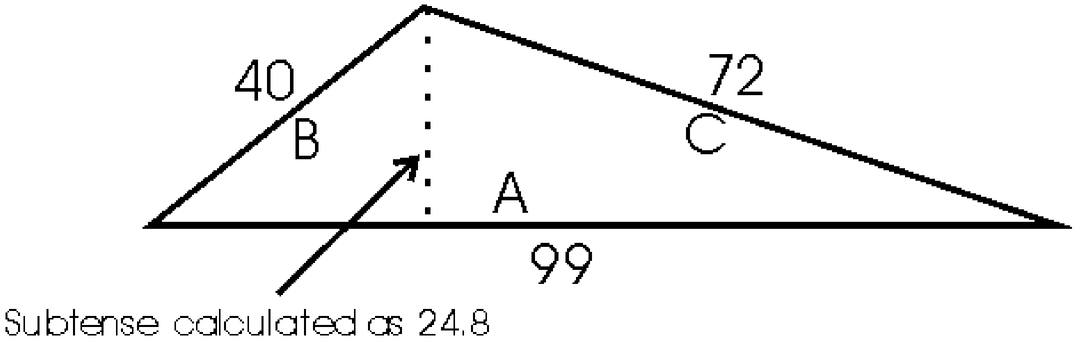
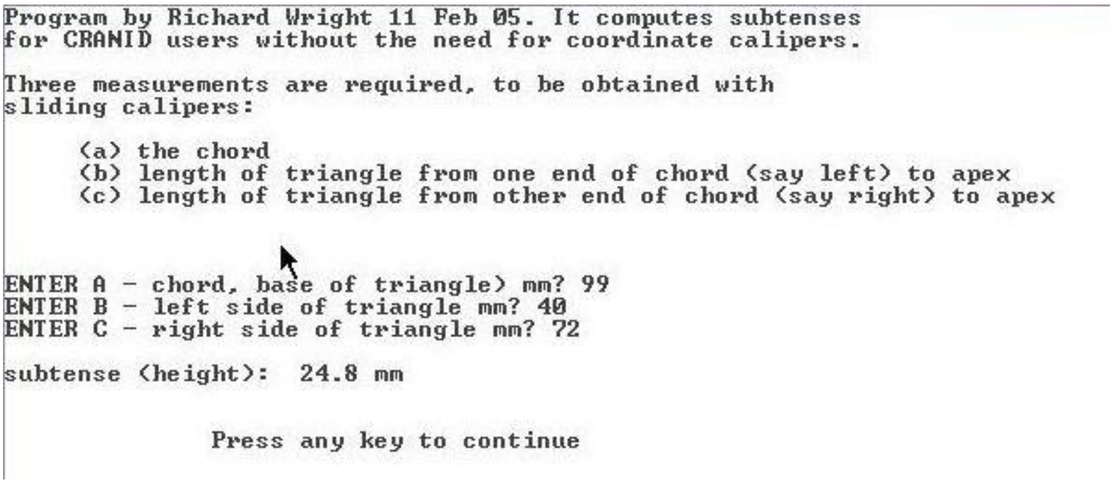


Figure 8. Trigonometric substitute for coordinate calipers.

Run SUBTENSE.EXE and enter the values as requested.

The SUBTENSE program screen for solving the above triangle shows the entry of the values for A, B & C and the result:



```
Program by Richard Wright 11 Feb 05. It computes subtenses
for CRANID users without the need for coordinate calipers.

Three measurements are required, to be obtained with
sliding calipers:

     (a) the chord
     (b) length of triangle from one end of chord (say left) to apex
     (c) length of triangle from other end of chord (say right) to apex


ENTER A - chord, base of triangle) mm? 99
ENTER B - left side of triangle mm? 40
ENTER C - right side of triangle mm? 72

subtense (height):  24.8 mm


                Press any key to continue
```

Table 14. Input and output of program SUBTENSE.EXE, which Substitutes for coordinate calipers.

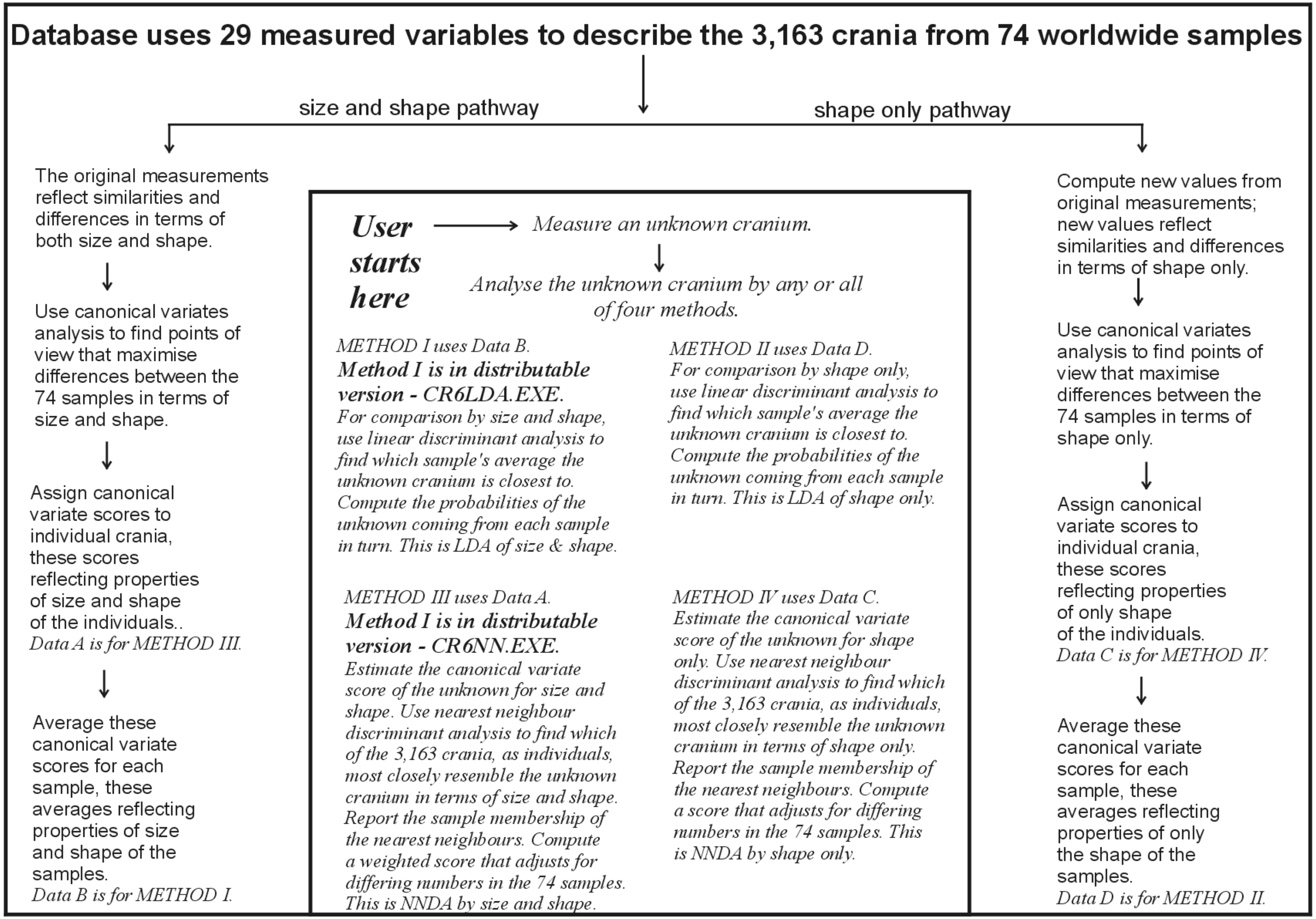
NOTES ON THE FULL CRANID ANALYSIS

Use of the CRANID distributable package has already been described. These notes now look beyond the distributable package to the entire CRANID approach and the niche that the distributable package occupies in that approach.

