

RADIOCARBON DATING OF THE TURIN SHROUD: NEW EVIDENCE FROM RAW DATA*

T. CASABIANCA† 

Ajaccio 20000, France

E. MARINELLI

Collegamento pro Sindone, Rome, Italy

G. PERNAGALLO 

Department of Economics and Business, University of Catania, Corso Italia 55, 95129 Catania CT, Italy

and B. TORRISI 

Department of Economics and Business, University of Catania, Corso Italia 55, 95129 Catania CT, Italy

In 1988, three laboratories performed a radiocarbon analysis of the Turin Shroud. The results, which were centralized by the British Museum and published in Nature in 1989, provided ‘conclusive evidence’ of the medieval origin of the artefact. However, the raw data were never released by the institutions. In 2017, in response to a legal request, all raw data kept by the British Museum were made accessible. A statistical analysis of the Nature article and the raw data strongly suggests that homogeneity is lacking in the data and that the procedure should be reconsidered.

KEYWORDS: ANOVA, HYPOTHESIS TESTING, OXCAL, TURIN SHROUD, WARD AND WILSON TEST

INTRODUCTION

The Turin Shroud (TS) is a piece of linen cloth approximately 4.4 m long and 1.1 m wide, considered by some to be the burial cloth of Jesus of Nazareth. The first certain historical record dates to the second half of the 14th century AD. In 1987, after a decade of negotiations, three laboratories (Arizona, Oxford and Zürich) were chosen by the Vatican authorities to perform a ^{14}C test using accelerator mass spectrometry techniques (AMS). On 21 April 1988, a sample was taken from one corner of the cloth, and pieces of the sample were delivered to the laboratories, along with three additional control samples. The results were centralized at the British Museum, where the statistical analysis was performed. In their *Nature* article, Damon *et al.* (1989) stated that there was ‘conclusive evidence’ that the linen of the TS was medieval (AD 1260–1390 with at least 95% confidence).

*Received 24 April 2018; accepted 15 February 2019

†Corresponding author: email tristancasabianca@yahoo.fr

© 2019 University of Oxford

Currently, non-specialists consider this general conclusion straightforward disproof of the hypothesis that the TS is an antique linen cloth (Ball 2017). However, since 2005, a growing number of studies have provided elements and arguments contradicting the medieval hypothesis (Rogers 2005; Poulle 2009; Fanti and Malfi 2014; Bevilacqua *et al.* 2014; Boi 2017; Casabianca 2017). In 2013, a new statistical study based on a regression analysis seriously called into question the reliability of the conclusions of the 1988 dating (Riani *et al.* 2013). Recently, we obtained the raw data and, for the first time, measured their convergence with the radiocarbon dates published in *Nature*.

The basic idea of this paper is to conduct a robust statistical analysis using the raw data and to compare these results with those obtained using the official data. We use statistical tests to determine whether the measurements *intra* and *inter* laboratories can be combined to obtain a calendar range. Based on the statistical results, we question the level of confidence of 95% attributed to the AD 1260–1390 calendar age range.

DATA AND METHODOLOGY

Since 1989, scholars interested in the TS have often requested the raw data from the laboratories, without success (Pourrat 1991). This denial of access led to controversy and hindered the assessment of the analysis performed by the British Museum (Marinelli 2012; Rinaldi 2012). In 2017, several Freedom of Information Act (FOIA) requests to the institutions involved in the TS radiocarbon dating were made. The British Museum replied favourably and made all its files, ‘not dated or arranged in any order’, available. On 18 July 2017, the British Museum electronically sent a file of 211 pages and in September 2017, one of the authors (T.C.) visited the British Museum and scanned more than 500 new pages.

We learned that Zürich performed 4×10 measurements for each of their five TS subsamples. Oxford performed five measurements, and while the detailed measurements were not sent to the British Museum, a mean was provided for two measurements, resulting in only three radiocarbon dates. The Arizona report shows that 40 measurements (5×8) were made. Several scholars already disclosed the eight Arizona radiocarbon dates (Van Haelst 2002), but these dates have never been confirmed. We are now certain that the eight disclosed dates (designated Arizona raw 2) were correct. However, previously, Arizona modified the errors of two of its eight dates (designated Arizona raw 1). In contrast to Zürich and Oxford, Arizona included in its report eight computer printouts along with the counts of the detected radiocarbon atoms (designated Arizona counts).

