

From Black, S. Aggrawal, A. & Payne-James, J. Age Estimation in the living, 2010, Wiley-Blackwell.

The presentation of results and statistics for legal purposes.

D. Lucy: Department of Mathematics & Statistics, Lancaster University, Lancaster, LA1 4YF, United Kingdom.

Introduction

In January of 2003 a young Turkish immigrant to Germany made his way into a Berlin woman's fitness centre, and shot three members of that centre's cleaning staff. The offender in this instance was a young man named Timucin 'O'. He alleged he had been paid the equivalent sum of £3,250 to kill the spouse of an associate. He also claimed he had killed the other two women in an attempt to cover his tracks by obscuring the motive for this particularly vindictive attack (Boyes, 2003).

Timucin 'O' was soon in custody, the trial being held in Berlin in July of 2003. He was originally to be tried as a full adult, but, during the course of the trial it emerged that he might have been under 21 years of age at the time the offence was committed. German law precludes trial as an adult for those under 21 years of age, and consequently a maximum sentence of 10 years for an offence of this nature. However, Timucin 'O's documents stated that he had been born on the 26th January 1981, which would have made him over 21 at the time of the offence. Timucin 'O's legal representatives claimed he had in fact been born on the 26th of January 1982, and that he was under 21 years at the time of the offence. The defence claimed that it is common for migrants to Germany to mislead the authorities about their age, and that records kept in Turkey would be too

poor to establish age reliably.

To establish Timucin 'O's age the defence took the unusual step of requesting a medical examination to see whether age related observations could be used to establish Timucin 'O's age, thus see whether he was under 21 years, or 21 years or over 21 years at the time of the offence.

This particular case is a specific instance where the pathologists skills, expertise and judgement will be challenged in ways which may be initially unforeseen to them. Uncertainty in the age related features imply that any judgement as to the age of Timucin 'O' will likewise be shrouded in uncertainty, and it is the quantification of that uncertainty with which statistical science concerns itself.

The question, in the case of Timucin 'O', revolved around whether he was under 21 years of age, at the time of the offence. This is a peculiarly precise question to which the pathologist must address themselves, but it is by no means the only question asked of pathologists and other investigators who are concerned with estimating the age of an individual from the physical manifestations of age. The question of the chronological age of an immigrant to a sovereign state upon entry to that state is of interest to authorities, as indeed it was in the case of Timucin 'O', and in previous chapters in this volume we have seen other similar critical uses for human age estimation. Participation in the correct age classes in junior sport, validation of age in adopted children, the falsification of age for preferential educational treatment are some examples. More upsetting, are cases involving the sexual abuse of children, the legal obligations of sovereign states to abandoned children, forced child labour, and the use of young men and women as soldiers.

The need for accurate and precise age estimates for individuals with no believable documented age is obvious, however, the importance of how accurate, and how finely resolved

an age estimate must be to fulfil its purpose depends upon that purpose, and requires some consideration.

In the case of Timucin 'O' a finely resolved, and accurate, age is of utmost importance. In other instances this may not be the case. If an individual is claiming a pension, and for the United Kingdom the pensionable age for men is 65 years of age, then a 60 year old claiming that they are 65 in order to claim their pension some five years early will not be of such great importance. This is because the size of the pension will depend upon the number of years of contribution to the pension fund. If, by claiming a premature age they claim their pension early any net financial advantage they may gain will be offset to a certain degree through the reduction in their final pension through the lower payouts. The same is true, to a lesser degree, for the component of their income provided by the state pension.

The same questions as to the accuracy and fine resolution can be asked of many of the social processes associated with immigration of individuals to any sovereign state which may be new to them. So long as an individual can be clearly placed into some category, or not claiming membership of some age-related category they are obviously not members of, then some fairly large grained, inaccurate, estimate of an individual's age will suffice for most purposes.

Even in fairly extreme cases of identity of living individuals an accurate, and narrowly defined knowledge of chronological age, may be desirable, but is not as critical to know as in the case of Timucin 'O'. On April the 7th 2005, officers from Kent Constabulary found a young man in a soaking wet suit wandering around the streets of Sheerness in England. He was named "Piano Man" after playing said instrument for staff at the chapel at Medway Maritime Hospital. The quality of the recital was debatable (BBC News; 2005, Moyes & Kaila; 2005). After some four months of baffling the British authorities, and a strenuous

search by Police and public throughout Europe, the young man was confirmed by the German authorities as 20 year old Andreas Grassl from Prosdorf in Eastern Bavaria. At no point did the fact that Grassl was 20 years of age become critical, however, the knowledge that he was a young man was important to his eventual identification.

