

Identifying Misidentifications: Fordisc 3.1 Application to Korean Crania

Addison N. Springer

Department of Anthropology, University of Tennessee

May 15, 2025

Abstract

When cases are of unknown identity or race, determining the population of the individual significantly increases the chances of determining the identity of that victim. Forensic anthropologists have been using skeletal methods to estimate population affinity for decades. Fordisc (Jantz and Ousley 2005) is the primary method used by forensic anthropologists when building biological profiles for forensic cases. The results from Fordisc are assumed by users to be accurate, however, Fordisc can often yield misleading results or incorrect affinities. This means that the use of Fordisc to estimate population affinity is potentially misleading the biological profiles of forensic cases. Fordisc has only thirteen populations to compare cranial measurements to (five female and eight male), so if a case is outside of those populations, it is impossible for Fordisc (Jantz and Ousley 2005) to provide an accurate affinity. This study aims to find a pattern in the outputs produced when the cranial measurements are from a population not within the populations available. The measurements of 365 cases, 219 males and 146 females, were provided by the National Forensic Service in South Korea for use in this study. The craniometric data will be entered in the application following standard protocols. We expect the craniometric data to yield low probabilities or indecisive results. We hypothesize that the cranial measurements will affiliate with Japanese, Guatemalan, and Chinese crania due to the close regions and morphological history of the populations. We will conduct a statistical analysis of the typicality probabilities and posterior probabilities provided to find potential patterns in the output that indicate the affiliation could be incorrect. Even though we are unable to fix the programming of the application, if forensic anthropologists can begin to notice the possibility of misleading outputs, then the biological profiles will not be impeding the investigations.

Keywords: Population affinity, Fordisc

Identifying Misidentifications: Fordisc 3.1 Application to Korean Crania

According to the Federal Bureau of Investigation's Crime Data Explorer (CDE) there were 14,195 homicide victims in 2024, of which 454 remained unidentified (CDE, 2024). The National Missing and Unidentified Persons Systems (NamUs) reported over 15,000 active unidentified persons cases as of July 2024, with 24% lacking a clear population affiliation. Forensic anthropology is the application of modern biological and physical anthropological methods to unidentified remains that aids law enforcement in their identification efforts. The primary role of a forensic anthropologist (FA) is to build a biological profile that consists of age, sex, stature, and population affinity estimations. These estimations are done through metric and morphological methods and help law enforcement narrow their search parameters to improve identification outcomes. Population affinity is a crucial factor when it comes to searching for these identities. When FAs are making these estimations, they are trying to provide the most accurate information possible, however, the methods they use can often have high error rates. When the outcomes of these methods result in misclassifications it can be detrimental to search efforts and skew age, sex, and stature estimations as well.

One of the most popular methods used by FAs is Fordisc, an application developed by Drs. Jantz and Ousley at the University of Tennessee. The program uses 42 standardized cranial measurements and runs them through a discriminant function analysis to compare them to a reference database. This predicts whether a case identifies with a group based on a linear combination of interval predictor variables (Stockburger, 2018). The program provides two key outputs: a posterior probability (PP) and three different typicality probabilities (TP). The posterior probability is the relative probability that a sample belongs to a population within the program and is measured against all selected populations at once. The sum of the posterior

probabilities will always equal one because of this. The standard threshold for PP is 0.8 or greater. Another significant output is the TPs. The TPs are based off of different statistical analyses and the TP used in this study is called the Ranking typicality or Typ. R for short. This TP ranks the crania according to how typical or atypical it is when compared to one population at a time assuming it is from that particular reference population. The standard threshold for TPs is usually 0.01 or greater.

Limitations arise when using this program though as there are currently only eight populations (White M/F, Black M/F, Hispanic M/F, Chinese M, Guatemalan M, Japanese M, Vietnamese M, and Native American M/F) and it is unlikely that anymore will be added. This limitation results in a higher likelihood of misclassification when Fordisc on unrepresented groups and hinders investigations. This also means that if a case is from outside of those few represented populations, it is impossible to provide an accurate population affinity.