These notes cast their net wider than the distributable program, but they contain advice that will be useful in the interpretation of results obtained by it.

There follows a chart (Figure 9) that summarises the meaning of the various databases that make up a full CRANID analysis. The user's position is placed in a box that symbolically lies within a larger box that contains the various databases.

There are two distinct types of data and two distinct types of analytical method. These allow four possible analyses, which are now described.



Database uses 29 measured variables to describe the 3,163 crania from 74 worldwide samples

size and shape pathway

The original measurements reflect similarities and differences in terms of both size and shape.

Use canonical variates analysis to find points of view that maximise differences between the 74 samples in terms of size and shape.

Assign canonical variate scores to individual crania, these scores reflecting properties of size and shape of the individuals..
Data A is for METHOD III.

Average these canonical variate scores for each sample, these averages reflecting properties of size and shape of the samples.
Data B is for METHOD I.

User starts here → *Measure an unknown cranium.*

Analyse the unknown cranium by any or all of four methods.

METHOD I uses Data B.
Method I is in distributable version - CR6LDA.EXE.
For comparison by size and shape, use linear discriminant analysis to find which sample's average the unknown cranium is closest to. Compute the probabilities of the unknown coming from each sample in turn. This is LDA of size & shape.

METHOD II uses Data D.
For comparison by shape only, use linear discriminant analysis to find which sample's average the unknown cranium is closest to. Compute the probabilities of the unknown coming from each sample in turn. This is LDA of shape only.

METHOD III uses Data A.
Method I is in distributable version - CR6NN.EXE.
Estimate the canonical variate score of the unknown for size and shape. Use nearest neighbour discriminant analysis to find which of the 3,163 crania, as individuals, most closely resemble the unknown cranium in terms of size and shape. Report the sample membership of the nearest neighbours. Compute a weighted score that adjusts for differing numbers in the 74 samples. This is NNDA by size and shape.

METHOD IV uses Data C.
Estimate the canonical variate score of the unknown for shape only. Use nearest neighbour discriminant analysis to find which of the 3,163 crania, as individuals, most closely resemble the unknown cranium in terms of shape only. Report the sample membership of the nearest neighbours. Compute a score that adjusts for differing numbers in the 74 samples. This is NNDA by shape only.

shape only pathway

Compute new values from original measurements; new values reflect similarities and differences in terms of shape only.

Use canonical variates analysis to find points of view that maximise differences between the 74 samples in terms of shape only.

Assign canonical variate scores to individual crania, these scores reflecting properties of only shape of the individuals.
Data C is for METHOD IV.

Average these canonical variate scores for each sample, these averages reflecting properties of only the shape of the samples.
Data D is for METHOD II.

Figure 9. Chart showing relation between data and analytical method in the full CRANID approach, of which the distributable CR6Ind.EXE package is a part.

DETAILED ANALYSIS OF AN UNKNOWN CRANIUM BY CRANID.

The methods of CRANID assume that the unknown cranium to be analysed is within the range of variation of modern *Homo sapiens*.

The user must remember that the analyses are based on similarities and dissimilarities in cranial morphology. Ancestry is an explanation that is inferred from these morphological results.

Methods I & III are the methods used in the distributable version of CRANID. Methods II – IV are available as data and software on my computer. Methods I – IV comprise the full CRANID analysis.

The extra analyses are available by e-mailing me (see section below **Charges for Additional Analyses**).

METHOD I
Linear discriminant analysis by size and shape - LDA-SS. This is a method included in the distributable version of CRANID.

METHOD II
Linear discriminant analysis by shape only - LDA-Sh. Available on application.

METHOD III
Nearest neighbour discriminant analysis by size and shape - NNDA-SS. This is a method included in the distributable version of CRANID.

METHOD IV
Nearest neighbour discriminant analysis by shape only - NNDA-Sh. Available on application.

OUTLINE OF METHODS FOR A FULL ANALYSIS OF AN UNKNOWN CRANIUM BY CRANID

These methods are presented in increasing order of complexity of data preparation and computational time.

METHOD 1 Rapid screening of unknown.

The results using METHOD I will be adequate, provided the unknown’s ancestry is reasonably well catered for by the database.

This method takes account of both size and shape, so attribution of sex by METHOD I is more likely to be correct than by results of METHODS II and IV. This is because cranial difference between the sexes is partly a factor of absolute size.

You should look for signs that the unknown is properly catered for by the database. Signs of being *poorly* catered for will be seen in a low probability of membership of the closest sample, and a long succession of subsequent samples with probabilities greater than 0.05 and bearing little geographical relationship to each other.

Consider the following possible explanations.

A. above all, first consider the possibility of measurement or entry error. *ACTION - check the measurements and their entry.* Consider whether

> the unknown individual is erroneously measured (perhaps a defined measurement was not understood by the measurer)
>
> an error was made when entering a value into the CRANID form.

B. unknown does come from a geographical region that is included in the database, but is a cranium that is particularly small or large. *ACTION - check results with those of METHODS II and IV, which work on shape alone*.

C. unknown comes from a region that is not included in the database of 74 samples or is of mixed ancestry. *ACTION - compare the LDA results with the NNDA results of METHODS III and IV*.

D. unknown is the 'victim' of a badly behaved performance by LDA, remembering that LDA assumes the regularity of statistical distributions in the database. *ACTION - compare the LDA results with NNDA results of METHODS III and IV*.

METHOD II. Screening of unknown by eliminating absolute size differences and working on shape alone.

In LDA this method of using only shape produces a slightly inferior classificatory success rate than LDA on size and shape. It is nevertheless an important method for making sure that erratic results are not obtained because crania are particularly small or particularly large for their respective populations.

Size is eliminated, leaving us with only shape, by the simple procedure outlined by Manly (1994:88-89).[34] Because all the variables in CRANID are linear, each measurement on a cranium can be expressed as a ratio of the sum of all 29 variables measured on that cranium.[35]

In general, sex attribution by this method is inferior to results obtained by the analysis of size and shape used in METHODS I and III.

Again look for signs that the unknown is properly catered for by the database. Consider the same possible explanations outlined for METHOD I above, except that of unusual size.

METHODS III and IV. Detailed examination of distance of unknown cranium from individuals in the database - nearest neighbour discriminant analysis.

Unlike linear discriminant analysis, nearest neighbour analysis makes no assumptions about the distributions in the underlying data.

Also unlike discriminant analysis, nearest neighbour analysis does not categorise the unknown cranium in terms of the mean values of the 74 samples. In other words, nearest neighbour analysis is individualistic. It compares the unknown with each of the individuals in the database, not with their information summarised as average values of samples.

