# Chapter 6.1: Bayesian inference personal identification

C.E.H. Berger[1,2], M. van Wijk[1], H.H. de Boer[1,3]

1 Netherlands Forensic Institute, The Hague, The Netherlands

2 Institute for Criminal Law and Criminology, Leiden University, The Netherlands

3 Dept. of Pathology, Amsterdam University Medical Center, University of Amsterdam, The Netherlands

**Abstract**

Providing an opinion on (elements of) the identity of heavily decomposed, often skeletonized remains is generally the mainstay of a forensic anthropologist's daily work. Often, forensic anthropologists accompany such opinions with a measure of uncertainty, such as a confidence interval. Such statements, however, give posterior probabilities without taking into account all the other (non-anthropological) evidence and information in the case. The application of logic, through the use of Bayes' theorem can provide a solution for this issue. This chapter explores how a Bayesian approach can be applied to interpret features observed during the examination of skeletal identifiers. It specifically focusses on two basic elements of the forensic anthropological biological profile; one with a binary outcome (sex estimation) and one with a categorical or continuous outcome (age estimation). Besides others, the formulation of propositions, the calculation of likelihood ratios, the choice of reference data, and the combination of evidence are discussed.

**Keywords:** Forensic anthropology, identification, sex estimation, age estimation, investigative, evaluative, Bayesian inference, criminalistics.

# Introduction

The identification of heavily decomposed, often skeletonized remains is an important part of a forensic anthropologist's work. Irrespective of the context of the case, such an identification effort basically entails observing the identifying features of the remains and using these observations to make an evidential statement on the identity of the decedent. Depending on the circumstances, such a statement may be rather general (e.g. focusing on providing sex or approximate age group of the decedent) or very detailed (e.g. focusing on a specific individual).

The evidential strength of the observations can vary and as a result, some opinions may be very strong while others may be much less robust. It is therefore important to accompany an opinion with a statement on the strength of the evidence. Often, forensic anthropologists accompany their opinions with a measure of uncertainty, such as a 95% confidence interval resulting in statements like "the estimated age of the decedent is between 36 and 48 years". Such a statement, however, relates to a posterior probability, without taking into account all the other (non-anthropological) evidence and information in the case. Why this is problematic is discussed in Chapter 3.2. In other instances, forensic anthropologists may revert to statements such as "consistent with", "in line with" or "cannot be excluded" to give some idea of the evidential strength of their conclusion. However, such statements are vague and suggest much more than they actually convey (see Chapter 3.2 for more detail).

The application of logic by using the likelihood-ratio framework and Bayes' theorem can overcome the above-mentioned shortcomings. This chapter explores how the Bayesian approach can be applied during the examination of skeletal identifiers. In the first part, we focus on the estimation of two elements of the forensic anthropological biological profile, namely sex and age. These two elements are chosen since they represent different outcomes:

a binary one (male vs. female), and a categorical or continuous one (an age group or age-at-death). In the latter part, the possibilities and challenges related to combining results in order to present a comprehensive biological profile or evidential strength are discussed.

Readers of this chapter are expected to have a basic understanding of Bayesian inference in forensic science. Otherwise, it is best to first read Chapter 2.3 and 6.2 in this book, or Chapter 2, 3, and 5 of “Interpreting Evidence” (Robertson et al. 2016).

# Sex estimation

Sex estimation is one of the key aspects of the biological profile. First, because an accurate estimation of sex virtually rules out 50% of the world’s population as potential sources of the skeletal material under study. Also, the other estimation methods for age or ancestry for example are mostly sex-specific. Inaccurate sex estimation therefore has a considerable impact on the accuracy of the biological profile as a whole.

When the likelihood-ratio framework is applied, the interpretation starts with the formulation of two mutually exclusive hypotheses (H1 and H2). The propositions are implicitly guided by the research question, so in this case they are:

*H1: The remains are female.*

*H2: The remains are male.*

The likelihood ratio (LR) of a skeletal feature for this pair of propositions is given by the probability of observing the feature under proposition H1, divided by the probability of observing the same feature under proposition H2. A skeletal feature that is strongly dependent on sex will have a high evidential strength. In contrast, a skeletal feature that is independent of sex will be equally distributed between females and males, and result in an LR close to 1.