Literature Review

Previous studies have highlighted significant limitations in the use of Fordisc for ancestry estimation. For example, Elliott and Collard (2009) conducted a study evaluating Fordisc's classification accuracy using 200 cranial samples. When the analysis was performed using varying combinations of ten cranial measurements and unspecified sex, Fordisc correctly classified only 2% to 17% (2–34 out of 200) of the cases when applying a posterior probability threshold of 0.5 and a typicality probability threshold of 0.01. When the threshold was raised to 0.8 PP while keeping the same measurement and sex parameters, correct classifications dropped further to between 0.5% and 4.5% (0–9 out of 200 cases). However, when all cranial measurements were included, all population groups were considered, and sex was specified, the

correct classification rate improved significantly to 78% (156 out of 200 cases) at the $PP \geq 0.8$ and $TP \geq 0.01$ thresholds.

These findings demonstrate that classification success is highly dependent on the completeness of the data and the specification of sex. This presents a considerable limitation for forensic casework, where trauma, fragmentation, pathological changes, or taphonomic damage often result in incomplete or compromised cranial measurements. As a result, Fordisc may be less reliable in realistic forensic contexts, where perfect data is rarely available and suggests that previously classified cases may have been misidentified.

Dr. Dudzik and Dr. Jantz (2016) did a study on the misclassifications of Hispanics and Asians in Fordisc and compared 29 Korean males and 15 Korean females to the known populations in Fordisc. Their study resulted in the conclusion that a Japanese classification cannot rule out the possibility of the individual being Hispanic due to the admixture of past populations and debated origins (Dudzik & Jantz 2017). They also noted their limitations due to a lack of data available because of discrimination towards Asians and Hispanics (Dudzik & Jantz 2017). This could mean cases previously identified using Fordisc were wrongly classified. Similarly, Hughes et al. (2019) conducted a study on the misclassification of Hispanic or Latino individuals where they hypothesized that Fordisc's population affinity results were directly related to craniometric trends that are and were influenced by the admixture of past populations. They found that almost 70% of their cases would be likely be affiliated with an incorrect population (Hughes et al., 2019). For the purpose of their study, a case was considered as classified correctly when it was affiliated with either Hispanic or Guatemalan and disregarded the typical standards of a PP greater than 0.8 and a TP greater than 0.01.

Methods

This research aims to study the outputs to determine if there is a pattern that can indicate a case has the potential to be from an outside population. As well as if a case is from an outside population, will the program assign the sample to a population geographically close to the sample's true population. To investigate these limitations, this study analyzed 319 Korean forensic cases to assess Fordisc's classification reliability and misclassification patterns.

For the study, 319 Korean forensic cases were provided by the National Forensic Service. There were 172 males and 147 females. The mean age for males was 46.7 years old with a standard deviation of 13.3 years and for females 44.9 with a standard deviation of 14.5. The total age range for all cases was 20 to 86 years old. The crania had computed tomography scans done and seventeen measurements were taken where possible (Table 1).

- **GOL** – Maximum Cranial Length
- **XCB** – Maximum Cranial Breadth
- **ZYB** – Bizygomatic Breadth
- **BBH** – Basion-Bregma Height
- **MAB** – Palate Breadth
- **MAL** – Palate Length
- **AUB** – Biauricular Breadth
- **UFHT** – Upper Facial Height
- **WFB** – Minimum Frontal Breadth
- **UFBR** – Upper Facial Breadth
- **NLB** – Nasal Breadth

- **EKB** – Biorbital Breadth
- **FRC** – Frontal Chord
- **PAC** – Parietal Chord
- **OCC** – Occipital Chord
- **FOB** – Foramen Magnum Breadth
- **DKB** – Interorbital Breadth
- **MDH** – Mastoid Height

Due to the incomplete nature of some crania, the number of measurements available varied across cases. Fordisc's internal filtering selected only reference samples with matching measurement sets for analysis. These measurements were input into the program following best practice protocols, the sex was specified, and the crania were run against the Forensic Data Bank (FDB) sample populations. The stepwise option was chosen when analyzing the crania as well.

For consistency only the top three populations were recorded based on the PP and the corresponding TP for each case. Since it is known that there are no Korean cranial measurements in the FDB, the Korean crania will identify with a population that is 100% incorrect. Therefore, this study focused on identifying patterns of consistent misclassifications. Additionally, a more conservative threshold of a PP of 0.9 or greater and a typicality between 0.01 and 0.05 was applied to evaluate whether stricter criteria would reduce the error risk.

