Two Biostatistics Seminars

David Balding

Schools of BioSciences and of Maths & Stats University of Melbourne.

Vic Biostat seminar, 26 February 2015

Seminar 1: How to evaluate the probability that bones found in a carpark are from a specified dead king

For further details see:

King TE et al. Identification of the remains of King Richard III. *Nat. Commun.* **5**:5631 doi: 10.1038/ncomms6631 (2014).

King Richard III

- Last king of England to die in battle, at Bosworth field, in 1485, aged 32. This ended the 300-year rule of the Plantagenets, replaced by the Tudors.
- Accounts of his death imply that his skeleton would show substantial signs of injury.
- Richard's remains were brought back to Leicester and buried in the choir of the church of the Grey Friars.
- Friary dissolved in 1538, most buildings torn down soon after and their exact locations were lost.
- ▶ 125 years later a rumour arose that he had been disinterred following the dissolution, and thrown into the river Soar.
- This account was no longer widely believed and recently an archaeological dig was undertaken to seek his remains at the presumed site of the friary.



- Richard was described by contemporaries as having slim build and one shoulder higher than the other.
- One of the earliest portraits of RIII, held by the Society of Antiquaries of London. It has not been affected by significant overpainting, and is thought to be one of only two portraits painted during his lifetime.

日本《檀本《国本《国本

Skeleton 1

In September 2012, Skeleton 1 was excavated at the presumed site of the friary, whose appearance was consistent with the remains being those of RIII.

- The skeleton was that of a male aged 30 to 34 years,
 - with severe scoliosis which would have rendered one shoulder higher than the other, and
 - with indications of numerous perimortem battle injuries.
- radiocarbon dating gave a 95% interval of 1456 to 1530, which overlaps his lifespan (1452 - 1485).



How to evaluate weight of evidence for S1 to be RIII?

◆□▶ ◆□▶ ◆目▶ ◆目▶ 目 のへぐ

How to evaluate weight of evidence for S1 to be RIII?

Likelihood ratio:
$$LR = \frac{P(\text{evidence}|H1)}{P(\text{evidence}|H2)}$$

where

$$H1 =$$
 Skeleton 1 is RIII
 $H2 =$ not $H1$

H1 is s simple hypothesis, but many alternatives are grouped together under H2.

- Of particular interest is the alternative that S1 was a matrilineal relative of RIII (within a few tens of mother-child links). Then matching mtDNA sequences are likely even though H1 is false.
- 81 contemporary, male, matrilineal relatives of RIII were excluded from having participated in the Battle of Bosworth, no relevant record was found for one other.

Background information and evidence

We take as time point for the distinction between background information that informs the prior probability of H1, and evidence that is explicitly evaluated in the LR, the moment that Skeleton 1 was uncovered in Sept 2012, and recognised to be human but no further details were yet noted. So

- location and nature of the grave are background information, along with historical documents and relevant scientific facts (DNA mutation rates etc),
- signs of disease and wounds on Skeleton 1, as well as sex and age at death, are evidence.

The radiocarbon dating



Radiocarbon Date Density

The isotope analysis revealed high levels of seafood in the diet of Skeleton 1, and so a compromise between marine and terrestrial calibration curves was used to obtain a probability distribution for date of formation of the bones.

- Probability mass assigned to the lifespan of RIII: 0.19.
- Probability mass assigned to the lifespan of the friary: 1.00.

Under H1, we assumed that the (mean) radio carbon date was U(1452.76, 1485.64) – the lifespan of RIII. So:

L(x|H1) = 1/32.9 if $x \in (1452.76, 1485.64)$ and 0 otherwise.

Under H2, we chose U(1227, 1538) – the friary lifespan. So:

L(x|H2) = 1/311 if $x \in (1227, 1538)$ and 0 otherwise.

Then

$$LR = \frac{0.19/32.9}{1.00/311} = 1.84,$$

corresponding to limited support for H1.

Osteoarchaeological analysis:

Skeleton 1 was of a male aged late 20s to early 30s. Under H1: L(age, sex|H1) = 0.95, allowing for some inaccuracy in the technique.

Under *H*2: From 706 skeletons with age and sex assignments at Grey Friars and two similar priories, 126 were found to be male and in the age class 26 to 35.

 For all count data we used pseudo-counts to bias low relative frequencies upward;

So we used L(age, sex|H2) = 127/708 and so

$$LR = \frac{0.95}{127/708} = 5.3,$$

again corresponding to limited support for H1.