For this chapter, we focus on the generally accepted sex estimation method of Phenice (1969), as revised by Klales et al. (2012), consisting of scores for the ventral arc (VA), subpubic concavity (SPC) and medial aspect of the ischiopubic ramus (MA). First, we consider how scoring a single element of the Klales' method can be used to calculate an LR for that specific observation. For this, we use the data in Table 1, derived from readily available published data (Kenyhercz et al. 2017).

**Table 1.** *Relative frequencies (%) of the expressions of the ventral arc (VA), suprapubic concavity (SPC) and medial aspect of the ischiopubic ramus (MA) for males and females in a pooled (black and white) US population (adopted from Kenyhercz et al. 2017). The possible scores for the expressions are 1 to 5.*

| | **true sex** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|---|
| **VA feature** | **female** | 65.1% | 26.1% | 6.9% | 0.6% | 1.2% |
| | **male** | 0.3% | 3.3% | 13.1% | 40.2% | 43.1% |
| **SPC feature** | **female** | 55.4% | 32.0% | 8.1% | 4.0% | 0.6% |
| | **male** | 0.2% | 3.6% | 7.7% | 52.7% | 35.9% |
| **MA feature** | **female** | 9.9% | 32.4% | 46.9% | 8.9% | 1.8% |
| | **male** | 0.1% | 1.0% | 17.2% | 39.9% | 41.8% |

In this example, the LR is given by the relative frequency of a given feature in females, divided by the relative frequency of that same feature in males. If the pelvis under study is for instance scored as 2 for the SPC, the LR for the above-mentioned pair of propositions will be 32/3.6 = 9. In other words: finding a SPC score of 2 is 9 times more likely if the remains are female, than if they are male. The LRs per score per feature are provided in Table 2.

Since the relative frequencies vary per SPC score, different SPC scores will come with different LRs. When the features are almost evenly distributed between the sexes, the LR will gravitate towards 1, and thus the evidential strength of the feature would be minimal. In our example, this can be seen with the SPC score of 3, which is almost equally frequent between males and females and thus has an accompanying LR of about 1. SPC scores 4 and 5 are more likely found in males than in females, so these scores favor the alternative proposition (H2). This finding is also reflected by the LR: an SPC score of 5 comes with an LR of 0.017. The reciprocal of this number gives the LR for H2 versus H1. In other words, the findings are 60 times (1/0.017) more likely if the remains are male (H2), than if they are female (H1).

**Table 2**. *Likelihood ratios resulting from the data in Table 1 for the remains being female (H1), or male (H2). Score 0 refers to a non-observed feature, giving an LR=1.*

| | **0** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|---|
| **VA feature** | 1 | $2.2\times10^{2}$ | 7.9 | 0.53 | 0.015 | 0.028 |
| **SPC feature** | 1 | $2.8\times10^{2}$ | 8.9 | 1.1 | 0.076 | 0.017 |
| **MA feature** | 1 | 99 | 32 | 2.7 | 0.22 | 0.043 |

Comparing LRs shows which of the scores and features provide the strongest evidence. In our example, scores of 4 and 5 for SPC produce LRs that point more strongly in the direction of the remains being male than the same scores for MA. For these scores SPC thus produces higher evidential strength.

We can use LRs to combine the evidential strength of the three main features observed, even in cases where not all three features can be observed. The latter situation is indicated in Table 2 with zero score, which refers to a non-observed feature resulting in an LR equal to 1 (neutral evidence). Since all three LRs address the same pair of propositions, we can combine them. If we can assume the features to be conditionally independent, we can simply multiply the LRs to obtain the LR for the combined features.