---

[34] Manly, B.F.J. (1994). *Multivariate Statistical Methods: a Primer.* Chapman & Hall: London.

[35] Do not do this to your unknown cranium and then put the re-expressed shape variables into the distributable package. The result will produce a nonsense. The procedure requires completely new discriminant functions to be computed from a database from which size has also been removed. Contact me for advice.

The unknown cranium is looked at in the space of the 29 dimensions of canonical variate scores of the crania in the database. Its distance to each of the 3,163 crania is measured and those individuals in the database with the shortest distances from the unknown (i.e. with greatest resemblance to the unknown) are recorded.

The search for nearest neighbours stops when the total number of hits reaches 56, which is a conventional stopping point derived from the square root of the total number of individuals.

The report lists the samples in terms of the 56 hits received, and then computes a weighted score that adjusts for differences in the numbers of crania in each sample - obviously a large sample can expect, by chance, to get more hits than a sample with a more meagre membership.

Though intuitively transparent, these methods of nearest neighbour analysis are highly computer intensive. Linear discriminant analysis is itself computationally difficult, but one might argue that it was developed as a less computer intensive substitute for nearest neighbour analysis. Only with the advent of fast processors has nearest neighbour analysis become possible. Originally statisticians modeled the classificatory method of linear discriminant analysis and in doing so had to assume the normality of the data and the equivalence of variance in the samples. These assumptions are not needed by nearest neighbour analysis.

Nearest neighbour analysis (whether METHOD III for size and shape or METHOD IV for shape only) is competent at showing overlap of individuals from separate samples in the database. It is therefore useful for suggesting mixed ancestry.

The principles of *size & shape* versus *shape only* results apply to nearest neighbour analysis as they do to linear discriminant analysis.

COMMENTS ON ANALYSIS OF INCOMPLETE CRANIA

As previously mentioned, the freely distributable package does not analyse incomplete crania, for which the 29 variables are not all available (or responsibly estimated). I am able to analyse such crania.

The question is often put to me - what is the minimum number of variables that is viable for a CRANID analysis.

The answer is that the number of variables is less important than the morphological coverage of the variables.

Let us consider some metric factors that we can consider in a cranium, and the 29 CRANID variables that define them.

Cranial length: GOL NOL

Cranial breadth: XCB, XFB, [AUB]

Cranial height: BBH

Facial projection: BPL

Facial height: NPH

Facial breadth: ZMB, FMB, JUB, EKB

Here we see that some factors are described by only one variable (e.g. cranial height) but other factors have more than one variable describing them (e.g. facial breadth).

Ideally all factors should be represented in the variable list of an incomplete cranium.

The further we wander away from having all factors represented the more fuzzy the results of classification are likely to be.

This is not to say that I would decline to analyse fragmentary crania that do not have all factors present. However in such cases we proceed at our own risk.

GRAPHICAL METHODS OF PRESENTATION

I have available the average canonical variate scores for the samples, pooled for sex. There are two graphical methods that can make use of these averaged scores. To use these graphical methods, the unknown individual's canonical variate scores must first be estimated from the discriminant function coefficients.

The following two graphical methods are used (see Figures 10 and 11 for examples of the template charts on to which the individual is placed after its canonical variate scores have been estimated). The figures are derived from shape data.

> (1) Figure 10. A scattergram, showing the samples and the unknown cranium placed among the samples. This scattergram is in the space of only the first two canonical variates and so is using only part of the identifying information available – 36.2% (LDA and NNDA use all the information).
>
> In my experience this graphical methods is suitable for displaying the similarities of unknown crania to averages - but only when LDA and NNDA have already produced a clear result. Where LDA and NNDA produce unclear results then this graphic method may be misleading, given that it uses only 36.2% of the information in the database. The visual simplicity is methodologically seductive. Despite this caution, I make a scattergram available on application.
>
> (2) Figure 11. A dendrogram, showing the samples classified by cluster analysis. This procedure uses all the canonical variates, i.e. 100% of the identifying information, and not only the first two canonical variates of the scattergram approach.

CHARGES FOR EXTRA ANALYSIS

I am happy to do the extra analyses discussed above. I am also able to analyse incomplete crania, where one or more of the 29 variables are missing.

Prices for extra analyses and services will be negotiated. These extra analyses include a written report, and interpretation of the computer output.