Results

Of the 319 cases, 130 (40.7%) met the standard threshold of a PP of 0.8 or greater, suggesting high confidence in the assigned population despite the known absence of Koreans in Fordisc's database (Figure 1). 156 cases had the top TPRank between 0.003 and 0.103. This

indicates that while Fordisc assigned them to a population, the crania are generally more atypical than typical relative to the reference group. Only 29 cases (9.1%) fell below the standard threshold of 0.01 or greater (Figure 2). The three highest population affiliations for males (n=172) were:

1. American Indian (51)
2. Vietnamese (40)
3. Chinese (31).

The rest of the male cases affiliated with Japanese (28), Guatemalan (8), White (7), Hispanic (5), Black (2) (Figure 3). The three highest for females (n=147) were:

1. Japanese (57)
2. American Indian (41)
3. Hispanic (35).

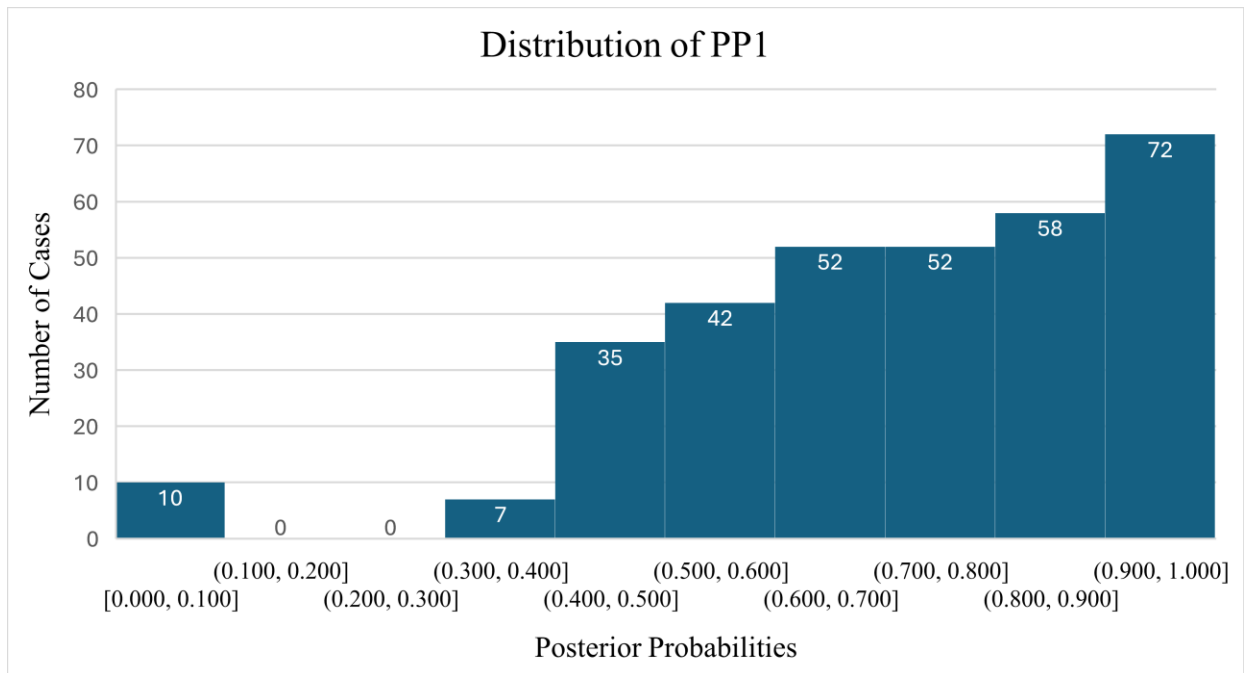
The remainder of the female cases affiliated with Black (10) and White (4) (Figure 3). Overall, 248 cases were classified as “Asian” Populations (American Indian, Japanese, Vietnamese, or Chinese). This is 77.7% of the total cases. The difference in male and female classification patterns may reflect morphological variation and is likely influenced by the limited female populations available in the program.

When the stricter threshold was applied of a PP of greater than 0.9 and a TP between 0.01 and 0.05 the error rate was reduced to 2.5% (n= 8, 5M, 3F), excluding cases labeled too dissimilar. These thresholds were chosen considering a TP shows the probability of the null hypothesis that an individual comes from a specific group. According to the manual, a TP above 0.05 should be ignored because “we do not have statistical grounds to reject the null hypothesis”