Conditional independence may require some additional explanation. Observations (e.g. ‘A’ and ‘B’) are conditionally independent given ‘C’, if observing ‘A’ does not change our expectations of observing ‘B’, given ‘C’. In our example, the VA, SPC and MA scores are conditionally independent given sex, if observing the VA score does not change our expectations for the SPC or MA scores, given the remains are male. Note that VA, SPC and MA scores are clearly dependent, since they all depend on sex, but they can still be *conditionally* independent.

Kenyhercz et al. (2017) use a logistic regression formula to combine the scores of SPC, VA and MA. The result provides a posterior probability of the sex of the remains. However, reporting posterior probabilities is problematic when case-specific prior probabilities cannot be assigned (see also Chapter 3.2). This problem can be circumvented by reporting an LR instead of a posterior probability. If conditional independence is a reasonable assumption, the data from Table 1 and Table 2 can be combined to obtain a combined LR for VA, SPC and MA. As an additional benefit, using LRs allows us to deal with missing features, for instance if taphonomical alteration precludes the scoring of a feature. For such missing features we can simply assign LRs equal to one. Calculating LRs also helps to deal with prior probabilities of the propositions that are different from the typical 50% (based on a population with an equal number of males and females). This is helpful in cases for which the general population is not the most relevant one, and one proposition is favored over the other, such as in South Africa, where in most cases the unidentified skeletal remains are expected to be male (Baliso et al. 2019).

Table 3 shows a number of examples of combinations of scores for the three pelvic features, their combined LR and the posterior probability of the remains being female if we assume the prior probability to be 50%. With equal prior odds (1) and assuming conditional independence, Bayes' theorem gives us:

$$posterior\ odds\ =\ prior\ odds \times LR_{VA} \times LR_{SPC} \times LR_{MA}\ .$$

The first example in Table 3 considers a score of 4 for the VA feature, a non-observed SPC feature and a score of 2 for the MA feature, which results in posterior odds of $1 \times 0.015 \times 1 \times 32 = 0.48$, and thus a posterior probability of the remains being female of $\frac{0.48}{0.48+1} = 0.33$ or 33%.

**Table 3.** *A number of (hypothetical) examples of feature scores (with a zero score indicating a non-observed feature), the resulting combined LR (assuming conditional independence and with H1: female; H2: male), and the probability of the remains being female (assuming prior probability of 50%).*

| VA feature | SPC feature | MA feature | LR | P(female) |
|---|---|---|---|---|
| 4 | 0 | 2 | 0.48 | 32.6% |
| 3 | 0 | 2 | 17 | 94.5% |
| 3 | 0 | 4 | 0.12 | 10.5% |
| 3 | 3 | 4 | 0.12 | 11.0% |
| 4 | 3 | 4 | 0.0035 | 0.3% |
| 4 | 3 | 1 | 1.6 | 60.9% |
| 4 | 3 | 0 | 0.016 | 1.5% |
| 1 | 3 | 0 | $2.3 \times 10^2$ | 99.6% |

One way to evaluate the method is to check in how many instances our Bayesian approach favors the correct proposition. For the U.S. pooled population in the study by Kenyhercz et al. (2017), the logistic regression equation by Klales et al. (2012) estimated sex correctly in 94.2% of cases (i.e. when classifying based on a 50% threshold to the posterior probability). Kenyhercz et al. (2017) recalibrated the equation, improving the success rate to 95.8%. Kenyhercz et al. (2017) shared their raw data for the purpose of this chapter[1]. When we

[1] Klales AR. MorphoPASSE: the Morphological Pelvis and Skull Sex Estimation Database. Version 1.0. Topeka, KS: Washburn, University. The MorphoPASSE data used in this research was funded by National Institute of Justice Grant 2015- DN-BX-K014. Opinions or points of view expressed in this research represent a consensus of the authors and do not necessarily represent the official position or policies of the U.S. DOJ, NIJ, or the grant PI.

applied our simple naive Bayes method to that data, we achieved a success rate of 95.5%. For a more thorough evaluation however, it should be applied to other data than those used to assess the LRs.