THE SCATTERGRAM

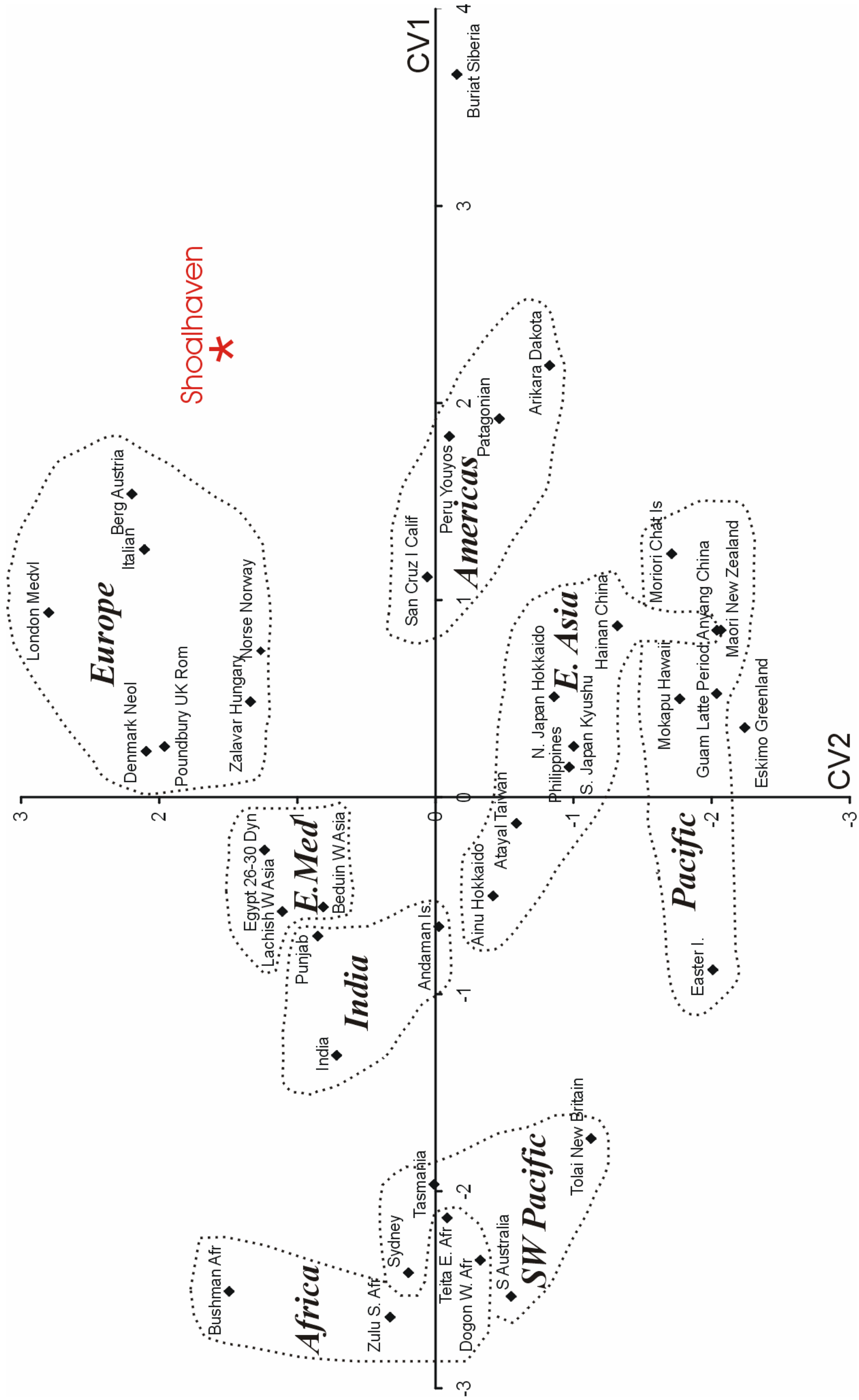


Figure 10 (previous page). Scattergram showing samples in the space of the first two canonical variates. The Shoalhaven cranium is placed on this scattergram, its canonical variate scores having been interpolated.

The scattergram of Figure 10, derived only from cranial morphology, bears some resemblance to a map of the world if examined with the CV2 axis horizontal.

THE DENDROGRAM

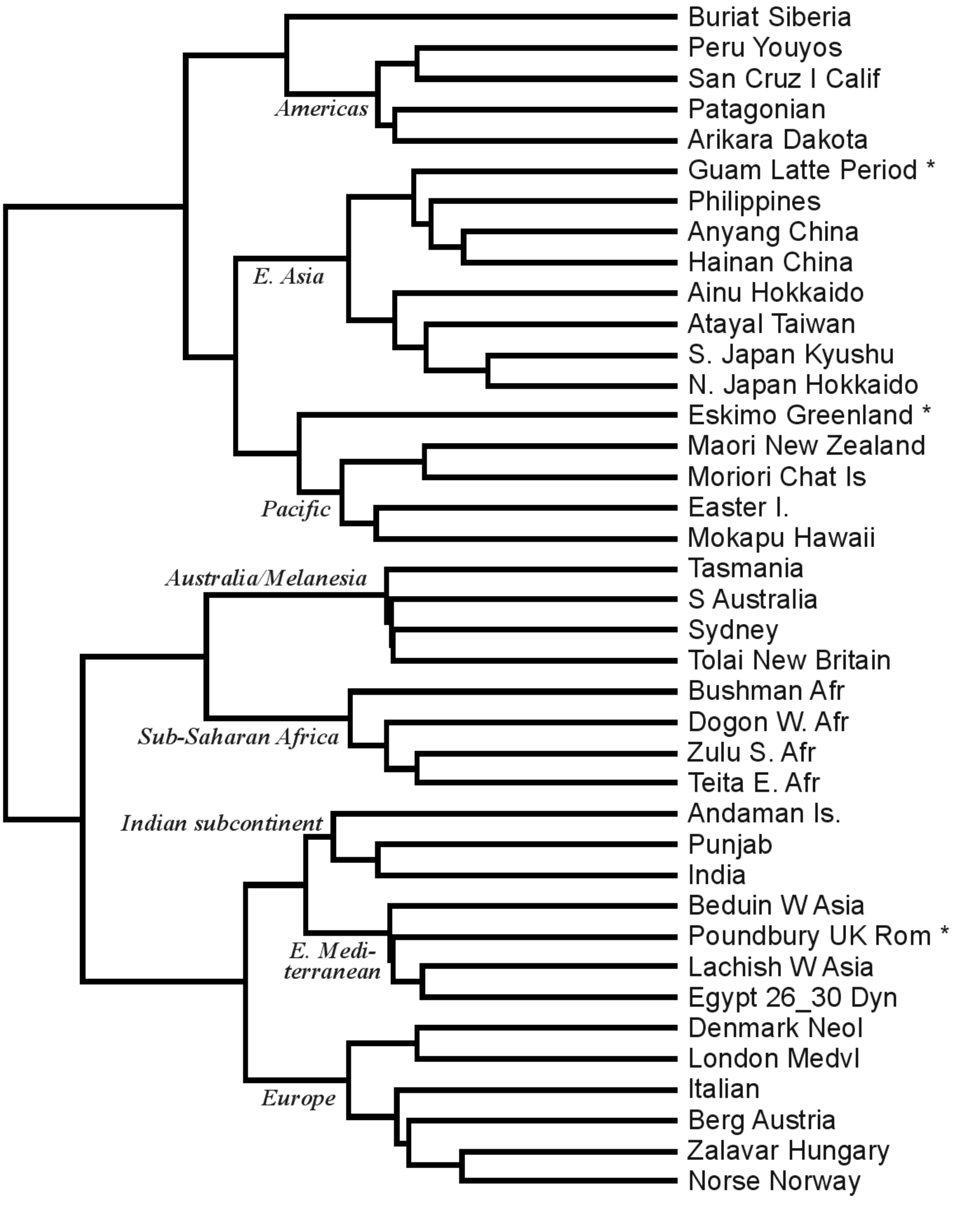


Figure 11. Dendrogram derived from scores on all 29 canonical variates. Asterisks mark samples that some might wish to argue are ‘misplaced’. However some anomalies seen in the scattergram of Figure 10 are now removed – for example there is separation between sub-Saharan African and SW Pacific samples.

The dendrogram of Figure 11 makes use of information in all 29 canonical variates, whereas the scattergram of Figure 10 used only the first two.

The clustering is, as with the scattergram, based only on cranial morphology. The structure of the dendrogram accords well with received belief about ancestry.

CORRELATION OF VARIABLES WITH SCATTERGRAM

Readers familiar with Principal Components Analysis (PCA) will know about *variable loadings*. These measure the correlation between the original variables and the principal components.

Similar statistics of correlation are available in linear discriminant analysis. They measure the correlation between the original variables and the canonical variates. These statistics are sometimes known as the *structure coefficients.*

Figure 12 shows a plot of the structure coefficients that are relevant to the groups of crania plotted on the scattergram of Figure 10.

In summary, groups that appear towards the extremes of CV1 or CV2 in Figure 10 are characterised by high values of the variables that appear towards the extremes of CV1 or CV2 in Figure 12.

A reverse interpretation is also to be made. Samples that appear towards the extremes of CV1 or CV2 in Figure 10 are characterised by *low values* of the variables that appear in the opposite position on the diagram.

The adjectives ‘high’ and ‘low’ refer to values relative to other variables, not to the absolute values. Thus the Buriat are characterised by relatively high values of AUB and low values of GOL, PAC, NOL, FMB, etc.. By contrast, Sub-Saharan Africans and people from the SW Pacific show the reverse of these characteristics.

When examining a scattergram like Figure 10, we are particularly interested in groups that lie away from the centroid - the position where the lines for CV1 and CV2 cross. By contrast, those groups that are close to the centroid are average for those CVs.

Likewise in Figure 12, variables that appear close to the centroid are not helping to distinguish the groups on the first two CVs. In fact five variables that do not reach a correlation value of 0.2 on either CV have been omitted from Figure 12.