The *Nature* and raw radiocarbon dates (Table 1) show that relevant differences exist among the estimates provided by the three laboratories: the Arizona minimum *Nature* estimate is 591, whereas the Oxford minimum is 730; the Arizona maximum is 701, whereas the Oxford maximum is 795. Most importantly, discrepancies exist even within the laboratories. For example, Arizona’s estimates range from 540 to 701, whereas Oxford’s estimates reach a value of 795. The computed error in the radiocarbon dates by Arizona presents differences (the error ranges from 30 to 41 in *Nature* and from 37 to 57 in the raw dates), whereas Oxford appears to have produced consistent estimates (homogenous radiocarbon estimates between *Nature* and the raw dates), but the computed errors in the raw radiocarbon dates are lower than the published errors (the most relevant is 730 ± 30 in the raw data against 730 ± 45 in *Nature*). Regarding Zürich, the problem is related to both the estimates and computed errors: the estimates are incongruent between the raw and *Nature* dates (a clear difference in the fifth observation, which is equal to 595 in the raw data and 679 in *Nature*). From a statistical perspective, these differences do not

Table 1 Radiocarbon dates before present (1950) of the Turin Shroud (\pm error)

<i>Arizona Raw 1</i>	<i>Arizona Raw 2</i>	<i>Arizona Nature</i>	<i>Oxford Raw</i>	<i>Oxford Nature</i>	<i>Zürich Raw</i>	<i>Zürich Nature</i>
606 \pm 41	606 \pm 41		795 \pm 53	795 \pm 65	733 \pm 61	733 \pm 61
574 \pm 45	574 \pm 45	591 \pm 30	730 \pm 30	730 \pm 45	722 \pm 56	722 \pm 56
753 \pm 51	753 \pm 51		745 \pm 46	745 \pm 55	635 \pm 57	635 \pm 57
632 \pm 49	632 \pm 49	690 \pm 35			617 \pm 47	639 \pm 45
676 \pm 40	676 \pm 59				595 \pm 46	679 \pm 51
540 \pm 37	540 \pm 57	606 \pm 41				
701 \pm 47	701 \pm 47					
701 \pm 47	701 \pm 47	701 \pm 33				

Sources: Damon *et al.* (1989); Arizona, Oxford and Zürich reports to the British Museum (FOIA 2017).

create a problem if they are irrelevant, but the data published in Damon *et al.* (1989) are affected by several problems (Brunati 1996; Van Haelst 1997, 2002; Riani *et al.* 2013).

The Arizona counts table (Table 2) represents the eight sessions (A1, A2, A3, ...) included in the report sent by Arizona to the British Museum. Two sessions were performed daily using the same standards (A1 and A2, A3 and A4, A5 and A6, and A7 and A8). Due to a technical incident encountered in the two first measurements of A2, the Arizona printout of A2 only includes four values, one of which (17 584) is the equivalent of two measurements. For the purpose of the analysis, we chose to keep five values for each session and we attributed the mean value (17 584/2 = 8792, in italics in Table 2) to the two problematic measurements.

The methodology adopted in this study is classical. A central argument proposed by critics of the TS radiocarbon dating results is that the computation of the *p*-value using the method in Ward and Wilson (1978) contains an error (Brunati 1996; Van Haelst 1997). The Ward and Wilson method is used to test whether the radiocarbon determinations can be combined (Ward and Wilson 1978, 23). Following Van Haelst (1997), we present the results of the chi-square tests using the Ward and Wilson method to detect statistically significant inconsistencies in the raw radiocarbon dates and we perform ANOVA, parametric and non-parametric tests to determine possible significant differences between means in the raw data to eventually validate the results. To test the eventual existence of an interaction effect between the types of data (raw vs. *Nature*) and

Table 2 Counts of radiocarbon atoms in the Turin Shroud subsamples by Arizona

<i>A1</i>	<i>A2</i>	<i>A3</i>	<i>A4</i>	<i>A5</i>	<i>A6</i>	<i>A7</i>	<i>A8</i>
8 226	8 792	8 315	9 812	10 030	10 502	5 128	6 838
9 033	8 792	7 637	10 020	9 450	11 123	6 319	6 841
9 033	8 767	7 762	9 927	9 363	11 335	7 744	7 089
9 238	8 750	7 637	10 010	9 444	11 409	8 132	7 191
9 206	8 843	7 432	9 288	9 201	11 327	8 287	7 882

Source: Arizona report sent to the British Museum (FOIA 2017, 110–7).

the laboratory that produced the data (Arizona, Oxford or Zürich), we perform a two-way ANOVA. Hereafter, a 5% level of significance is assumed for all procedures.

Finally, due to the current importance of radiocarbon procedures, we present results obtained with the software OxCal version 4.3 (OxCal 2018), which was used here to statistically analyse specific versions of the radiocarbon dates. OxCal is an open-access diagnostic tool based on a Markov chain Monte Carlo developed by Christopher Ramsey (Oxford) in the mid-1990s and is currently used by many radiocarbon specialists (Ramsey and Lee 2013). The individual agreement, model agreement and overall agreement indexes should generally all be over 60% in the case of a homogeneous result. Several individual measurements below the 60% level combined with model and overall agreement indexes below 60% indicate the presence of at least one problem in the data, probably due to a flawed measurement or contamination.

RESULTS

The ANOVA on *Nature* radiocarbon dates (Table S2) shows that at least two groups' means are significantly different from each other, providing evidence that the differences among the laboratories are not irrelevant, especially for the Arizona/Oxford *Nature* dates (Table S3). The Ward and Wilson test is also above the critical value ($8,60 > 5,99$). The OxCal 4.3 software confirms these results with an overall agreement index of 41.8%, strongly below the usual threshold of 60%. On 12 radiocarbon dates, three have a low individual agreement index (Table 3 for the Ward and Wilson tests and the OxCal analysis).

Figure 1 shows a box plot representation of both the *Nature* and raw radiocarbon dates. Since the columns of the raw radiocarbon dates provided by Arizona are only distinguished by the computed errors, we just include one box plot for the Arizona raw radiocarbon dates (measured age), whereas the estimates from Oxford are equal for both the raw and the *Nature* dates (except for the related errors). The box plots show that the raw and *Nature* radiocarbon dates are highly similar when referring to the same laboratory. However, notable differences can be observed among raw dates with different provenances. Furthermore, the Arizona raw radiocarbon dates show a prominent range, and the distribution appears to be slightly skewed (skewness = -0.13), whereas the skewness of the Zürich and Oxford raw data is evident (respectively, 0.40 and 1.36). The range is higher and the skewness more apparent in the raw radiocarbon dates than in *Nature*.

Table 3 Ward and Wilson chi-square test and OxCal 4.3 analysis for the TS radiocarbon dates

	Ward and Wilson test (critical value in brackets)	OxCal 4.3 overall agreement index (number of individual dates below 60% in brackets)
Arizona <i>Nature</i> vs. Oxford	8.60	41.8%
<i>Nature</i> vs. Zürich <i>Nature</i>	(5.99 for 3–1 df)	(3/12)
Arizona Raw 1 vs. Oxford	10.75	18.1%
<i>Nature</i> vs. Zürich <i>Nature</i>	(5.99 for 3–1 df)	(6/16)
Arizona Raw 2 vs. Oxford	8.55	28.4%
<i>Nature</i> vs. Zürich <i>Nature</i>	(5.99 for 3–1 df)	(5/16)
Arizona Raw 1	19.24	21.4%
	(14.07 for 8–1 df)	(2/8)
Arizona Raw 2	14.45	34.6%
	(14.07 for 8–1 df)	(2/8)

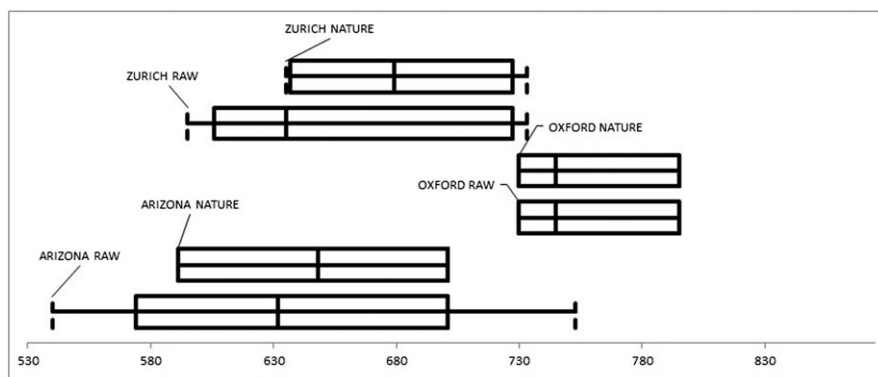


Figure 1 Box plots based on Table 1.

An understanding of the relationship that exists between the raw and the official radiocarbon dates is important. To consider the simultaneous effect on the years BP estimates of the type of data (*Nature* report data or raw radiocarbon dates) and the laboratory that produced the estimates (Arizona, Oxford or Zürich), we performed a two-way unbalanced ANOVA (via a regression; Table 4).

The test of Factor A (variable on rows) and the interaction effect failed, whereas it was significant for Factor B (variable on columns); that is, the laboratory in which the assessment was performed is relevant in explaining the observed differences.

To isolate the effect of the eventual significant differences among the raw dates, we replicate the ANOVA only for raw data, using Arizona Raw 1, Oxford Raw and Zürich Raw.

The results of the ANOVA of the raw radiocarbon dates indicate that the null hypothesis cannot be rejected (p -value=0.0749, Table S6). Therefore, the observed differences are not statistically significant (at least at 5% significance). The Ward and Wilson chi-square for Arizona Raw 1, Oxford Raw and Zürich Raw, using the corresponding errors, indicates that the estimates are not consistent (15.99, with a critical value of 5.99). Thus, the three raw assessments given by each laboratory are, with 95% of confidence, representative of the same phenomenon, but the age estimated by the laboratories is not consistent.

We focused our attention on the Arizona raw radiocarbon dates and compared these results with the results reported by Zürich and Oxford in Damon *et al.* (1989). With the Ward and Wilson test (Table 3), significant differences exist among the three laboratories when we compare

Table 4 Two-way ANOVA (via regression, $\alpha = 0.05$)

	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i> -value	<i>sig</i>
Rows	286.5473	1	286.5473	0.0841	0.7745	No
Columns	47 367.44	2	23 683.72	6.9527	0.0046	Yes
Inter	740.575	2	370.2875	0.1087	0.8975	No
Within	74 940.61	22	3 406.391			
Total	124 740.1	27	4 620.004			

the radiocarbon dates of Oxford *Nature*, Zürich *Nature* and Arizona Raw 1 ($10.75 > 5.99$), and the radiocarbon dates of Oxford *Nature*, Zürich *Nature* and Arizona Raw 2 ($8.55 > 5.99$).

When we use OxCal to analyse the radiocarbon dates of Arizona Raw 2, Oxford *Nature* and Zürich *Nature*, we notice that five of the 16 dates are below the threshold of 60%. The overall agreement index is 28.4%. With Arizona Raw 1, Oxford *Nature* and Zürich *Nature*, the results are even less satisfying (overall agreement index = 18.1% and six of the 16 dates below the threshold).

The same rationale applies to the intra-laboratory differences. We also computed the Ward and Wilson test for the raw radiocarbon dates of Arizona, and in both cases (raw 1 and raw 2), the null hypothesis was rejected. Using OxCal for Arizona Raw 2, the overall agreement index (34.6%) is below the threshold (with 12.8% for A3 and 43.0% for A6), whereas for Arizona Raw 1 the overall agreement index is lower (21.4%). Based on these results, a relevant problem emerges in the consistency between the Arizona raw radiocarbon dates and the published results from the other laboratories.

The analysis of the Arizona counts showed further interesting aspects. The eight counts of the Arizona data were categorized into four groups (A1 and A2, A3 and A4, A5 and A6, and A7 and A8) because they were executed on the same day using the same standards. The non-parametric Kruskal–Wallis test (Table 5; see also Table S10 for the assumptions) shows highly statistically significant differences even if we consider the eight counts both separately and gathered (p -values < 0.0001).

The results show that the different assessments produced by the same laboratory (raw vs. *Nature*) are not statistically significant, whereas the analysis of the raw radiocarbon dates confirmed that the different laboratories produced different assessments and that these differences are, in most cases, statistically significant.

DISCUSSION

The conclusions of the various statistical methods applied to both the *Nature* and raw results *intra* and *inter* laboratories are concordant. Even when we apply the Bonferroni correction, a very conservative choice that increases the probability of type II errors (Perneger 1998), the results of Arizona Raw 1 and Arizona counts are still significant (Box S1).

Some raw data are missing, even if they do not have a noticeable impact on our overall conclusion. The aggregation of the five Oxford measurements to obtain the three Oxford raw and Oxford *Nature* results ($2+2+1$) is probably in favour of a homogenization of the radiocarbon dates. Notably, as pointed out by Bray in his report to the British Museum dated 27 September 1988, the enlargement of the Oxford computed errors is probably the consequence of a more

Table 5 The Kruskal–Wallis test for the Arizona counts

Statistics	Arizona, four gathered counts	Arizona, eight counts
H-stat	25.7020	35.5932
H-ties	25.7092	35.6032
df	3	7
p -value	<0.0001	<0.0001
α	0.05	0.05
sig	Yes	Yes

inductive than deductive evaluation process (Turin Shroud Archive). The error of the TS subsample O1.2b has been changed from 730 ± 30 to 730 ± 45 , whereas Damon *et al.* (1989, 613) affirm that 'Oxford errors below 40 are rounded up to 40'. Two changes were made to 50 years, one for the subsample O3.2b of control sample 3 (from 1990 ± 33 to 1990 ± 50) and the other for the subsample O4.2u of control sample 4, from 785 ± 35 to 785 ± 50 (Table S11 presents the raw radiocarbon dates of the control samples).

On 31 August 1988, Zürich explained to the British Museum its change regarding the 'current dependent effect' (FOIA 2017, 136). Nevertheless, the conditions of this adjustment remained unclear to Arizona. In a letter to the British Museum dated 8 September 1988, Donahue and Damon affirmed that they 'do not understand how such a systematic calculational error could have changed the values of their uncertainties' (Turin Shroud Archive).

After the change in two uncertainties from Arizona Raw 1 to Arizona Raw 2, the chi-square becomes almost acceptable (Table 3). This adjustment is unusual, since none of the radiocarbon dates of the control samples were modified by Arizona. In its report, Arizona explained that

[for] these results the average of the ratios of OXII/OXI standards was different from more than two standard deviations from its current value. (The average differed from the correct value by 1.0%.) An additional error of $\pm 0.5\%$ was added quadratically to the regular standard deviation of these measurements.

The two modified radiocarbon dates, which were achieved using the same standards, were clearly not identical within errors. After the change, they appear compatible (Table S12).

Our statistical results do not imply that the medieval hypothesis of the age of the tested sample should be ruled out. The two extreme *Nature* radiocarbon dates (591 ± 30 and 795 ± 65) were proposed as outliers, yielding a calibrated calendar age range of AD 1281–1302 with at least 95% of certainty (Christen 1994; Christen and Pérez 2009). But, as emphasized by Ramsey *et al.* (2010, 959), 'the inclusion and exclusion of outliers should not be seen as a black box to cover up poorly understood problems in the data'.

Each TS raw and published radiocarbon date indicates a medieval interval for the fabric. Nevertheless, this reasoning would simply assume a constant amount of ^{14}C atoms among the subsamples. This basic assumption is not supported by the heterogeneity of the TS raw data, the consistent ages of the control samples, the significant statistical trend in the TS radiocarbon dates and the amount of foreign material found by the laboratories.

The hypothesis of a statistical significance only due to some difference in measurements among the laboratories is weakened by the fact that the results were correct and consistent for the three control samples (Damon *et al.* 1989; Riani *et al.* 2013). In addition, for Arizona and Zürich, Egyptian linen samples were tested and not mentioned in Damon *et al.* (1989). Moreover, our statistical analysis of the raw data supports the conclusion of Riani *et al.* (2013). They used the known locations of the tested samples in each laboratory and showed a significant decrease in the radiocarbon age as one gets closer to the centre of the sheet (in length, from the tested corner). This variability of the *Nature* radiocarbon dates in a few centimetres, if linearly extrapolated to the opposite side of the TS, would lead to a dating in the future. Many hypotheses, not mutually exclusive, have been proposed to explain this lack of homogeneity (Marinelli 2012). The environment and the history of the sample could be part of the explanation. Despite the close visual inspection of the TS by textile experts and the loss of weight of approximately 25% after the cutting (FOIA 2017, 162), Oxford found and removed several textile fibres of different colours, including one identified by a textile laboratory to be cotton, 'possibly of Egyptian origin and quite old' (Anonymous 1988; FOIA 2017, 104). Oxford mentions that in one subsample there may have been 'glass', perhaps sodium

chloride crystals (Wilson 1995, 18; FOIA 2017, 103). In the original draft, Arizona indicated that ‘a red thread and three blue threads’ were removed from one of their subsamples (Turin Shroud Archive). In 2010, Arizona recognized that they had kept one piece of an undated TS subsample. On this subsample, the authors identified foreign material invisible to the naked eye, including a blue substance described as ‘apparently wax’ (Freer-Waters and Jull 2010, 1522) and some cotton fibres. Zürich may have found an assortment of debris (Marinelli 2012, 26).

The protocol finally adopted in the 1988 TS test was heavily criticized at the time. Harry E. Gove, the inventor of the AMS method, affirmed that ‘[if] one of the three laboratories obtained an outlier result [...] it would be impossible statistically to identify it and the three measurements would all have to be included in the average thereby producing an incorrect result’ (Gove 1989, 237). Our statistical analysis confirms that this criticism was warranted. The documentation provides a practical example of complex procedures that vary for each radiocarbon laboratory and illustrates the way in which they can affect the reliability of a calendar age interval. As suggested by one of the referees, there is a question as to whether the results are combined to give a weighted average or calibrated individually. This will be the subject of further research.

The discussed statistical analysis reinforced the argument against the goodness of the radiocarbon dating of the TS, suggesting the presence of serious incongruities among the raw measurements. Our results, which are compatible with those previously reported by many other authors (Brunati 1996; Van Haelst 1997, 2002; Riani *et al.* 2013), strongly suggest that homogeneity is lacking in the data. The measurements made by the three laboratories on the TS sample suffer from a lack of precision which seriously affects the reliability of the 95% AD 1260–1390 interval. The statistical analyses, supported by the foreign material found by the laboratories, show the necessity of a new radiocarbon dating to compute a new reliable interval. This new test requires, in an interdisciplinary research, a robust protocol. Without this re-analysis, it is not possible to affirm that the 1988 radiocarbon dating offers ‘conclusive evidence’ that the calendar age range is accurate and representative of the whole cloth.

REFERENCES

- Anonymous [World News Network], 1988, Rogue fibres found in the Shroud, *Textile Horizons*, **8**(12), 13, available at <http://www.sindone.info/TEXTILE.PDF>.
- Ball, P., 2017, Material witness: Is this holy relic preserved? *Nature Materials*, **16**, 503.
- Bevilacqua, M., Fanti, G., D’Arienzo, M., and De Caro, R., 2014, Do we really need new medical information about the Turin Shroud? *Injury*, **45**, 460–4.
- Boi, M., 2017, Pollen on the Shroud of Turin: the probable trace left by anointing and embalming, *Archaeometry*, **59**, 316–30.
- Brunati, E., 1996, La corrispondenza con ‘radiocarbon’ sulla datazione della Sindone, available at <http://www.sindone.info/BRUNATI2>
- Casabianca, T., 2017, Turin Shroud, resurrection and science: one view of the cathedral, *New Blackfriars*, **98**, 709–21.
- Christen, J., and Pérez, S., 2009, A new robust statistical model for radiocarbon data, *Radiocarbon*, **51**, 1047–59.
- Christen, J. A., 1994, Summarizing a set of radiocarbon determinations: a robust approach, *Applied Statistics*, **43**, 489–503.
- Damon, P. E., Donahue, D. J., Gore, B. H., Hatheway, A. L., Jull, A. J. T., Linick, T. W., Sercel, P. J., Toolin, L. J., Bronk, C. R., Hall, E. T., Hedges, R. E. M., Housley, R., Law, I. A., Perry, C., Bonani, G., Trumbore, S., Woelfli, W., Ambers, J. C., Bowman, S. G. E., Leese, M. N., and Tite, M. S., 1989, Radiocarbon dating of the Shroud of Turin, *Nature*, **337**, 611–5.
- Fanti, G., and Malfi, P., 2014, Multi-parametric micro-mechanical dating of single-fibers coming from ancient flax textiles, *Textile Research Journal*, **84**, 714–27.
- FOIA (Freedom of Information Act), 2017, Freedom of Information request on the Turin Shroud to the British Museum, The British Museum to T.C. (electronic communication), 18 July, PDF file, 211 pp.
- Freer-Waters, R. A., and Jull, T. A. J., 2010, Investigating a dated piece of the Shroud of Turin, *Radiocarbon*, **52**, 1521–7.
- Gove, H. E., 1989, Letter to the editor: The Turin Shroud, *Archaeometry*, **31**, 235–7.

- Marinelli, E., 2012, The setting for the radiocarbon dating of the Shroud, in 1st International Congress on the Holy Shroud, Valencia, Spain, available at <http://www.sindone.info/VALENC-3.PDF>
- OxCal version 4.3, 2018, <https://c14.arch.ox.ac.uk/oxcal.html> (accessed 20 March 2018).
- Perneger, T., 1998, What's wrong with Bonferroni adjustments, *BMJ*, **316**, 1236–8.
- Poulle, E., 2009, Les sources de l'histoire du linceul de Turin, revue critique, *Revue d'histoire ecclésiastique*, **104**, 747–82.
- Pourrat, O., 1991, Shroud dating still questioned, *Nature*, **349**, 558.
- Ramsey, C. B., and Lee, S., 2013, Recent and planned developments in the program OxCal, *Radiocarbon*, **55**, 720–30.
- Ramsey, C. B., Dee, M., Lee, S., Nakagawa, T., and Staff, R. A., 2010, Developments in the calibration and modeling of radiocarbon dates, *Radiocarbon*, **52**, 953–61.
- Riani, M., Atkinson, A. C., Fanti, G., and Crosilla, F., 2013, Regression analysis with partially labelled regressors: Carbon dating of the Shroud of Turin, *Statistics and Computing*, **23**, 551–61.
- Rinaldi, G. M., 2012, La statistica della datazione della sindone, available at http://sindone.weebly.com/uploads/1/2/2/0/1220953/nature_statistica.pdf
- Rogers, R., 2005, Studies on the radiocarbon sample from the Shroud of Turin, *Thermochimica Acta*, **425**, 189–94.
- Turin Shroud Archive [uncatalogued documentation], Conservation and Science Department, British Museum, London.
- Van Haelst, R., 1997, Radiocarbon dating the Shroud: a critical statistical analysis, available at <https://www.shroud.com/vanhels3.htm>
- Van Haelst, R., 2002, Radiocarbon dating of the Shroud of Turin: a critical review of the nature report (authored by Damon *et al.*) with a complete unbiased statistical analysis, available at <http://www.sindone.info/VHAELST6.PDF>
- Ward, G. K., and Wilson, S. R., 1978, Procedure for comparing and combining radiocarbon age determinations: a critique, *Archaeometry*, **20**, 19–31.
- Wilson, I., 1995, From a forgotten memorandum: A visit to the Oxford Research Laboratory 7 July 1988, *British Society for the Turin Shroud Newsletter*, **41**, 15–8, available at <https://www.shroud.com/pdfs/n41part4.pdf>.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

- Table S1. Shapiro–Wilk test (a) and Levene's and Bartlett's tests (b) (*Nature* radiocarbon dates).
- Table S2. ANOVA (*Nature* radiocarbon dates).
- Table S3. T-tests (a) and Mann–Whitney tests (b) (*Nature* radiocarbon dates).
- Table S4. Shapiro–Wilk (a) and Levene's tests (b) (raw radiocarbon dates).
- Table S5. Shapiro–Wilk (a) and Levene's and Bartlett's tests (b) for a two-way ANOVA.
- Table S6. ANOVA (raw radiocarbon dates).
- Table S7. Mann–Whitney tests (intra-laboratory data: *Nature* vs raw radiocarbon dates).
- Table S8. Mann–Whitney tests (comparison among laboratories using raw radiocarbon dates).
- Table S9. T-tests for mean differences (*Nature* radiocarbon dates).
- Table S10. Shapiro–Wilk test (a) and Levene's tests (b) (Arizona gathered counts).
- Table S11. Control samples (raw radiocarbon dates). Legend: (a) average (Zürich report); (c) weighted mean corrected (Arizona report). Source: Oxford, Arizona and Zürich reports (FOIA, 2017).
- Table S12. Compatibility between two radiocarbon dates (Arizona Raw 1 and Arizona Raw 2).
- Box S1. Bonferroni correction.