The performance of a method producing LRs can be further optimized by adjusting the LRs afterwards, to improve their discrimination and calibration properties (Robertson et al. 2016, Ch 7.2). For example, we can optimize the above-mentioned success rate of 95.5% to 95.95% by simply multiplying all combined LRs with a factor of 0.6. Note that this success rate only looks at whether the LRs point towards male or female, and not at the strength with which they do so. It therefore only addresses the discrimination property. There are generic methods available to also improve the calibration property of the LRs (Zadora et al. 2014, p. 240 and references therein).

Finally, if conditional independence is not a reasonable assumption, then a Bayesian Network (BN) could be constructed to take the conditional dependencies into account (Taroni et al. 2006).

# Age estimation

Applying the Bayesian framework to the other elements of the biological profile (i.e. age, stature, and ancestry) is less straightforward than for sex estimation. This is mainly due to the fact that age, stature, and ancestry are non-binary, i.e. they are expressed as categorical or continuous values. Age, for instance, can be expressed in terms of age groups (categorical) or as a continuous age-at-death.

## Age estimation for investigative purposes

When operating in *investigative* mode (Jackson et al. 2016), the expert aims to provide information to guide the police investigation (see also Chapter 6.2). For age estimation, this means that the estimated age will be used to indicate the age (or age group) of the individual in order to narrow the list of potential sources of the skeletal material. Naturally, such estimation of age is not only dependent on the age-specific features of the remains. Even before the skeletal analysis, there is information available that has an effect on the probability of age-at-death of the remains. Many countries for instance have easily accessible basic information on the distribution of the general population over various age groups. This information for the Dutch population in 2019 is provided in Figure 1. Such information can be used to assign a prior probability of the remains belonging to someone from a specific age group.

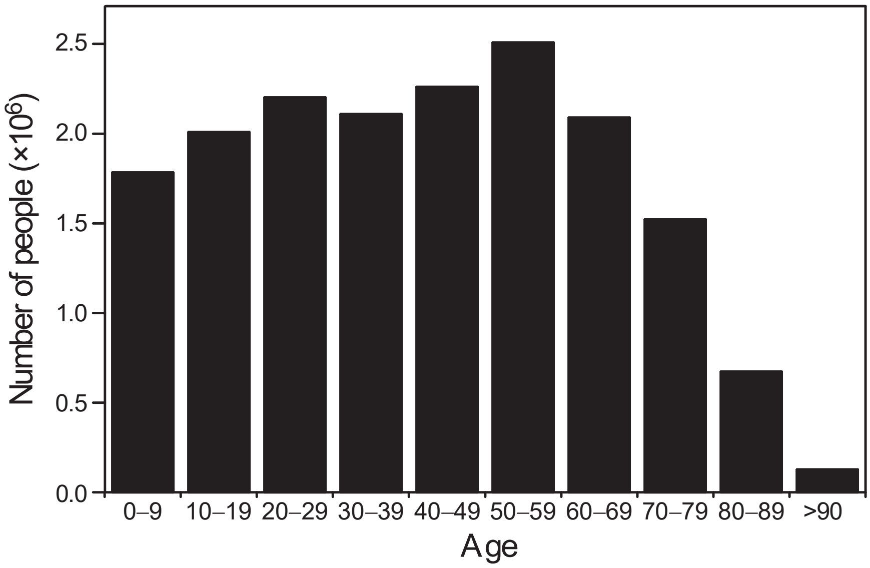


**Figure 1.** *Distribution of the total Dutch population by 10-year age cohorts in 2019, as registered by Statistics Netherlands (CBS, 2019).*

Before calculating such prior probabilities, one should consider the suitability of the reference data. For instance, the reference data from a living population (as provided in Figure 1) produce different prior probabilities compared to data containing the age-at-death of the individuals in the same population. The latter data may seem more suited for forensic anthropological investigation, since they originate from deceased individuals. But, depending on what is known about the case, these data may not be representative at all because they mostly relate to people dying of natural causes.

The circumstances of a case can significantly change our prior expectations for the age of the individual. If, for instance, it is thought that the remains belong to a mountaineer rather than a random member of the general population, this narrows the age distribution since the probability of very young and very old ages is reduced. Note that for assessing the circumstances of the case in this kind of investigative work, professional judgment and common sense are important. Even so, finding appropriate reference data to reflect the prior probability may not always be crucial, because if the anthropological evidence is rather strong it will lessen the influence of the prior distribution.

The combination of the prior probability and the evidence from the skeletal analysis narrows the age estimation, which can then be used to guide the investigation. As in sex estimation, we need relevant and mutually exclusive propositions. The age cohorts in Figure 1 could be used as basis for a series of proposition pairs, with every age cohort being used for H1, and the remaining age cohorts for H2. The prior odds (derived from the data in Figure 1) can be updated with evidence from the skeletal analysis.

In our example, we focus on age estimation on the basis of pubic symphysis as developed by Brooks and Suchey (1990) and revised by Hartnett (2010). This method differentiates seven phases, which can be used as evidence for the individual belonging to one of the age cohorts (Table 4).

Note that Table 4 provides the probability of age given the observed phase. To be able to assess the evidential strength of the features, however, we need the probability of the observed phase given the age. For this, we needed the original data of Hartnett (2010), which contained the observed phase for 619 individuals with known ages. Subsequently, the individuals were grouped in cohorts and the relative frequencies of the observed phases per age cohort were calculated. Table 5 shows the probability of the phases given the age cohorts.

**Table 4.** *Descriptive statistics for the revised pubic symphysis phase descriptions as adopted from Hartnett-Fulgitni (Hartnett 2010).*

| | Females | | | Males | | |
|---|---|---|---|---|---|---|
| **Phase** | **n** | **Mean age in years (SD)** | **Age Range (in years)** | **n** | **Mean age in years (SD)** | **Age Range (in years)** |
| **1** | 5 | 19.8 (1.33) | 18-22 | 14 | 19.29 (1.93) | 18-22 |
| **2** | 5 | 23.2 (2.38) | 20-25 | 14 | 22.14 (1.86) | 20-26 |
| **3** | 25 | 31.44 (5.12) | 24-44 | 36 | 29.53 (6.63) | 21-44 |
| **4** | 35 | 43.26 (6.12) | 33-58 | 69 | 42.54 (8.8) | 27-61 |
| **5** | 32 | 51.47 (3.94) | 44-60 | 90 | 53.87 (8.42) | 37-72 |
| **6** | 35 | 72.34 (7.36) | 56-86 | 34 | 63.76 (8.06) | 51-83 |
| **7** | 56 | 82.54 (7.41) | 62-99 | 96 | 77 (9.33) | 58-97 |

**Table 5.** *Summary of the observed phases for 619 individuals with known age (obtained from Kennyhercz et al., see Footnote 1). The table gives the probability of observing a phase given the cohort.*

| **Phase** | **10-19 years** | **20-29 years** | **30-39 years** | **40-49 years** | **50-59 years** | **60-69 years** | **70-79 years** | **80-89 years** | **90-99 years** |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.833 | 0.194 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **2** | 0.000 | 0.278 | 0.016 | 0.000 | 0.000 | 0.013 | 0.000 | 0.000 | 0.000 |
| **3** | 0.083 | 0.264 | 0.206 | 0.161 | 0.076 | 0.013 | 0.012 | 0.000 | 0.000 |
| **4** | 0.083 | 0.222 | 0.571 | 0.441 | 0.352 | 0.289 | 0.146 | 0.111 | 0.053 |
| **5** | 0.000 | 0.042 | 0.175 | 0.297 | 0.410 | 0.355 | 0.268 | 0.333 | 0.053 |
| **6** | 0.000 | 0.000 | 0.032 | 0.102 | 0.152 | 0.329 | 0.573 | 0.556 | 0.895 |
| **7** | 0.000 | 0.000 | 0.000 | 0.000 | 0.010 | 0.000 | 0.000 | 0.000 | 0.000 |

Now that we have assigned the probabilities of the observations given the propositions, we can apply the LR framework to update prior odds with evidential strength. We will calculate the prior odds, LRs and posterior odds with propositions based on the various cohorts.

With each cohort as a proposition for the age of the person, the prior odds for each cohort $C_i$ are given by dividing the number of individuals $N_i$ in the cohort $C_i$ by the number of individuals in other cohorts combined:

$$prior\ odds(C_i) = \frac{N_i}{\left[\sum_j N_j\right] - N_i}.$$

Observing a phase provides evidence for updating these prior odds. The evidence is evaluated by assessing the probability of the evidence if the person were in a specific cohort $C_\mathrm{i}$ (of which to update the odds), versus the probability of the evidence if the person were in any of the other cohorts. For the numerator of $LR_i$, Table 5 provides P(phase|$C_i$). For the denominator, we need to combine the probability of the observed phase for all the other cohorts. In this combination the relative importance of the various alternative cohorts is taken into account by weighting the probability of the evidence for the cohorts with the prior probability of those cohorts.

The strength of the phase evidence for every age cohort $C_\mathrm{i}$ is given as likelihood ratio ($LR_i$):

$$LR_i = \frac{P(phase|C_i)}{\left[\left(\sum_j P(phase|C_j)\cdot P(C_j)\right) - P(phase|C_i)\cdot P(C_i)\right] / \left[\left(\sum_j P(C_j)\right) - C_i\right]}.$$

Multiplying the prior odds for each cohort $C_\mathrm{i}$ with $LR_\mathrm{i}$ results in the posterior odds:

$$posterior\ odds(C_i|phase) = LR_i \times prior\ odds(C_i).$$

As such, depending on the strength of the evaluated evidence (the phase of the pubic symphysis) the probability distribution for the age cohort changes. Strong evidence narrows the probability distribution considerably, while weak evidence has a limited effect.

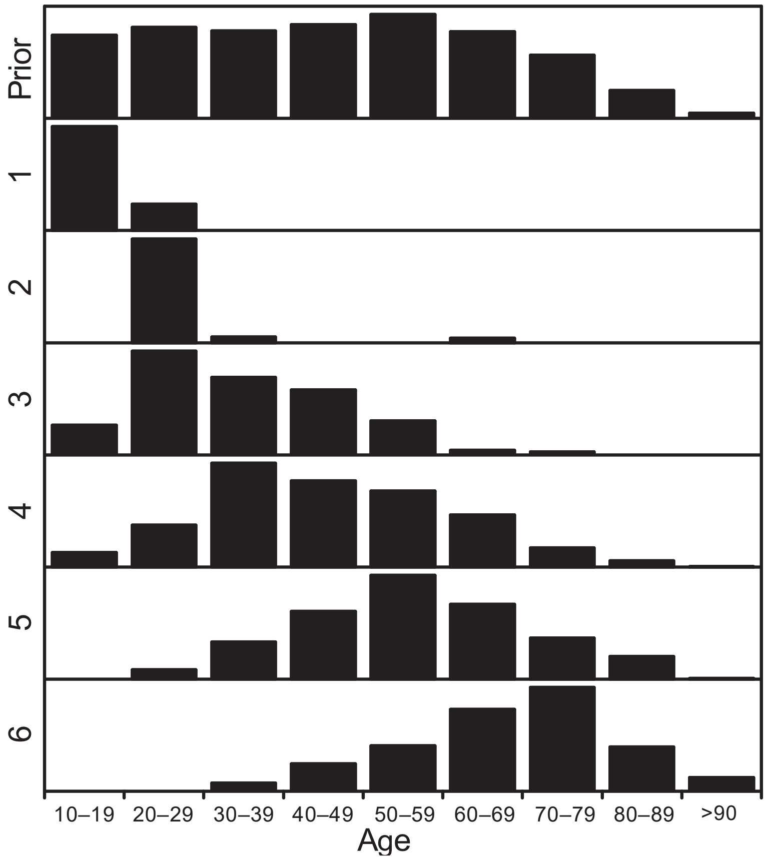


**Figure 2.** *Prior probability distribution based on the age cohorts of the Dutch population in 2019 (top). The graphs below show the posterior probability of the age cohorts of an individual after observing pubic symphysis phase 1 to 6 respectively.*

Figure 2 shows an example prior probability distribution for the unidentified individual belonging to a given age cohort based on age cohorts of the Dutch population in 2019, and updated probability distributions after observing pubic symphysis phase 1 to 6, respectively. It gives an overview of how observing the various phases narrows the probability distribution for the age cohorts.

It is clear that observing phases 1 or 2 provides strong evidence; the posterior probability of one (young) age cohort is relatively high, whereas the other age cohorts are far less likely or even excluded. In contrast, observing phase 3-6 provides relative weak evidence; the probability differences between the cohorts are far less than when observing phase 1 or 2. This finding is expected since younger individuals present larger changes in skeletal features and within a shorter time span than older individuals. It is for this reason that an accurate estimation of age at death becomes more difficult with increasing age.

## Age estimation for evaluative purposes

During the investigation, information may become available that suggests a particular identity of the remains, for example, Mrs. Johnson, aged 34 years. With this new information, we formulate a set of relevant and mutually exclusive propositions and the identification effort now becomes evaluative. The propositions are:

*H1: The remains belong to Mrs. Johnson.*

*H2: The remains belong to someone else.*

Please note that these propositions aim at identifying the remains by using age-related observations as evidence. To evaluate that evidence for these propositions, the age of the remains does not play a role in a direct way. This is why age is not an explicit element in the propositions.

Because we are operating in evaluative mode, we are not concerned with the prior odds for these propositions. Those prior odds would depend on all kinds of information about Mrs. Johnson, which has nothing to do with forensic anthropology. We only aim to report what the anthropological evidence means for the relevant propositions in the case, in other words, its evidential strength as expressed in an LR.

Suppose that pubic symphysis phase 4 was observed. To evaluate this evidence, we first assess the probability of observing phase 4 if indeed it was Mrs. Johnson (H1), who is in the

30-39 cohort. Table 5 gives a value of 57% for this probability. Next we assess the probability of observing phase 4 if the remains belong to someone else, from any of the cohorts (including that of Mrs. Johnson). We combine the probabilities of observing phase 4 for each cohort, weighting them with the prior probability of 'someone else' being in that cohort (again based on the Dutch population). This results in a probability of observing phase 4, given H2, of 27%.

The resulting likelihood ratio is therefore:

$$LR = \frac{P(phase\ 4|cohort\ Mrs.Johnson)}{\left(\sum_j\ P(phase\ 4|C_j)\cdot P(C_j)\right)/\sum_j\ P(C_j)} = \frac{0.57}{0.27} = 2.1.$$

This means that whatever the odds are that the remains belong to Mrs. Johnson as opposed to someone else, the anthropological observation doubles those odds. If the observation would have been phase 6, then the LR would be 0.20. In other words, observing phase 6 would decrease the odds of the remains belonging to Mrs. Johnson by a factor of 5.

It is also possible to make the alternative proposition more specific based on the case circumstances. What do the case circumstances tell us about the part of the population that the remains could belong to? For example, suppose that the remains are found in a desolate, mountainous area, only accessible to well-trained and self-reliant hikers[2]. The propositions may therefore be specified as:

*H1: The remains belong to Mrs. Johnson.*

*H2: The remains belong to another physically fit adult.*

[2] It is hoped that the examples given are sufficiently realistic, even if the name Johnson and mountainous areas are not particularly common in the Netherlands.

To calculate the probability of observing a phase given the alternative proposition, ‘a physically fit adult’ needs to be defined in terms of age. This could be done based on expert knowledge and reference data; e.g. assuming that a physically fit adult is between 18 and 60 years of age. This would change the weighting we applied previously in assessing the probability of the evidence given H2.

# Combining evidence

The use of likelihood ratios facilitates the combination of separate pieces of evidence in order to arrive at a single (combined) evidential strength for a pair of propositions. As such, the results of various age estimation techniques or the results of different elements of the biological profile (e.g. sex and age) can, in principle, be combined. Also, the forensic anthropological evidence can be combined with the results of other identification methods.

For a simple combination of evidence, a few prerequisites must be met. First, all LRs need to relate to the same propositions. This is one of the reasons why propositions should not depend on the specific scientific method applied. Second, if the items of evidence are independent, the prior odds can be subsequently updated by the first LR, and then by the second. In other words, the posterior odds for the first item of evidence become the prior odds for the second. The combined evidential strength is the multiplication of the separate evidential strengths (i.e., $LR_{tot} = LR_1 \times LR_2 \times LR_3 \times \ldots LR_i$). This approach can also be used if the pieces of evidence are *conditionally* independent. If the pieces of evidence are not (conditionally) independent, the combination of evidence is more complex. A Bayesian Network can then be used to combine all the available evidence.

Even if all the forensic anthropological evidence is eventually combined into one overarching LR, the forensic anthropologist should refrain from giving an opinion on the posterior probability of the propositions, since there may still be other, non-anthropological evidence. In this evaluative situation, the posterior odds are not up to the forensic anthropologist but up to the decision maker. Similarly, it is not the forensic anthropologist's task to decide on an identification, since this requires both the posterior odds and defining the threshold for identification, which depends on the costs and benefits of wrong and right decisions (Robertson et al. 2016, Ch. 6).

**Acknowledgement**

The authors of Kenyhercz et al. (2017) and Hartnett (2010) are gratefully acknowledged for sharing the raw data we requested.

# References

- Baliso A, Finaughty C, Gibbon VE, Identification of the Deceased: Use of Forensic Anthropology at Cape Town's Busiest Medico-legal Laboratory. FSI Reports (2019), 1, 100042. (DOI)
- Brooks S, Suchey JM, Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods, *Human evolution* 5.3 (1990): 227-238.
- Hartnett KM, Analysis of age-at-death estimation using data from a new, modern autopsy sample—part I: pubic bone, *Journal of forensic sciences* 55.5 (2010): 1145-1151.
- Jackson G, Aitken CGG, Roberts P, Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice. 4. Case Assessment and Interpretation of Expert Evidence. Royal Statistical Society's Working Group on Statistics and the Law, 2016.

- Kenyhercz MW, Klales AR, Stull KE, McCormick KA, Cole SJ, Worldwide population variation in pelvic sexual dimorphism: A validation and recalibration of the Klales et al. method, Forensic Sci Int. (2017), 277, 259.e1-259.e8. (DOI)
- Klales AR, Ousley SD, Vollner JM, A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods, Am J Phys Anthropol. (2012), 149, 104-114. (DOI)
- Phenice TW, A newly developed visual method of sexing the os pubis, Am J Phys Anthropol. (1969), 30, 297-301. (DOI)
- Robertson B, Vignaux GA, Berger CEH, Interpreting Evidence: Evaluating Forensic Science in the Courtroom. 2nd edition, 2016, Wiley. (DOI)
- Statistics Netherlands (CBS) https://opendata.cbs.nl/statline/#/CBS/nl/dataset/7461bev/table?ts=1573564373297, accessed December 10, 2019.
- Taroni F, Aitken CGG, Garbolino P, Biedermann A, Bayesian Networks and Probabilistic Inference in Forensic Science (John Wiley & Sons, 2006).
- Zadora G, Martyna A, Ramos D, Aitken CGG, Statistical Analysis in Forensic Science: Evidential Value of Multivariate Physicochemical Data (John Wiley & Sons, 2014).