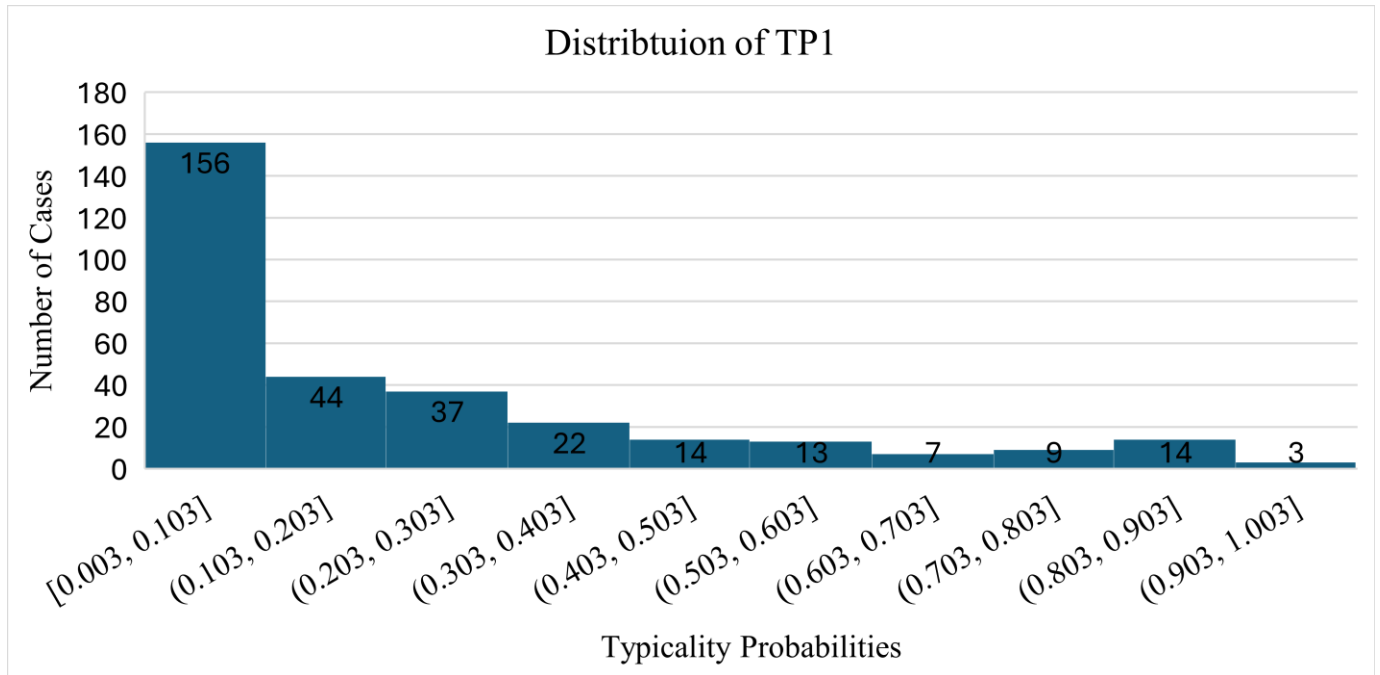
because the assumption is that a case does belong to one of the populations (Jantz & Ousley, 2005).

Figure 1

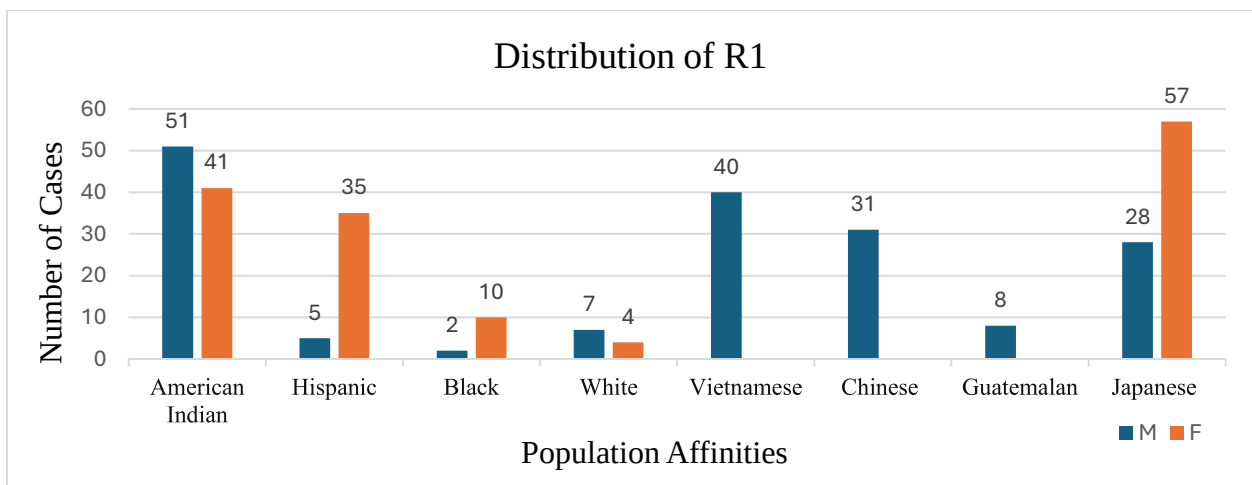
Distribution of PP 1.



Note. The distribution of the PPs of the top population that the program assigned to each case.

Figure 2*Distribution of TP Rank 1*

Note. The distribution of the TP of the top population that the program assigned to each case.

Figure 3*Distribution of R1*

Note. The top population that Fordisc assigned to each case for both males and females.

Discussion

Although the tool produces seemingly confident posterior probabilities, and the typicality values often reveal significant atypicality, there were no clear indications in the outputs that a case could be from an outside population. As shown in Figure 1, the majority of cases surpassed the 0.8 PP threshold, yet the low TP values depicted in Figure 2 indicate these assignments were largely atypical. The low TP values observed in the majority of cases serve as a critical indication for forensic anthropologists to proceed with caution because while Fordisc may assign a population with high PP, the low typicality indicates that the individual does not fit well within that population's morphological variation. Therefore, FAs should interpret Fordisc results carefully and contextually, and consider geographical, historical, and morphological factors beyond just the numerical outputs. Applying more conservative thresholds can reduce misclassification rates, but not enough to significantly reduce the risk of error to confidently estimate population affiliation based off of Fordisc's results alone.

The interpretation of the various geographic regions represented in Fordisc could prove useful. A Japanese affiliation could technically be an incorrect estimation, but if the FA building the biological profile interprets it as an Asian population affinity and includes similar geographic regions, the results could be considered "correct". Future improvements to forensic classification tools should prioritize expanding reference datasets to include currently underrepresented populations as well as a broad range of morphological variation.

Conclusion

Fordisc does not provide users with a clear indication that a case could be from an outside population, but careful observation and broad interpretations of population affinities could still

prove useful in identification efforts for law enforcement. The warning signs in built into the program should always be taken into consideration as well, such as TPs lower than 0.01 highlighted in red or PPs distributed relatively even across all populations.

References

- Dudzik, B., & Jantz, R.L. (2016). Misclassifications of Hispanics using Fordisc 3.1: Comparing cranial morphology in Asian and Hispanic populations. *Journal of forensic sciences*. 61.5: 1311–1318. Web.
https://utk.primo.exlibrisgroup.com/permalink/01UTN_KNOXVILLE/k96n06/cdi_wiley_primary_10_1111_1556_4029_13123_JFO13123
- Elliott, M., and Collard, M. (2009). FORDISC and the determination of ancestry from cranial measurements. *Biology letters* 5.6: 849-852.
<https://royalsocietypublishing.org/doi/full/10.1098/rsbl.2009.0462>
- Hughes, C. E., Dudzik, B., Algee-Hewitt, B., Jones, A., & Anderson, B.E. (2019). Understanding (mis)classification trends of Latin Americans in Fordisc 3.1: Incorporating cranial morphology, microgeographic origin, and admixture proportions for interpretation. *Journal of forensic sciences* 64.2: 353–366. Web. <https://doi-org.utk.idm.oclc.org/10.1111/1556-4029.13893>
- Jantz, R. L., & Ousley, S. D. (2005). FORDISC 3: Computerized forensic discriminant functions, Version 3.0. *The University of Tennessee, Knoxville*.
- Mariana, E., Collard, M. (2009). Fordisc and the determination of ancestry from cranial measurements. *Biol. Lett.* 5:849–852. <http://doi.org/10.1098/rsbl.2009.0462>
- Ousley, S, & Jantz, R.L. (2013) Fordisc 3: Third generation of computer-aided forensic anthropology. *Rechtsmedizin (Berlin, Germany)* 23.2: 97–99. Web.
https://utk.primo.exlibrisgroup.com/permalink/01UTN_KNOXVILLE/k96n06/cdi_crossref_primary_10_1007_s00194_013_0874_9
- Stockburger, D. (2018). Discriminant function analysis. In *The SAGE encyclopedia of educational research, measurement, and evaluation* (Vol. 4, pp. 526-532). SAGE Publications, Inc., <https://doi.org/10.4135/9781506326139.n204>