Evidence and intelligence

Undoubtedly the quickest and most accurate method of age estimation is the location of some documentary evidence upon the individual for whom it is desired that chronological age be known. Unfortunately, a small portion of those for whom age needs to be known have concealed their true age and identity, so documentary methods tend to be prone to fraud and deception. A universal adoption of biometric based records may help to combat this in the medium to distant future, but in the short term accurate documentation which can be relied upon in all situations will not be available for all individuals.

In the absence of reliable documentary evidence physicians, dentists and other health workers can, on the whole, make extremely appropriately estimates of age, particularly if those persons for which age needs to be known are from a population with which the dentist, or health professional, happens to be familiar. A health professional using their expertise will make finely balanced judgements based upon experience with many hundreds, or even thousands, of individuals for which they do know the chronological age, and with whom they have intimate day to day contact. These experts will use many age related observables, some of which may be used only by that particular expert, and the expert may combine these observations to make some estimate for an age in a way which their experience best dictates. Anecdotal evidence suggests that an expert making a well informed estimate will often, within a large sample, outperform even the best of

anthropological measures.

However, what may be the most accurate estimate is not by necessity the best estimate to use in a legal context. In most jurisdictions the criminal and civil laws impose a burden of proof based upon evidence. Evidence is a form of information, but a very special form in that it is a public knowledge, or one which can be used to bear upon the truth, or otherwise, of some specific proposition in a very public way. This is still knowledge, but one which is distinct from a knowledge based upon instinct and unquantified experience which can be a form of private knowledge known only to the mind which apprehends it. Evidence must be fully explicable, that is it must be based upon observations which are open so that all may make those same observations. The years and years worth of accumulated observations which exist in the mind of an expert do not, as a rule, constitute a set of data which is open to the sort of examination which would satisfy others as to its veracity. So, despite potentially being a more accurate estimate, estimates made by the experience and intuition of experts do not satisfy the criterion of being publicly demonstrable.

So it is for the expert to somehow make an estimate based upon observations, and using a method, which can be used so others can come to the same, or very similar, conclusions as to an individual's chronological age. When seen with this light it is not so much a wonder that commonly used age estimations methods are imprecise, but that they manage to produce any estimates at all given they are based upon small, carefully measured, highly controlled training sets, and a very limited number of observables which are at best poorly related to the chronological age of the individual.

Statistical methods in age estimation

To satisfy the public nature of evidence some method by which a set of observations which are age related can be turned into an estimate of age is required, and because any estimate of age will have associated with it a level of uncertainty, those methods will inevitably be statistical, or statistically based, methods.

There are broadly two classes of approach to using more formal quantitative methods for the estimation of the age of a human from the observation of their physical characteristics. From a statistical point of view, there is nothing special about the case where an individual is alive, or whether only the mortal remains of the individual are open to inspection by the pathologist, or anthropologist. The only vital requirement is that there is some feature which can be observed which is related to age, and that for a given state of the feature the observations made by investigators will be the same, or different in some known way. The process of observation and inference is not, in principle, different between the deceased individuals seen by pathologists and physical anthropologists, and live individuals, although the actual features observed may be different.

The two sets of approaches, for want of better terminology, are:

1. classical approaches, sometimes termed frequentist approaches, is a view in which, amongst other tenants, a probability is seen as some function of ratios of frequencies. For frequentists the emphasis for inference tends to be on long runs of repeatable experiments (Efron: 2005).
2. Bayesian approaches by contrast tend to think of parameters as random variables in the same way that data are random variables. The emphasis for Bayesians is the estimation of probabilities, and probabilities tend to be expressions of degrees

of belief.

Few applied statisticians would consider themselves as working exclusively in either the two paradigms briefly characterised above. Most would tend to think in whatever manner best suited the nature of the immediate problem with which they were faced. There is a theorem called Bayes' theorem, which is the basis of any Bayesian approach. However, rather confusingly, frequentists may also use Bayes' theorem. This is because what is called Bayes' theorem is a small extension of the third law of probability, and, given the axioms of statistics, is equally true for both frequentist and Bayesian schools of thought.

Both approaches have been used by pathologists and forensic scientists for making estimates of age for humans for whom age was unknown.

Classical, or frequentist, approaches

If it has been observed that for a sample of individuals that some specific feature changes in some systematic manner with age, then by appealing to induction, one can say that due to some inherent biological similarity the wider population of humans will respond in a similar fashion. If that change happens to be on a continuous scale of variation, that is can, in principle at least, take on any value, and is not confined to groups, or to integer, or whole number, values, then some form of model may be considered an appropriate device by which to make estimates of the age for an individual of unknown age.

This process, at it's simplest, assumes a linear relationship between age and the observable quantity, and the parameters for the model itself can be calculated by the minimisation of errors in one, or the other. In which dimension errors should be minimised is a subject of debate (Aykroyd et al., 1997; Samworth & Gowland; 2007), but what is

less contentious is the form of any estimate of age made from such a model. Usually an estimate takes the form of a point, which represents the most likely age for the individual of unknown age, and an associated “confidence interval” which is some measure of the uncertainty associated with the estimate. However, it is the nature of the confidence interval which has a strict interpretation which makes such estimates, in many circumstances, unsuited to forensic contexts.