Though these five variables are not discriminators in Figure 12, we must remember that the first two CVs merely give us a summary look at the analytical results. These five variables may well help to discriminate groups in the remaining 27 CVs.
d
In similar vein, we can note that groups that are near the centroid on Figure 12 may be well discriminated on one or more of the remaining CVs.

All the discussion in this section is relevant to discrimination based on shape alone, and shown in the scattergram of Figure 10. The discussion is not relevant to LDA and NNDA included in the freely distributable package, and which are based on both size and shape and all CVs.

This sort of interpretation is to be made with care. Further explanation is beyond the scope of this manual, but the reader is referred to literature that deals with the use of biplots in general (though Figures 10 and 12 are not strictly speaking biplots).[36]

---

[36] Gabriel, K.R. (1971). "The biplot graphic display of matrices with application to principal component analysis". *Biometrika* **58** (3): 453-467.

Gower, J.C. and Hand, D.J. 1996. *Biplots.* London: Chapman & Hall.

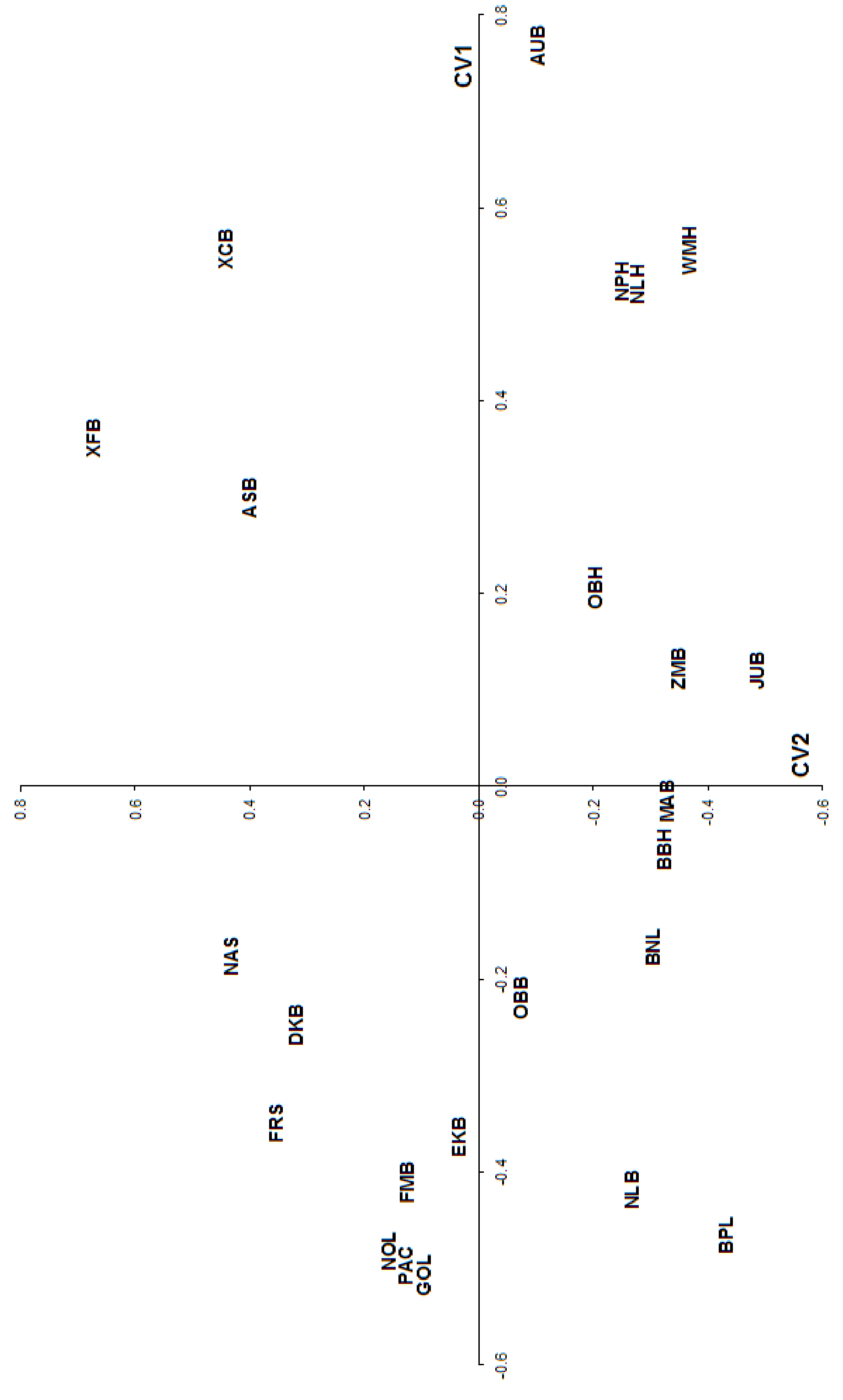


Figure 12. The correlation between the most influential variables and the scattergram of Figure 10.

APPENDIX 1

HOW TO LOWER THE RISK OF MEASUREMENT ERRORS

Measurement errors lead to erroneous CRANID results. In effect, the user is analysing a non-existent cranium.

Most errors are caused by users not understanding the definitions of landmarks and measurements. Some are caused by misreading calipers or erroneously transliterating entries into the spreadsheet Cr6Form.csv.

This section offers a method of lowering the risk of error in the entries in Cr6Form.csv.

The statistics used to define suspect measurements.

The standard deviation can be used to show how far measurements on an unknown cranium are away from the mean for the database of 3,163 crania. This method can be used for each of the 29 measurements in turn.

On the assumption that great distance from the mean may signify an erroneous measurement, we are interested in easily finding measurements that are both suspiciously small or suspiciously large.

Percentages under the normal curve.

In the case of a normal curve, the following percentages apply to distances from the mean:

1. 68.3% of individuals lie within one standard deviation.

2. 95.5% of individuals lie within two standard deviations.

3. 99.7% of individuals lie within three standard deviations.

Turning now to the CRANID database of 3,163 individuals, we have the following approximate distribution.

1. 2,159 crania lie within one standard deviation, leaving a residue of 1,004 that lie beyond one standard deviation.

2. 3,019 crania lie within two standard deviations, leaving a residue of 144 that lie beyond two standard deviations.

3. 3,154 crania lie within three standard deviations, leaving a residue of nine that lie beyond three standard deviations.

These numbers are approximate, because of slight deviations from normality in each of the variables in the CRANID database. The existence of skewedness, which some have sought to correct by log transformation in craniometric data, is minimal for all the 29 measurements. In fact analysis indicates that log transformation introduces rather more negative skewedness into the measurements than the positive skewedness in the raw data. So the measurements are best used in their raw state, as indeed CRANID does.

The lack of skewedness is shown in Table 13. Positive skewedness produces a median that is greater than the mean. In Table 13 it is obvious that the median and the mean are close to each other for all 29 measurements in the database of 3,163 crania.

	Mean	Median
GOL	179.2	179.0
NOL	176.8	177.0
BNL	99.1	99.0
BBH	131.8	132.0
XCB	137.0	137.0
XFB	114.1	114.0
AUB	120.5	120.0
ASB	106.9	107.0
BPL	97.2	97.0
NPH	65.8	66.0
NLH	49.9	50.0
OBH	33.6	34.0
OBB	39.4	39.0
JUB	114.6	114.0
NLB	26.0	26.0
MAB	63.1	63.0
ZMB	94.5	94.0
SSS	23.2	23.0
FMB	96.9	97.0
NAS	16.6	17.0
EKB	97.2	97.0
DKB	21.6	22.0
WMH	22.5	22.0
FRC	109.6	109.0
FRS	25.6	26.0
PAC	110.8	111.0
PAS	23.8	24.0
OCC	95.3	95.0
OCS	27.7	28.0

Table 13. Shows how mean and median, for each of the 29 variables in the database, are close to each other.

How to use the statistics of standard deviations.
To apply this information to the 29 measurements on an unknown cranium, and if we are looking for measurements that are suspected to be in error, we can use the following statistics given in Table 14

			<	<	<		>	>	>	
Variable	MyFile.csv value	GROSSLY SUSPICIOUS	3 Std Dev smaller SUSPICIOUS	2 Std Dev smaller WORTH CHECKING	1 Std Dev smaller	Mean	1 Std Dev larger	2 Std Dev larger WORTH CHECKING	3 Std Dev larger SUSPICIOUS	GROSSLY SUSPICIOUS
GOL	183		153.4	162.0	170.6	179.2	187.7	196.3	204.9	
NOL	180		152.8	160.8	168.8	176.8	184.8	192.8	200.9	
BNL	106		82.0	87.7	93.4	99.1	104.8	110.5	116.2	
BBH	134		110.5	117.6	124.7	131.8	138.8	145.9	153.0	
XCB	144		114.4	121.9	129.5	137.0	144.5	152.1	159.6	
XFB	124		94.3	100.9	107.5	114.1	120.7	127.3	133.9	
AUB	126		98.1	105.6	113.0	120.5	127.9	135.4	142.8	
ASB	117		89.2	95.1	101.0	106.9	112.8	118.7	124.6	
BPL	101		78.1	84.5	90.8	97.2	103.6	110.0	116.4	
NPH	73		49.3	54.8	60.3	65.8	71.3	76.8	82.3	
NLH	54		38.0	42.0	45.9	49.9	53.9	57.8	61.8	
OBH	37		26.8	29.0	31.3	33.6	35.8	38.1	40.3	
OBB	39		33.3	35.3	37.3	39.4	41.4	43.5	45.5	
JUB	119		95.5	101.9	108.2	114.6	120.9	127.2	133.6	
NLB	24		18.8	21.2	23.6	26.0	28.3	30.7	33.1	
MAB	66		50.5	54.7	58.9	63.1	67.3	71.5	75.8	
ZMB	98		76.8	82.7	88.6	94.5	100.4	106.3	112.3	
SSS	25		13.3	16.6	19.9	23.2	26.6	29.9	33.2	
FMB	101		83.2	87.8	92.3	96.9	101.5	106.1	110.6	
NAS	18		8.8	11.4	14.0	16.6	19.2	21.8	24.4	
EKB	101		84.1	88.5	92.8	97.2	101.6	106.0	110.4	
DKB	24		14.0	16.5	19.0	21.6	24.1	26.6	29.1	
WMH	26		13.4	16.4	19.5	22.5	25.6	28.6	31.6	
FRC	104		93.0	98.5	104.1	109.6	115.1	120.6	126.1	
FRS	20		16.7	19.6	22.6	25.6	28.6	31.5	34.5	
PAC	110		90.3	97.1	104.0	110.8	117.6	124.5	131.3	
PAS	25		14.9	17.9	20.8	23.8	26.8	29.8	32.8	
OCC	95		77.3	83.3	89.3	95.3	101.4	107.4	113.4	
OCS	26		16.7	20.4	24.0	27.7	31.4	35.1	38.8	

Table 14. This table helps find erroneous measurements. The second column contains the measurements of the Shoalhaven cranium, the results for which were used to illustrate the output from Cr6b.

With Table 14, we can use the following procedures when looking for possibly erroneous measurements:

1. Accept as unsuspicious all measurements that lie within one standard deviation (the green zone).

2. Perhaps check all measurements that lie between one and two standard deviations (the buff zone).

3. Treat as suspicious all measurements that lie between two and three standard deviations (the pink zone).

4. Treat as grossly suspicious all measurements that lie beyond three standard deviations (the red zone).

If we examine the values for Shoalhaven, we see that none of them gives us reason for suspicion when we compare them with Table 14. All but two measurements lie within the green zone. The only two worth checking are:

> an ASB of 117 mm lies within the 'worth checking' zone of being rather, but not suspiciously, large;
>
> an FRS of 20 mm lies within the 'worth checking' zone of being rather, but not suspiciously, small.

We must note that applying this procedure does not ensure that all measurements within the green zone on the unknown cranium are correct. Nor does it demonstrate that measurements within the pink and red zones of Table 14 are necessarily incorrect.

The purpose of the Table 14 is to arouse suspicion of error, not provide proof.

How to enter the values for your unknown cranium into Table 14.

Applying this analysis should be part of the user's standard operating procedure before relying on CRANID results.

To this end, the package Cr6b includes the file *Checking table.xls.*

The user:

> opens *Checking table.xls*;
>
> copies the column of 29 measurements from a MyFile.csv file;

pastes the column into the second column in *Checking table.xls.*

Checking of each of the 29 measurements is then done by inspection.

* * * * * * * * * * * *

APPENDIX 2

NOTES ON CRANID MEASUREMENTS FOR BONES CLONES CAST BC-110

INTRODUCTION

The purpose of these notes is to help CRANID users take their 29 measurements correctly.

Users of CRANID live around the world. Up to now there has been no single skull for all users to get ready access to. The reason for having a single skull is that users can test its accredited set of measurements against their own. Thereby they will detect their errors.

Most errors come from two unrelated sources:

misunderstanding of landmarks and measurements;

misreading of calipers.

Both these sources of error can be mitigated by practising on a skull that has an accredited set of measurements.

After consultation with several users, I chose the cast made by Bone Clones Inc, and labelled BC-110 (African male). It is already available in many institutions. It can be readily purchased from the supplier:

http://boneclones.com/

The supplier is long established in the business of making and selling casts of bones.

Though some landmarks on the cast of BC-110 are somewhat unclear, any resulting problem can be readily dealt with by users of these notes.

A suggested set of 29 measurements for BC-110

The following set of measurements has been arrived at in consultation between myself and other CRANID users. To that extent they are accredited. Some measurements may be open to slight debate, and I shall be happy to hear of any objections users may have.

It is appropriate here to make an aside on the ancestry of BC-110. This aside does not affect its usefulness as a cast to practise on, however it so happens that a CRANID analysis of BC-110 shows that it occupies an unusual position in 29 dimensional space. Its morphological space is populated by both African and SW Pacific skulls. We already know that African and SW Pacific skulls sometimes resemble each other when the 29 CRANID measurements are used.

In this connection, remember that CRANID is not offered as the only tool for determining ancestry. Other observations should be used as checks.

LIST OF 29 accredited MEASUREMENTS FOR BC-110

The 29 accredited measurements of BC-110 in mm are:

GOL	177
NOL	173
BNL	99
BBH	123
XCB	127
XFB	110
AUB	113
ASB	99
BPL	97
NPH	65
NLH	53
OBH	34
OBB	40
JUB	113
NLB	34
MAB	62
ZMB	99
SSS	27
FMB	100
NAS	19
EKB	98
DKB	24
WMH	20
FRC	99
FRS	25
PAC	114
PAS	25
OCC	83
OCS	25

SOME PROBLEMS ASSOCIATED WITH BC-110

BC-110 offers some problems for measurement. Most of these relate to slight blurring of detail in the cast, and are unlikely to reappear as problems on a real skull.

Definitions of all 29 measurements are given, by description and illustration, on pages 36 to 51 in this manual. In this section I do not rehearse them all - only those that relate to slight problems with BC-110.

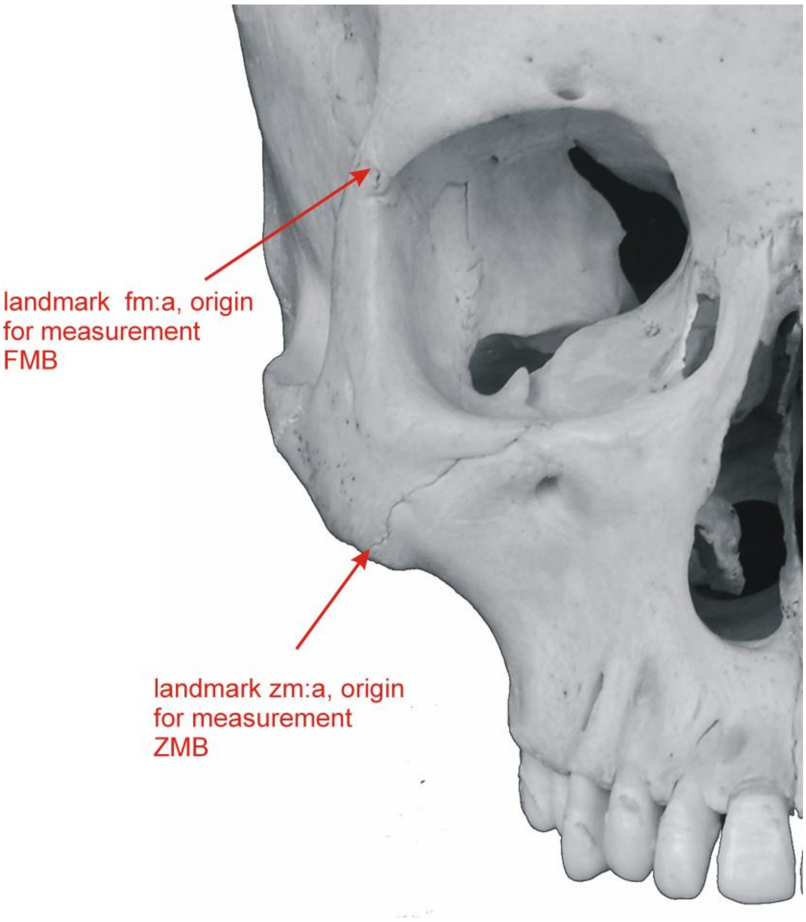
FMB and ZMB

ZMB is measured between zm:a and zm:a.

FMB is measured between fm:a and fm:a.

Unfortunately, the sutures required for both measurements are virtually invisible in BC-110. This poor definition appears to be due to smoothing during moulding or casting.

No matter. In most real skulls that you will measure, the sutures will be clearly visible.



Therefore I have provided illustrations for zm:a and fm:a taken from a real skull.

The relevant definitions from Howells are:

Zygomaxillare anterior zm:a DI 5.

"The intersection of the zygomaxillary suture and the limit of the attachment of the masseter muscle, on the facial surface."

17. ZMB Bimaxillary breadth, DI 5

"The breadth across the maxillae, from one zygomaxillare [anterior] to the other."

Howells elaborates on the definition of ZMB by pointing out that zm:a is on the facial surface, and not (as are other measures of bimaxillary breadth) on the lower edge of the zygomaxillary suture. The two measures are, however, close to each other.

Frontomalare anterior fm:a DI 1, 7.

"The most anterior point on the fronto-malar suture. It may be found with the side of a pencil lead held in the transverse plane."

19. FMB Bifrontal breadth, DI 7

“The breadth across the frontal bone between frontomalare anterior on each side, i.e., the most anterior point on the fronto-malar suture.”

In the case of BC-110 I have estimated where the landmarks zm:a and fm:a are, and marked them. A photo of this marked version of BC-110 is shown above.

Whether or not these are the best estimates may be open to debate, but for the purpose of these notes debate is not required. This is because the sutures will almost certainly be clear in any real skull you measure.

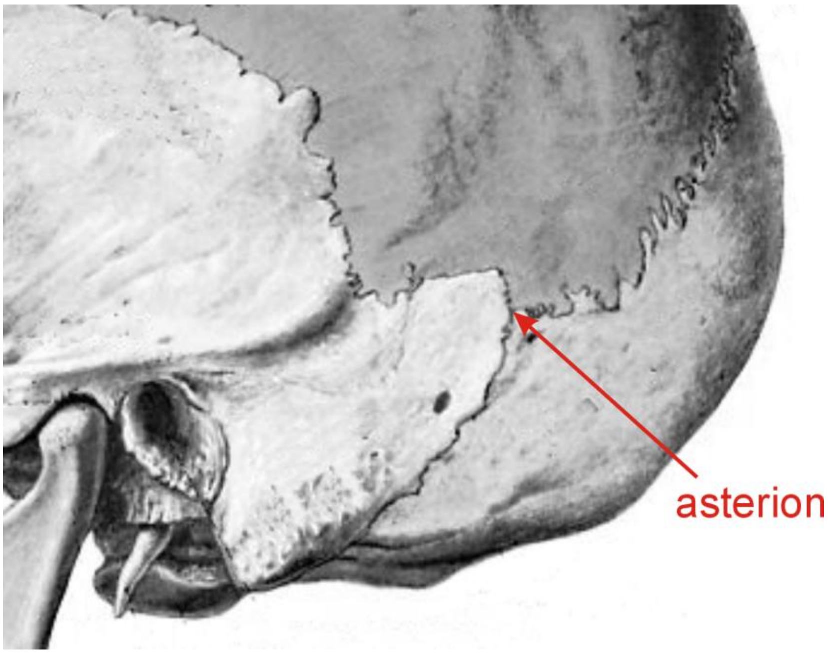
All the user needs to know about BC-110 is that the resulting accredited measurements in mm, using these marks, are as follows:

ZMB	99
SSS	27
FMB	100
NAS	19

When practising these four measurements on BC-110, I suggest you take your sliding calipers, and lock them on the measurement of 99 mm for ZMB. Use BlueTac, if you do not have a locking screw. Then get the feel for what ZMB is by applying the calipers to the cast of BC-110. Measure SSS.

Repeat this exercise for FMB and NAS.

The asterion



Adapted from Sobotta

In the case of BC-110, the asterion is poorly defined, particularly on the individual’s left side.

This problem is not so much a problem with the casting as with the rather chaotic growth of minor supernumerary bones in the original skull.