Scoliosis

Skeleton 1 had severe idiopathic adolescent-onset scoliosis, which is consistent with the asymmetric shoulder observation. We identified two other medical conditions that could explain the observation: Erb's Palsy and Sprengel's deformity.

Using current UK data for the latter 2 conditions + observation of 5 cases of scoliosis among 1476 UK skeletons, we obtained:

L(Scoliosis|H1, asymmetric shoulders) = 0.9

= Scoliosis rate (5/1476) divided by the sum of the 3 rates.

We multiplied this by 0.95 to allow for the possibility that the report that RIII had asymmetric shoulders was incorrect.

Under H2, we used the (biased) scoliosis fraction (5+1)/(1476+2):

$$LR = \frac{0.95 \times 0.90}{6/1478} = 212,$$

corresponding to moderately strong support for H_{1} , H_{2} ,

Wounds

Skeleton 1 had 11 perimortal wounds; two under the base of the skull would have been fatal.

- These are consistent with accounts of RIII's death; we assigned L(Wounds|H1) = 0.9 to allow for possible exaggeration in these reports.
- Under H2, we identified 1 skeleton with comparable wounds among 91 in the choirs of 8 priories active in a similar period.
 - only used priory choirs, which are prestige locations; much additional data available for other priory/church sites.

These lead to:

$$LR = \frac{0.9}{2/93} = 42,$$

corresponding to moderate support for H1.

The Skeleton 1 Y haplotype doesn't match any of the 5 presumed patrilineal relatives of RIII (according to *Burke's Peerage*).

- ► under H1, none of these 5 can be a true patrilineal relative, and ≥ 2 false paternity events (FPE) have occurred.
- all 5 are presumed descendants of the C18 Duke of Beaufort, who is apparently a 15th-generation descendant of Edward III;
- RIII is apparently a 4th generation descendant of EIII. The FPE required under H1 could have occurred in either lineage.
- 4 of the 5 share a Y haplotype, which is presumably that of the Duke of Beaufort. The other must have resulted from another FPE, among the 22 father-son transmissions in the lineages descending from the Duke.

Patrilineal descendants of Edward III



The Y haplotypes

and subgroups Richard III Som 1,2,4,5 Som 3 R1a1 I2a2 5 22 5 U, Ξ z á ш M285 M222 M253 P287 P214 U198 M153 M167 S145 U152 U106 S116 L11 M269 M198 M173 M242 M231 M304 M9, P128, P131, P132 M201 M170 M96 M89

R1b1b

FPE rates

There have been many published estimates of FPE rates,

- none over time periods relevant here or among the aristocracy;
- grounds for suspicion about many of these estimates;
- we chose the lowest estimate among those we identified: a study reporting 8 FPEs in 936 putative father-son links; to these we added the 1 FPE among presumed descendants of the Duke of B.

Thus our FPE rate was (8+1)/(936+19) = 9/955 (no pseudo-counts because we are working under H1). Then

▶ L(Y data $|H1) = 1 - (1 - 9/955)^{19} \times$ Pr(observed Y hap)

2nd term = L(Y data | H2), so cancels in the LR, leaving

$$LR = 1 - (1 - 9/955)^{19} = 0.16$$

which corresponds to limited support against H1.

Wendy Duldig (ML2)

(1452-1485) Richard III Katherine Crathorne (c.1555-1605) Frances Winstanley (c.1703-1766) Sabrielle Whitehorn (1928-2004) Frances Wentworth (1631-1693) Jorothy Grantham (1659-1717) Everhilda Constable (c.1535-?) Frances Gardiner (1828-1907) Harriet Villebois (1774-1821) Marjorie Moore (1891-1954) Frances Truman (1726-1801) Harriet Plunkett (1807-1864) Sophia Lysaght (1861-1945) Everhilda Maltby (1605-?) Frances Read (1750-1820) ⁻rances Holt (1681-1771) Everhilda Creyke (?-?) 3arbara Gough Calthorpe (1746-1826) Charlotte Vansittart Neale (1817-1881) Charlotte Vansittart Frere (1846-1916) Catherine Manners (c.1510-c.1547) Margaret Babthorpe (c.1550-1628) Barbara Constable (c.1530-c.1561) Barbara Cholmley (c.1575-1618) Barbara Yelverton (c.1692-1724) Barbara Calthorpe (c.1716-1782) Barbara Belasyse (1609-1641) Barbara Talbot (1665-1763) Muriel Stokes (1884-1961) Anne St Leger (1476-1526) Barbara Slingsby (1633-?) Ann Spooner (1780-1873) Joy M Brown (1926-2008) Michael Ibsen Anne of York (1439-1476) (ML1)

Richard Plantagenet (1411-1460), Duke of York = Cecily Neville (1415-1495)

Matrilineal descendants of Anne of York (sister of RIII)

C

Mitochondrial DNA sequences



Summary of mtDNA sequencing results:

Michael Ibsen: full mtDNA sequence match with Skeleton 1.