A confidence interval involves some level, usually denoted α , as being equal to some percentage, or proportion, score. For instance such an estimate, in the case of an age, may cite “an $\alpha = 0.95$ confidence interval of 30 to 50 years”, which a strict interpretation means that *in the long run, were the same experiment repeated many times, on 95% of those occasions the quantity of interest would fall within the confidence interval.*

To interpret a confidence interval as somehow standing in for a distribution representing uncertainty about the point estimate would be incorrect. It would, for instance, be a misinterpretation to take the 68% confidence interval and treat it as one of the parameters of a corresponding normal distribution, then use that normal to calculate arbitrary percentile probabilities for other quantities of interest. A confidence interval should not really be treated in that manner, although many scientists do.

Bayesian approaches

An alternative to ordinary least squares regression, and frequentist, based methods are Bayesian methods. These have been explicitly developed for population age profile estimation (Konigsberg & Frankenberg: 1992), and giving estimates of age for an individual (Lucy *et al.* 2002).

Bayesian methods calculate a posterior distribution for the quantity of interest given the particular states for some set of quantities which may be observed. Some workers feel uneasy about using Bayesian approaches as, to calculate the posterior distribution, some prior distribution is required. The notion of a prior distribution for some scientists can be seen as a weakness in the Bayesian approach, although few statisticians will regard it as such.

A Bayesian approach will take some prior distribution for the age of an individual for whom age is unknown, use observations of that individual, and other individuals who are of known age and thought to have some underlying biological similarity with the individual for whom age is unknown. Then the application of Bayes' theorem turns the observations, and the prior distribution into a posterior distribution. It is the posterior distribution which is a probability distribution, and can be regarded in the Bayesian paradigm as the degree of belief for the age for the unknown individual given the observations and prior belief.

The key to Bayesian approaches is the production of a posterior probability distribution for the age of interest. This is in fact what many scientists think they are calculating when they conduct some form of regression analysis and calculate confidence intervals. Interpreting a posterior probability distribution is unproblematic. If the probability density function can be defined simply, say for instance it is normal, then any percentile range can be expressed as an age range.

The relevance to age estimation

The discussion above should make it clear as to why statisticians have, especially in the last few years, been making many inroads into producing more and more Bayesian

approaches to the various problems with which they are presented. If one's purpose is to calculate the value for a parameter of a model involving some natural process, and one is basing the estimate upon data from a repeatable experiment, then a 95% confidence interval is a perfectly reasonable way to couch the uncertainty of the estimate. A trivial example might be human height with age in the prepubescent phase of life. Overall stature of the individual will increase with age at a given rate, and that rate would be a parameter which one could calculate from some least squares fit of height to age, where height for any given age for a cross section of the developing population, may be considered to vary with some known random function. From this model the gradient would be known with some confidence interval, and the researcher would know that when, for other samples from this same population, 95% of the point estimates for the gradient would fall within the 95% confidence interval calculated in the first instance, for that is what a 95% confidence interval is.

It is this feature of frequentist statistics which can make the paradigm simply unsuited to answering questions of evidence. The notion of a "long run" of identical, or nearly identical, experiments, has an appeal in sciences such as physics, where there are infinite potential observations to be made, or in chemistry where an experiment can be repeated as many times as the experimenter has patience to do so, but in the forensic sciences the idea of class of similar event has a more difficult meaning. In the context of an estimate made of age for an unknown individual there is precisely that, an individual. The experiment is not open to replication in the same way that the observation of the force of gravity is, or the products from a certain chemical reaction are. Even though upon each occasion an observation in a physical science is made there is a univocity, or ontological singularity (Deleuze; 1990, p177) in a very technical sense, the instances have sufficient similarities that for practical purposes they can be regarded as a class of events. The singularities between forensic events are much more difficult, if not impossible, to regard as

classes. For instance, Criminal courts are concerned with the specific and particularistic, not the general. This manifests itself in a disinterest in whether a particular suspect has been found to commit offences of a certain nature in previous instances, and in fact any such evidence would be dismissed as prejudicial, instead any criminal court would focus on *this* suspect, and *this* offence on *this* particular occasion.

A confidence interval may be appropriate for any intelligence role of age estimation, where for instance an investigator may simply need a rough idea of the age of an individual, but in a case such as that of Timucin “O” an answer involving a “long run of similar observations” is no answer to the question of whether Timucin O was over, or under, 21 years of age as the question is highly specific, and repeated experiments are out of the question.

Does a Bayesian estimate offer a better solution to the problem? The answer is a rather ambivalent ‘yes’ and ‘no’. Yes, because a posterior probability distribution given the observables, and the prior distribution, is a measure of belief to the specific question as to the age of the individual whose age is otherwise unknown. It has none of the baggage of repeatability of the confidence interval, however it does carry the requirement for a prior belief. A prior belief should not be seen as an insurmountable obstacle to the calculation of Bayesian estimates for age. It is relatively easy to show that in the case of some linear models the posterior estimate for the distribution for the age of an individual is normal, with parameters the same as the point estimate and confidence interval. The restriction is that the prior for the age distribution is the same as the age distribution for the training set from which the parameters for model were calculated. There are many instances in which frequentist estimates, and confidence intervals, can be given a Bayesian interpretation with certain strong assumptions about the prior distribution, and age estimates can usually be restated along these lines.

To show a Bayesian estimate has certain properties which may be more suited to making any specific age estimate for forensic purposes rather than an equivalent estimate made from a frequentist tradition is not the same as answering whether Bayesian solutions are appropriate solutions as this depends very much on the nature of the individual estimates.

Returning to the case of Timucin 'O', if one were requested to make an estimate for age in which all the probability was so confined that it fell between 19 and 20 years of age, then the estimate would answer the question as to whether Timucin 'O' was under 21 years of age at the time the offence was committed, and would be affirmative. Similarly were the posterior probability for age to be all above 25, then the question would be equally adequately answered, this time in the negative. However, an estimate of a 95% probability for Timucin 'O' of between 20 and 30 years of age doesn't really answer the question. One might use the posterior density function to calculate the relative posterior probabilities of him being under, and over, 21 years, then use that to calculate the posterior odds of him being under, or over, 21. This would be a form of answer which would encompass all the uncertainties within a publicly demonstrable form, the only disadvantage being that the prior distribution for age has to be agreed and stated.

Whether an age estimate answers the question or not can be seen to be dependent on the precision and accuracy of that age estimate. If, by some chance, the age estimates were highly precise and accurate, then from a practical point of view the exact nature of the treatment of the observations is immaterial. However, until some pathologist, or forensic anthropologist, makes some new discovery of an age marker which is highly correlated with chronological age, then forensic science will have to make the best of the somewhat vague age estimates which can be calculated from the current set of age related features.

likelihood ratio approaches

Part of the problem with the provision of an age estimate as an answer to some questions which might be posed in the course of a forensic investigation is that the age estimate itself verges, in some cases, on being a *non sequitur*. The legal question in the case of Timucin ‘O’ was “was this individual 21 years of age, or greater, or under 21 years of age at the time he committed the offence”. In some senses to answer “he was 20 to 30 years of age” fails to answer the legal question. If, as suggested above, one could answer “he was precisely 20 years of age”, or “he was above 30”, then the legal question would, by implication, be unambiguously settled. The salient point is that an extra inference is necessary to answer the legal question, but, in the case of an accurate and precise age estimate this would be an unproblematic inference.

The estimate of age given as some range, or point estimate with some indication of uncertainty, is an attempt to answer any universal question pertinent to the age of the individual, and, if precise enough for the purpose in hand, will do so without the fact finder making a fairly solid inference from the information presented by the expert. However, for certain forms of question, it is possible to give an answer to the legal question in a more direct way without having to provide an age estimate, or make any attempt to provide some universal answer to the question.

In recent years statisticians who work in forensic science have been looking closely at a quantity called a “likelihood ratio” as a measure of evidential value in forensic contexts (Aitken & Taroini: 2004). Although likelihood ratios are not new quantities, the uptake in their use over the last decade has been rapid. Likelihood ratios are, at the moment, seen as the measure of choice for evidential interpretation in DNA (Buckleton *et al.* 2005), glass (Curran *et al.* 2000), handwriting (Bozza *et al.* 2008), human identification (Steadman *et al.*, 2006), and more general multivariate matching problems (Aitken &

Lucy, 2004; Aitken *et al.*, 2007).

Likelihood ratios, in a simple form, are relatively easy to calculate, do not necessitate many prior assumptions, and answer questions involving dichotomous propositions in a direct way. The derivation and mathematical justification is usually approached from a Bayesian viewpoint, however, the use of a likelihood ratio, and its interpretation, are neither specifically Bayesian, nor frequentist. A likelihood ratio is:

$$\text{likelihood ratio} = \frac{\text{probability of the evidence given one proposition}}{\text{probability of the evidence given the other proposition}} \quad (1)$$

and it is probably better to explain what it is, and how to calculate and interpret it, by example. Chapter 10 of this volume refers to a pioneering paper in age estimation (Gustafson, 1950), and the appendix to this chapter has a table with a subset of two of the variables observed by Gustafson in that paper, and given as Table 3.

The terms in Equation 1 refer to probabilities given the truth of some propositions. In the case of Timucin ‘O’ the legal questions revolved around whether he was 21 years, or under, or over 21 years of age, so the propositions could be stated:

- Timucin ‘O’ was younger than 21 years at the time of the offence
- Timucin ‘O’ was 21 years, or older, at the time of the offence

These two propositions are exclusive and exhaustive in that they cover all possibilities for the age status of Timucin ‘O’, and they cannot both be true together. These propositions can be fitted into Equation 1, so that it becomes:

$$\text{likelihood ratio} = \frac{\text{probability of the evidence given Timucin ‘O’} \geq 21\text{years}}{\text{probability of the evidence given Timucin ‘O’} < 21\text{years}} \quad (2)$$

Which proposition appears on the numerator, and which the denominator is a matter of personal preference. It is customary to have propositions which suggest the prosecution

case as the numerator, here that is Timucin ‘O’ was 21 years, or greater, at the time of the offense, thus can be tried as an adult. Consequently propositions which suggest the defence case form the denominator, which here is that Timucin ‘O’ was under 21, and imply that he cannot be tried as an adult.

Next we need some evidence. Evidence can be, and in this case would be, a matter of simple observation, so suppose that an appropriate tooth was examined from Timucin ‘O’, and found to have a root dentine score, on the Gustafson scale, of zero. So the evidence, at the moment, could be said to comprise the observation that that Timucin ‘O’s root dentine translucency score was zero.

This new information can be inserted into Equation 2:

$$\text{likelihood ratio} = \frac{\text{translucency score} = 0 \text{ given Timucin 'O'} \geq 21\text{years}}{\text{translucency score} = 0 \text{ given Timucin 'O'} < 21\text{years}} \quad (3)$$

To put numbers into Equation 3 we need to appeal to data. Any appropriate database should suffice, for the purposes of the examples presented here those data come from Gustafson (1950). Here we do need two assumptions about the nature of the sample, and those individuals which comprise the sample:

1. the sample has observations made upon individuals who, using a form of the uniformitarian assumption, have some underlying biological similarity to the individual from whom the evidence has been extracted.
2. the sample has sufficient observations so the population frequency of the evidence can be estimated with sufficient accuracy from the data.

In this particular instance there may well be all sorts of reasons for doubting whether these two assumptions are true. The Scandinavian population upon which Gustafson made the original observations may have some fundamentally different rates of biological

change with age in the dental organs, and Gustafson’s sample was based upon only 41 teeth. However, as this example is purely illustrative we shall continue without addressing these potential shortcomings in the database.

Table 1: Tabulation of root dentine translucency score from Gustafson (1950), the data for which are given in Table 3. Here the rows represent those individuals who, in Gustafson’s sample were under 21, and 21 years, and over respectively. The columns correspond to the root dentine transparency scores. The numbers in the cells are the numbers of individuals from the age category with the given root dentine translucency state, so, for example, there are 7 individuals from Gustafson’s sample who were under 21 years of age, and who had a root dentine translucency score of 0, and 7 individuals who were 21, or over 21 years of age, and who had a root dentine score of 2.

age group	transparency score		
	0	1	2
age < 21	7	0	0
age \geq 21	8	19	7

Estimation of the two probabilities in Equation 3 from Gustafson’s data can be made from the frequency table. Table 1 gives the numbers of individuals in each age category from Gustafson’s sample, with each stage of root dentine translucency. The numerator from Equation 3 calls for the probability of observing Timucin ‘O’s’ root dentine translucency score as zero, were it true that Timucin ‘O’ was 21 years, or of an age greater than 21 years. To find this from Table 1 we focus upon those who are at least 21, that is the bottom row of data, and see that there are $8 + 19 + 7 = 34$ individuals who were 21, or over. Of these 8 had a root dentine translucency score of 0. Were the requirements for the sample met, that is: Timucin ‘O’ has some form of underlying biological similarity to the sample, and that the observations from the sample sufficient in both quality, and quantity, to estimate the relevant probability, then the probability of making the observation of Timucin ‘O’s’ tooth having a root dentine translucency score of 0 could be said to be $8/34$.

The denominator in Equation 3 requires the probability of observing Timucin ‘O’s’ root

dentine translucency score of 0 were it true that Timucin ‘O’ were truly less than 21 years. Making the same assumptions about the relevance of Gustafson’s observations as for the numerator, we can say that, of 7 individuals from the data who were under 21, 7 had root dentine translucency scores of zero. An estimate for the probability of observing a root dentine translucency score of zero were it true that Timucin ‘O’ were truly under 21 years of age could be said to be $7/7 = 1$.

Inserting these two probabilities into Equation 3:

$$\text{likelihood ratio} = \frac{8/34}{1} = \frac{8}{34} \quad (4)$$

Which, in this instance, gives a likelihood ratio of $8/34$, or, $\approx 1/4$. The interpretation of this figure is quite straightforward and uncomplicated, and could be stated: “the observation of the root dentine translucency is $1/4$ times as likely were Timucin ‘O’ 21 years of age, or over, than were Timucin ‘O’ under 21 years”. It is quite legitimate to invert this statement to: “the observation of the root dentine translucency is 4 times as likely were Timucin ‘O’ under 21 years of age, than were he 21 years, or over, thus these observations offer limited support for the proposition that Timucin ‘O’ is under 21”.

Many observations which can be made of the human to establish chronological age occur in a multivariate case, and Gustafson’s (1950) observations were no different[§]. Table 2 has data from Gustafson (1950) for the joint occurrences of attrition *and* root dentine translucency for those under 21, and those 21 or over.

[§]Gustafson (1950) handled the multivariate nature of the observations by using an unweighted sum to reduce six observations from each tooth to a single observation. This made a great deal of sense in the light of Gustafson’s subsequent use of a linear model, and produced a simplification which led to a highly manageable method of age estimation.

Table 2: Tabulation of root dentine translucency score and attrition score from Gustafson (1950), the data for which are given in Table 3. Here the left hand table represents those individuals who, in Gustafson’s sample under 21 years, and the right hand table those 21 years or over. The columns correspond to the root dentine transparency scores. The rows the attrition scores. The numbers in the cells are the numbers of individuals from the age category with the given root dentine translucency state and the given attrition score, so, for example, there are 6 individuals from Gustafson’s sample who were under 21 years of age, and who had a root dentine translucency score of 0, and attrition score of 0. There were 7 individuals who were 21 years of age, or over, and who had a root dentine translucency score of 1 and an attrition score of 1.

attrition	age < 21			age ≥ 21		
	translucency	translucency	translucency	translucency	translucency	translucency
	0	1	2	0	1	2
0	6	0	0	3	1	1
1	1	0	0	4	7	3
2	0	0	0	1	10	3
3	0	0	0	0	1	0

Were the evidence from Timucin ‘O’ to consist of a root dentine translucency score of 0 *and* an attrition score of 1, then Equation 3 can be rewritten to take account of the additional information:

$$\text{likelihood ratio} = \frac{\text{translucency} = 0 \text{ and attrition} = 1 \text{ given Timucin 'O'} \geq 21 \text{ years}}{\text{translucency} = 0 \text{ and attrition} = 1 \text{ given Timucin 'O'} < 21 \text{ years}} \quad (5)$$

Making the same assumptions, and working in the same way as in the univariate case, the numerator in the can be estimated by observing that there are 4 individuals who had a translucency score of 0 *and* attrition score of 1, of the 34 individuals who were over 21 years of age. The numerator can therefore be written as 4/34. The denominator in likelihood ratio can be estimated by observing that there is only a single individual who has a translucency score of 0, and attrition score of 1, from the 7 individuals from the sample who were under 21 years. This gives an estimate for the denominator of 1/7. The likelihood ratio is the quotient of these, so:

$$\text{likelihood ratio} = \frac{4/34}{1/7} = \frac{28}{34} \quad (6)$$

and can be stated “the observation of an attrition score of 1, *and* a translucency score of 0

is $\approx 7/9$ times as likely were Timucin ‘O’ 21, or greater than 21, as were Timucin ‘O’ less than 21 years of age. Or, again it is perfectly legitimate to restate the inverse to frame the likelihood ratio in support of the defence case, and would read “the observations of a translucency score of 0, *and* an attrition score of 1 offer about 1.25 as much support for Timucin ‘O’ being under 21, than 21, or older than 21, thus these observations offer very limited support for the proposition that Timucin ‘O’ is under 21”.

The advantages for the forensic scientist of employing this sort of statement of the evidential worth, or value, of the observations, are not immediately apparent, but:

1. the scientist can confine their statements quite specifically to their observations, and the support those observations give to the various propositions, rather than forcing the scientist to make a comment upon the relative merits of the propositions. The truth, or otherwise, of a proposition is usually the province of the fact finders, juries and judges and the like, in most jurisdictions. A case may comprise of many elements, and much evidence. The scientist examines only the evidence in which they have some special expertise, and it is inappropriate, and premature, to comment upon the whole case from a single piece of evidence.
2. It allows the scientist to approach the question in hand directly, rather than giving an answer which requires further inference on the part of the fact finder to make the information provided by the scientist relevant to the question asked by the court.
3. Each and every assumption is clear and explicit. In this way a likelihood ratio easily satisfies every criteria for a “publicly demonstrable” knowledge, making it a way in which a forensic scientist can safely provide a court with the full facts in as compact a form as possible.
4. The framework is simple, and can act as a template within which many problems of forensic evidence may be hung. However complicated any single probability calcu-

lation may be, the framework requires only the calculation of only two conditional probabilities, which may be estimated by any legitimate means.

Chapter 1 of this volume points the sorts of questions with which the medical examiner or forensic scientist may be faced when approached to comment on the age of an individual. Many of these questions revolve around the ages of legal maturity, and can best be answered in any specific case by a likelihood ratio. However, not all questions of age can be addressed this way. Those questions tend to be where, for some form of record keeping, the age itself is required. So for passports an age estimate may be the best solution, however, for some document which is age dependent, such as a driving licence, the question may be “is the individual qualified by age to hold some licence”, then the report of a likelihood ratio may be the best solution.

Errors of interpretation

It would be all too easy to use some table such as Table 1 to find a posterior probability for the age group of any individual given the age related observations. So, for the first example Timucin ‘O’ was said to have a root dentine translucency score of 0. From Table 1 there are $7 + 8 = 15$ individuals who have a root dentine translucency score of 0, 7 of whom were under 21 years. A statement which could be made might be, given the same assumptions about the sample as before, is “the probability that some individual is under 21 years, given the observation of a root dentine translucency score of 0, is $7/15$, or just under half”. A consequence is that the probability that Timucin ‘O’ is 21 years, or over, given the observation of root dentine translucency, is just over half. These two would, on the face of it, offer support for the proposition that Timucin ‘O’ was 21, or over 21, however the forensic scientist should be on their guard for this sort of evaluation

of the evidence.

The probability of 7/15 for Timucin 'O' being under 21 years, is the posterior probability of age *given* the observations. It is a Bayesian estimate for the posterior probability for him being under 21 years, but it is easy to show that this Bayesian estimate has within it an implied prior probability for Timucin 'O' being under 21 years, which is equal to the probability of observing an individual from the sample who is under 21 years. As there are 41 teeth in the sample, 7 of which came from individuals who were under 21, this would make the prior probability of Timucin 'O' 7/41, which is quite a strong prior, and would require quite a lot of data in favour of the younger age group to make the posterior probability support that case.

It is this sort of confusion which modern statistical evaluation of forensic evidence hopes to avoid by using likelihood ratio based measures of evidential value.

Another error which scientists can be all too tempted by is to confuse the likelihood ratio for a posterior probability ratio, so, some statement such as: "the observations of a translucency score of 0, *and* an attrition score of 1 offer about 1.25 as much support for Timucin 'O' being under 21, than 21 or older" gets confused for a comment such as "it is 1.25 times as probable that Timucin 'O' was under 21, than 21 or over". This statement has left out the important clause about the ratio being about observations, rather than propositions. This sort of error is called a "transposed conditional", and is where the probability of observations given propositions is mistaken for the probabilities of propositions given observations. It is also sometimes referred to as "the prosecutors fallacy" because often it takes the form of a low probability for some evidence given the innocence of some suspect is mistaken for a low probability of innocence.

Evetts (1995) suggests that all statements of probability should include the terms "were",

or “given” to indicate precisely what the conditioning is for any probability, and forensic scientists can thus avoid making fundamental, and sometimes dangerously so, errors of interpretation and evaluation.

Concluding comments

Were it the case that pathologists, physical anthropologists, and forensic scientists, could assign an accurate and highly precise chronological age on the basis of some physical, or other features, for human beings, then statistical calibration could largely be neglected. Unfortunately this is not the case, and until some near miraculous set of age related features is discovered, age estimates will always have some degree of uncertainty associated with them, and that uncertainty will, for the foreseeable future, be large.

Traditional frequentist methods of calibration have a precise interpretation involving repeatability and samples from “long runs” of identical events which, although appropriate in many areas of science, can conflict with the ontological singularities which exist at a fundamental level in the forensic sciences. To circumvent these difficulties it is recommended that Bayesian estimates are used whenever possible which have an interpretation which invokes a “degree of belief” given the observations.

An estimate for age, followed by some interval which describes the uncertainty for that estimate can be seen as a universal solution for questions surrounding age, however, it is not always the best solution in any given case and specific question within a legal framework. Many of the more critical questions set by case work can be dichotomised into sets of mutually exclusive and exhaustive propositions. Questions such as how old a young female found working in a brothel in London really is, from the legal view, more questions of whether the young female is younger than, or older than, some arbitrarily

defined chronological age stated in law. These questions can best be answered, and presented in court, as a likelihood ratio gives a measure of the strength of support of evidence for any given proposition without the forensic scientist having to comment upon their opinions upon the propositions themselves. Steadman *et al.*, although more interested in human identification, age being part of identification, point to how likelihood ratios can be used with continuous data, rather than the ordinal data illustrated in this chapter.

Age estimation will always require data upon which to base estimates of probabilities, and/or parameters. These data should be conditioned correctly for the exact circumstances for the questions under consideration. For instance, any calibrated value of Timucin 'O's chronological age, or estimate of a likelihood ratio for age class, should be based in so far as possible upon samples of humans drawn from those individuals of known age thought to somehow resemble him at some basic biological level. Many of the critical ages with which the law in various jurisdictions operates are at the younger end of the spectrum of human age, and it is this end which, from Chapter 7 of this volume, is most prone to systematic change through secular trends in the way in which humans mature. This will be a significant factor in human age estimation regardless of the way in which an age is represented to the fact finder, and of the purpose to which the age estimate is to be put. However, it is unlikely to prove a major problem so long as pathologists and forensic scientists are aware of the magnitude of the secular trends and account for them by either modelling them, or using data from a temporally relevant sample.

Timucin 'O' was found to be guilty by a court in Munster, and was given a life sentence.

References

Aitken, C.G.G. & Lucy, D. (2004) Evaluation of trace evidence in the form of multivariate data, *Applied Statistics*, **53**, 109-122.

Aitken, C.G.G. & Taroni, F. (2004) *Statistics and the Evaluation of Evidence for Forensic Scientists*, Wiley.

Aitken, C.G.G. Zadora, G & Lucy, D. (2007) A Two-Level Model for Evidence Evaluation, *Journal of Forensic Sciences*, **52**(3), 412-419.

Aykroyd, R.G. Lucy, D. Pollard & A.M Solheim, T. (1997) Technical note: Regression analysis in adult age estimation. *American Journal of Physical Anthropology*. **104**(2); 259-265.

BBC News (2005) Fantastic response to “piano man”, 16th May 2005.

BBC News (2005) “Piano man” flies back to Germany, 22nd August 2005.

Boyes, R. (2003) Murderer says he is too young to grow old in jail, *The Times*, 8th July.

Bozza, S. Taroni, F. Marquis, R. & Schmittbuhl. M. (2008) Probabilistic evaluation of handwriting evidence: likelihood ratio for authorship. *Applied Statistics*, **57**, 329-341.

Buckleton, J. Triggs, C.M. & Walsh, S.J. (eds) (2005) *Forensic DNA evidence interpretation*, CRC Press.

Curran, J.M. Hicks, T, N. & Buckleton, J.S. (2000) *Forensic Interpretation of Glass Evidence*. CRC Press.

Deleuze, G. (1990) *The Logic of Sense*. Continuum International Publishing Group.

Efron, B. (2005) Bayesians, frequentists, and scientists. *Journal of the American Statistical Association*, **100**(464), 15.

Evetts, I.W. (1995) Avoiding the transposed conditional, *Science & Justice*, **35**(2), 127-131.

Gustafson, G. (1950) Age determination on teeth. *Journal of the American Dental Association*, **41**, 45-54.

Konigsberg, L.W., & Frankenberg, S.R. (1992). Estimation of age structure in anthro-

pological demography. *American Journal of Physical Anthropology*. **89**, 235-256.

Lucy, D. Aykroyd, R.G. & Pollard, A.M. (2002) Nonparametric calibration for age estimation. *Applied Statistics*, **51**(2); 183-196.

Moyes, S. & Kaila, J. (2005) Exclusive: piano man sham, *The Mirror*, 22nd August.

Samworth, R. J. & Gowland, R. L. (2007), Estimation of adult skeletal age-at-death: statistical assumptions and applications, *International Journal of Osteoarchaeology*, **17**, 174-188.

Steadman, D.W. Adams, B.J. & Konigsberg, L.W., (2006) Statistical basis for positive identification in forensic anthropology, *American Journal of Physical Anthropology*, **131**(1), 15-26.

A Gustafson's data

Table 3: Age related data from Gustafson(1950). These observations were made from extracted teeth, and are stages, or points, awarded to predefined stages in the development of quantities whose underlying distribution is continuous. As such, despite being from extracted organs, the discretised nature of these observations are similar to many of the age related features used by pathologists and forensic scientists to make estimates of chronological age. Gustafson originally looked at six features from these data, here, for the purposes of illustration, we will only examine occlusal attrition and root dentine translucency.

age	attrition	translucency	age	attrition	translucency
11	0	0	48	1	1
12	0	0	48	0	1
12	0	0	49	1	2
13	0	0	49	1	2
15	0	0	50	2	2
16	0	1	51	2	2
17	0	0	51	2	1
23	0	0	51	1	2
23	0	1	52	1	2
25	0	0	52	1	1
28	0	1	52	2	1
35	1	2	53	2	0
37	0	1	55	1	1
37	0	0	55	2	2
38	1	2	59	1	2
38	1	1	64	1	2
39	1	1	64	1	0
39	0	2	65	1	2
45	1	2	69	1	3
45	1	1	69	1	1
48	2	1			