A clearly defined sketch of the location of an asterion - with a clarity that you are likely to meet in a real skull - is shown above.

The following definition of the asterion, and the consequent measurement ASB, are taken from Howells (and reproduced in the CRANID manual):

> **Asterion as DI 2, 3, 6.**
> "The common meeting point of the temporal, parietal, and occipital bones, on either side."
>
> **8. ASB Biasterionic breadth, DI 6**
> "Direct measurement from one asterion to the other."

The measurement of ASB on BC-110 is estimated as 99 mm. As advised above for ZMB and FMB, lock your calipers on 99 mm, lay them on the cast and get a feel for the measurement ASB.

To repeat - on any real skull you measure, both asterions are likely to be clear.

MAB

Howells defines MAB (palate breadth, external) as:

> The greatest breadth across the alveolar borders, wherever found, perpendicular to the median plane.

Note that this location is usually at M2, but is not necessarily so. In the case of BC-110, there is no M3 so the greatest width is at M2.

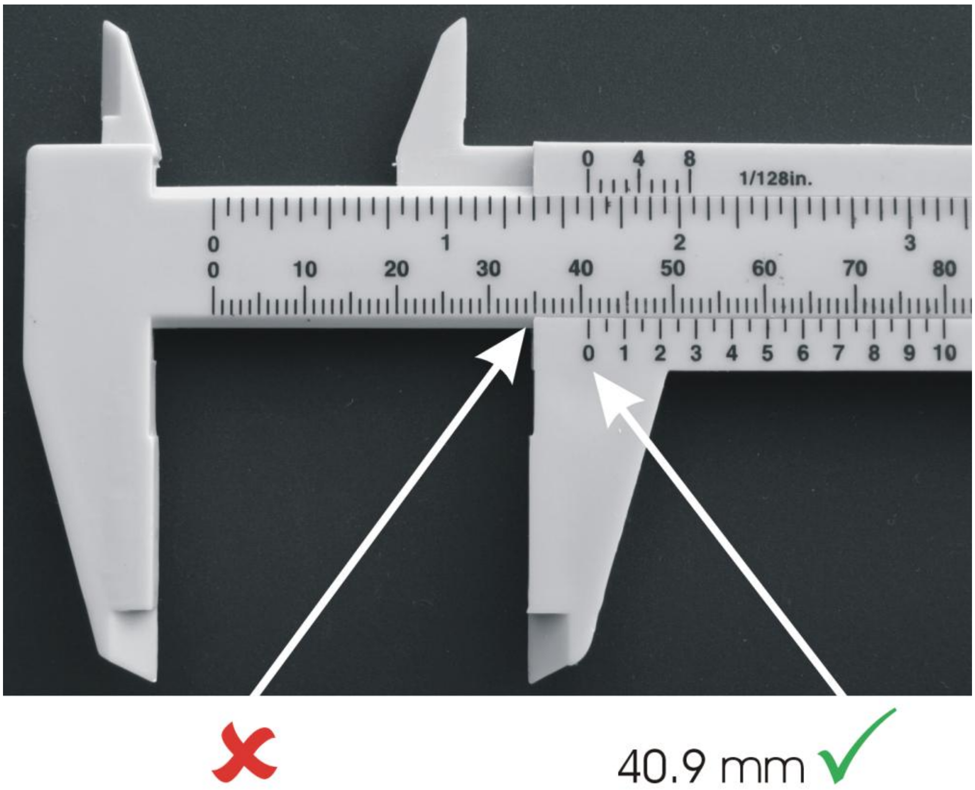
Howells further tells us to:

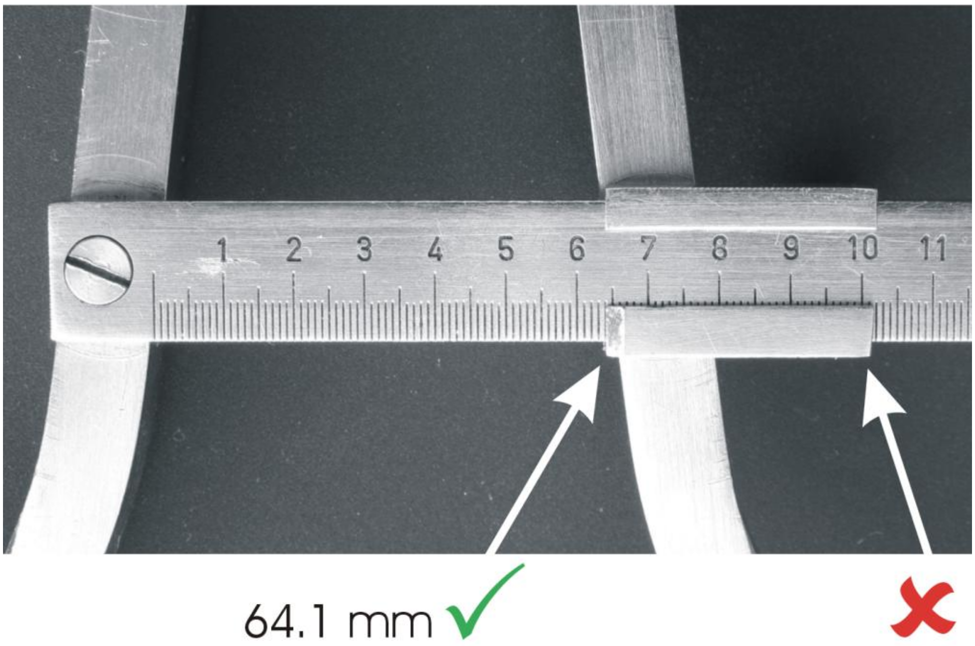
> 1. Measure to the bone, not the teeth, unless the roots are exposed at the widest point.
>
> 2. If there is some obvious special growth on one side, make allowance.

When Howells writes ‘special growth on one side’ I take it he is referring to pathology. In other words we should estimate what the width would be without the pathological growth. I do not interpret the bilateral growth of bone on BC-110 as pathology, though others may disagree.

CAUTION WITH CALIPERS

Dial and digital calipers lower the risk of misreading.





Dial type carbon fibre spreading calipers are readily available on eBay. Digital carbon fibre spreading calipers are not so readily available, though there is a requirement for them in other fields where scratching by sharp metal points is taboo (as it should be in osteology).

Digital spreading calipers have already come on the market (see page 53).

In spite of recent innovations in caliper design, these notes assume many users will be measuring with traditional types of sliding and spreading calipers.

The primary risk with sliding and spreading calipers is that the user reads at the wrong place on the sliding part.

The illustration above shows the correct places for making a reading (sliding above, spreading below).

The correct procedure for sliding caliper is self-evident. However in the case of spreading calipers, the correct location may not be so immediately obvious, though the correct end of the sliding bar is usually beveled.

Lapse of attention is possible, even with the most experienced user. I therefore recommend that the correct location for taking a reading be marked in red on the calipers themselves, using an indelible marker pen.

ACKNOWLEDGMENTS

Many people helped me in the decision to use BC-110 as the control skull for CRANID users.

I am particularly grateful to Ambika Flavel and Tania Kausmally for helping me with the process of getting an agreed set of measurements. However, I take responsibility for any minor errors that may exist in the final accredited set.

FINIS