Wendy Duldig: differed at 1 site.

Question: Given the Michael Ibsen evidence, how strong is the additional support given to *H*1 by the Wendy Duldig evidence? *Answer:*

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

Question: Given the Michael Ibsen evidence, how strong is the additional support given to H1 by the Wendy Duldig evidence? *Answer:* Nil.

- Given MI data, WD data have the same probability whatever the identity of Skeleton 1.
- ▶ We can ignore the WD data when calculating *LR*.
- So the effort to identify and sequence WD was of little value given that MI fully matched S1,
 - but it could have been useful if there had been a different outcome from MI.

MtDNA mutation

Under H_1 , the full matching of MI with S1 is somewhat unlikely because at least one mutation is plausible over 19 generations.

- mtDNA has a much higher mutation rate than nuclear DNA,
- but it is variable across sites,
- few whole-mtDNA-genome mutation rate estimates.
- DNA mutation rates are being revised downward because of ancient DNA sequencing results, but for our purposes higher estimates are conservative (do not favour H1).
- We used a high estimate based on 10 control region mutations in 327 generations using genealogical data.
- although only based on mtDNA control region, higher point estimate than recent whole-genome estimates based on ancient sequences.

 $L(\text{mtDNA sequence match of MI with } S1|H1) = (1-11/329)^{19} = 0.52.$

Under H_2 , we require the population fraction of the S1 mtDNA sequence. This is not as easy as it seems:

- ▶ We want the C15 population fraction, not contemporary
 - probably not a big problem.
- We can have bigger sample sizes if we restrict attention to the mtDNA control region, as whole-genome sequencing is still relatively novel. That implies neglecting much information in the full-sequence data.
- We can have bigger sample sizes if we use European data rather than English data;
 - ► mtDNA maternally inherited ⇒ possible geographical clustering of mtDNA types,
 - but the mobility of the female nobility was apparently high.

We chose the most conservative combination of

- only mtDNA control region, not full sequence
- only English and not European data.
- We observed 0 copies of the S1 sequence among 1823 sequences.
 - (European data: 0 observations from 26 127).

So these data suggest that the observed sequence is very rare

but we know it exists in MI, and consequently we expect a large number of copies of the sequence among his (possibly very many) matrilineal relatives.

Adding the sequence of MI to the English database, and using the usual pseudo-counts:

L(mtDNA sequence match of MI with S1|H2) = 2/1826,

$$LR = \frac{0.52}{2/1826} = 478$$

 \Rightarrow moderately strong support for H1.

Evidence not directly used in the calculation

- Isotope analysis: Skeleton 1 lived primarily in the North-West of England and enjoyed a high-status diet rich in seafood.
- Genotypes at hair and eye colour loci: IrisPlex and HIrisplex results were that Skeleton 1 had a 96% probability of blue eyes and a 77% probability of blond hair. Consistent with the SAL portrait (above), but substantial variation possible:



Assuming the distinct types of evidence are mutually independent:

 $LR = 1.84 \times 5.3 \times 212 \times 42 \times 0.16 \times 478 = 6.7$ million.

Extremely strong support for H1.

- ▶ DNA data only: $LR = 0.16 \times 478 = 79$. Moderate support.
- Using all the evidence except mtDNA (appropriate if alternative is that S1 is a matrilineal relative of RIII) LR = 14000. Very strong support for H1.

Prior?

- The archaeologists were sufficiently convinced of finding RIII that they invested considerable resources in the dig.
- grave was in the expected location and at about the right depth; absence of grave goods as expected.

So a reasonable prior could not be very small, we suggest $1/40 \Rightarrow$ **POSTERIOR PROBABILITY THAT S1 = RIII IS 0.999994**. There were many authors of the King et al (2014) paper responsible for various aspects of the research outlined above.

The probability calculation was performed in collaboration with Mark Thomas (UCL), with help on DNA data from Michael Hofreiter (York) and Walther Parson (Penn State). The team was led by Turi King (Leicester).

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <