Advances in Social Sciences Research Journal – Vol.4, No.13
Publication Date: July. 25, 2017
DoI:10.14738/assrj.413.3418.
Arachige, D. (2017). Revisiting the Analysis of 1988 Radiocarbon Dating of the Shroud of Turin. *Advances in Social Sciences Research Journal, 4(13)* 16-23.

Revisiting the Analysis of 1988 Radiocarbon Dating of the Shroud of Turin

Darshi Arachige
Independent Researcher
Melbourne, Victoria, Australia

ABSTRACT

This paper revisits the analysis of 1988 radiocarbon dating of the Shroud of Turin published in the *Nature* in 1989. The radiocarbon dating has been controversial as the sample taken from the Shroud is believed to be contaminated. Arguing from a novel statistical point of view, the current analysis, within the confines of the data provided in the *Nature* article, further confirms their radiocarbon dates for the Shroud of Turin. It also finds the opposing argument that their result is untenable due to the shortcomings of sample selection arising from the localised heterogeneity such as invisible mending in the selected sample, to be highly unlikely.

INTRODUCTION

Since the publication of the article of Damon *et al* [1], the conclusion of the research team that dated the Shroud of Turin to the medieval period has been controversial. People who believe that the Shroud is the authentic article which covered the body of Jesus are very reluctant to accept the radiocarbon dates falling in the medieval period. However, the sceptics hail this as the final nail on the coffin of an earlier origin for the Shroud. As these views are widely known and the authour's interest is purely data orientated, they are not revisited here. Some obvious statistical issues relevant to sampling raise valid questions. For an excellent discussion on many of the statistical issues relevant to sampling please see Riani *et al* [2]. In Figure 1 of Riani *et al* [2], they discussed the manner in which the original sample from the Shroud was subdivided for each of the three laboratories in Arizona, Oxford and Zurich. Only Arizona had two pieces from two different places of the Shroud sample so that each laboratory received Shroud material of equal weight. These authours also believe that *"the twelve results from the 1988 radio carbon dating... show surprising heterogeneity".* Another main concern for them was the locations of sub-samples from the Shroud material selected by the three laboratories. Christen [3] also reanalysed the data with Bayesian approach using Gibbs sampler and decided to remove the two determinations A1.1b and O1.1u as outliers to further confirm the age of the Shroud arrived at by the Damon *et al* [1].

The current analysis reviews the data provided in the *Nature* article of 1989 to see whether any earlier date would have been probable based on the published determinations. This paper is not intended as another high-end statistical analysis of the Radiocarbon 14 determinations on the Shroud, but a commonsense analysis to answer some questions not posed by the previous researchers. The following analysis acknowledges that the sample was taken from the 'only' possible place given the constraints on destructive sampling of the Shroud. Thus, it is possible that the rest of the Shroud could be different to this specific corner of the Shroud from where the sample was taken. [4] This possibility is out of the scope of this investigation. Similarly, it is very reasonable to believe that the material from the sample provided to three different laboratories is likely to capture a range of contaminations and invisible mending which is discussed by Benford and Marino [5]. Thus, in the current analysis, the author tries to

Copyright © Society for Science and Education, United Kingdom

establish whether the impact of internal inconsistency within the material of the sample resulting from mending or contamination has been captured in the heterogeneity of determinations. When allowed for such inconsistency, had a higher determination of dates (i.e. in BP terms) far removed from the medieval period been probable? These two questions have not been considered in previous analyses to the authour's knowledge.

METHODS

The current analysis solely rely on the data published in Damon et al [1]. Data available in their paper are based on the radiocarbon dating done on the Shroud sample. Together with the sample from the Shroud material, each laboratory was also provided with three control samples. For the purpose of the current analysis, the following labels for the samples will be used:

Sample 1 - the sample from the Shroud of Turin

Sample 2 - Linen sample from a tomb excavated at Qasr Ibrîm in Nubia (dated to 11th-12th centuries AD)

Sample 3 - Linen sample from an early second century AD mummy of Cleopatra (dated to 110 BC - AD 75)

Sample 4 - Threads from the corpse of St Louis d'Anjou (dated to AD 1290 - 1310)

More details on these samples are available from Damon et al [1] whose labels match the above.

As there was some inconsistency of the values reported in Table 2 of Damon et al [1], using the method described in Ward and Wilson [5], their table 2 was recalculated and presented in Table 1 below. Note that Ward and Wilson [6] uses the reciprocal of sample variance as weights for combining determinations to calculate the pooled mean and variance. Next step of the analysis minimises the impact of variability within the determinations on the samples provided to each laboratory and study the variability attributable to these samples from three laboratories. As these laboratories used different pre-treatment methods and possibly had other inherent differences, it was required to assess the likely impact of these procedures. For this purpose, the data from the three control sample were used.

Let us assume, the realisations of the random variable Y for the determination i,

$$y_i \sim N(\bar{y}_i, s_i) \text{ where } i = 1..n$$

To see the impact of laboratories on the mean of the determinations $\bar{y}_i$ associated with standard deviations s_i , for the i th determination for each sample as published in Table 1 of Damon et al [1], the least squares equation as defined below was minimised with respect to the β s defined below.

$$\text{Sum of Squares} = (SS) = \sum_{i=1}^{n} w_i(\bar{y}_i - \beta_1 - \beta_2 x_{2i} - \beta_3 x_{3i})^2$$

where i = 1..n; $w_i = 1/s_i$; $x_2, x_3 =$ indicator variables for Labs 2 and 3 ; β_1 .. β_3 = parameters for three Labs

The analyses were performed using R Statistical Software Package running in Windows 10 desktop environment [7] and the results are presented in the next section.

To further investigate the outcome from the weighted least squares approach above, box plots of the dispersion of the probable determinations for each laboratory were constructed using the process of generating normal random numbers described below. This analysis should provide further information about the distribution of data points between laboratories for the Shroud and the other three control samples while complementing the analysis using Ward and Wilson [6].

Following the methodology used in [8] with respect to several determinations of cranial capacities on the same skull, Monte Carlo simulations as given below were performed. The reason for the use of simulation is the unrealistic realisations from the analytical outcome of the asymptotic nature of statistical distributions considered here. For an example, normal distribution can theoretically produce a realisation falling between -∞ and +∞. The following simulation study while using same parameters as the theoretical distribution generates more of the most likely values. The final outcome in this study is based on the maximum of the most likely realisations an investigation of which is not possible with the analytical methods. Because of these reasons the conclusions drawn from this study will be referred to in probabilistic terms.

To facilitate the discussion of this methodology, let us also assume that we simulated a set of m realisations of y_i as

$$r_{ij} \sim N(\bar{y}_i, s_i) \text{ where } j = 1...m$$

The resulting **R** (n x m) matrix that, for the Shroud sample, had 12 rows and 10,000 columns was subset for each laboratory and used to produce the box plot at the top left hand corner in Figure 1. This plot is thus based on 12 * 10,000 data points subdivided for each laboratory. For an example, the box plot for Arizona is based on 40,000 data points. The same process was repeated for Sample 2 (n = 13), Sample 3 (n = 11) and Sample 4 (n =13) generating 10,000 normal random numbers for each determination. The box plots resulting from these simulations are also presented in Figure 1.

The final analysis performed was a Monte Carlo simulation to investigate the largest possible determinations of BP times if the determinations on the Shroud sample strayed towards the least probable maximal extremes.

Then, using the realisations of normal variates r_{ij} for the Shroud sample, a data set of maxima was constructed as $r_{(n)j}$ for each j. These $r_{(n)j}$ values were then used to find the maximum across all j as $r_{(n)(m)}$ which has been presented in Table 3 and sample histogram of $r_{(n)j}$ from one of the ten trials of 10,000 simulations is also given in Figure 2. Note that theoretically, for large n*m,

$$1 - \int_{-\infty}^{r_{(n)(m)}} f(r)\,dr \rightarrow 0$$

where f(r) represents the normal density function for the simulated random variable R the realisations of which are r_{ij}. Thus, for a normally distributed R the probability of observing a simulated value larger than $r_{(n)(m)}$ is highly unlikely.

Subsequently, it was assumed that the determinations observed by Damon et al [1] represent the maximum BP values for each sub-sample. If the determinations represented the maximum possible for each case, was there any chance of observing a date close to the first century AD even at the extreme? As Gumbel distribution is widely used to study rarely probable events, it was assumed,

$$y_i \sim \text{Gumbel}(\alpha_i, \theta_i)$$

where α_i and θ_i are location and scale parameters respectively that were initially estimated using $\bar{y}_i$, s_i and the first two moments of Gumbel distribution. Based on the random number generation described by Tanizaki [9], a set of Gumbel random numbers were generated as:

$$\mathbf{r}_{ij} \sim \text{Gumbel}(\boldsymbol{\alpha}_i, \boldsymbol{\theta}_i) \quad \text{where } j = 1...m$$

Then, using fitdist function in R Software and $\mathbf{r}_{ij}$ values generated above, updated maximum likelihood estimates were computed for $\boldsymbol{\alpha}_i$ and $\boldsymbol{\theta}_i$ and a new set of r_{ij} values were generated using these updated values as

$$r_{ij} \sim \text{Gumbel}(\alpha_i, \theta_i) \quad \text{where } j = 1...m$$

These were then subject to a process similar to the normal variable case above to find $r_{(n)j}$ and $r_{(n)(m)}$ for Gumbel distributed random numbers. Thus, building on the normal case, a probability of a maximum larger than $r_{(n)(m)}$ is again highly unlikely.

Fre'chet distribution might have resulted in a larger maximum value as its tail tapers off polynomially rather than exponentially to positive infinity.[10] However, given the arbitrariness around the shape parameter to be used with a known sample mean and standard deviation, it was decided to restrict the analysis to the Gumbel case. It is also known that the estimates of the radiocarbon age follows a normal distribution, especially, for age estimates less than 30,000 BP.[6] Given this understanding, use of Gumbel distribution defined by two parameters together with the normal distribution seems reasonable without artificially biasing the outcome.

RESULTS AND ANALYSIS

Reproduction of Table 2 of Damon et al [1] as presented in Table 1 below shows some differences. However, the impact of these differences on the conclusions they have drawn from the table are marginal. Most substantial one is the larger than published difference between laboratories for the Shroud sample (6.4 vs. 8.6 for $\mathbf{X}^2$). Thus, pooling all three sets of determinations from three laboratories to arrive at the final determination can be questioned. However, such pooling can only have impacted the confidence limits although the general conclusion would be still valid. The $\mathbf{X}^2$ values in Table 1 presented in this article also confirm that unlike the Shroud sample, the control samples did not display significant mean differences between laboratories. This further verifies that the localised contamination in Shroud sample was very likely to be captured by the Oxford team as their result was significantly different from the other two.

Table 1 Reworked summary table of mean radiocarbon dates and assessment of inter-laboratory scatter

Laboratory	Sample 1	Sample 2	Sample 3	Sample 4
Arizona	646 ± 17	927 ± 20	1,995 ± 20	722 ± 20
Oxford	749 ± 31	938 ± 29	1,977 ± 33	756 ± 26
Zurich	676 ± 24	941 ± 23	1,940 ± 29	685 ± 25
Unweighted mean*	691 ± 31	935 ± 4	1,971 ±16	721 ± 20
Weighted mean**	672 ± 13	934 ± 13	1,977 ± 15	720 ± 13
X^2 value (2 d.f.)	8.6	0.2	2.5	3.9
Significance * level (%)**	2	90	30	15

This authour argues that the non-homogeneity in the sub-samples between laboratories and even within a laboratory is very likely to represent any contamination or invisible mending in the sample and hence, should be left untouched. The outlying observations reported by Christen [3] are very likely to be real data points which should not be left out. Thus, in the regression analysis using weighted least squares analysis presented next, the non-homogeneity was welcome as it didn't impede the significance testing as shown by the analysis of residuals.

Table 2 shows the result of the regression analysis performed to see the impact of the laboratories on the determinations for the four samples. As the within sample variance was taken care of by the weighting , it was the normality of the residuals that can be influencing the parameter testing. As it can be seen from 'AD Test' column of Table 2, none of the residuals from the regressions on samples showed any significant departures from normality. Furthermore, only the determinations on Shroud sample from Oxford showed significant departure from the base sample of Arizona (β_1). This further confirms the result from Table 1 about the special significance of the Shroud sample. Together these two analyses on the control samples indicate that pre-treatments etc. did not significantly impact the outcome of the laboratories.

Table 2 Weighted Least Squares Analysis of individual determinations from three laboratories

Sample	t-value for Oxford^	t-value for Zurich	AD Test^^
Sample 1	2.52	0.88	0.65
Sample 2	0.25	0.36	0.30
Sample 3	-0.27	-0.91	0.39
Sample 4	0.70	-0.77	0.23

^Sample 1 t-value for Oxford was the only significant estimate at 5% level other than the intercept

^^Anderson Darling normality tests for residuals were not significant

Box plots in Figure 1 based on normally distributed random numbers, in contrast to Ward and Wilson [6], do not show any significant differences between laboratories. However, it shows a fairly high median for the Shroud sample analysed by the Oxford team, while all the controls

point to largely similar distributions of determinations between laboratories. Thus, the most likely source for the differences between the laboratories on the Shroud sample should be the sample itself. Therefore, the impact of internal inconsistency within the material of the sample from any mending or contamination seem to have been captured in the determinations. This makes the date range that Damon *et al* [1] assigned to the Shroud based on the unweighted mean of the determinations to have incorporated any within sample differences.

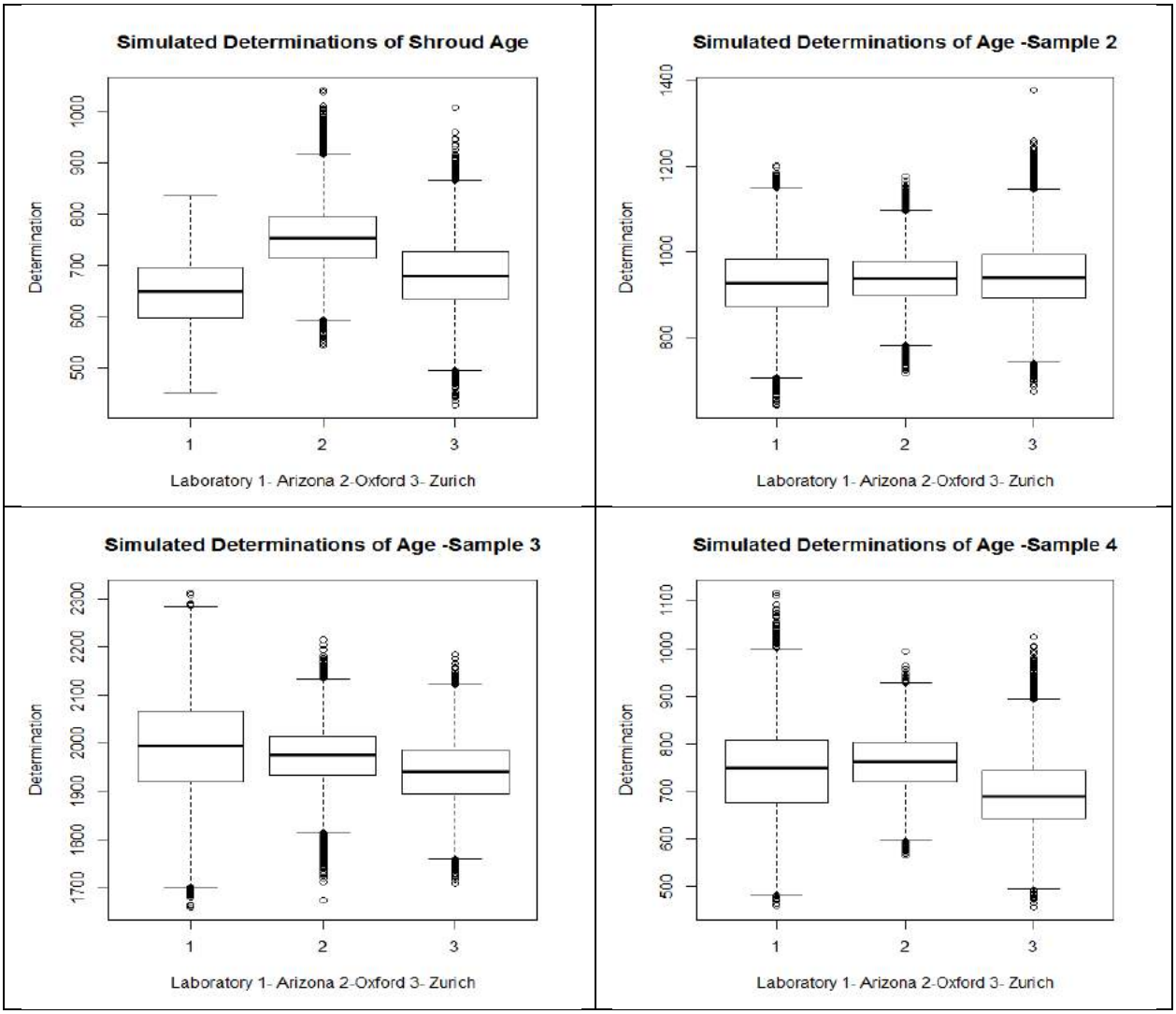


Fig 1. Box plots of probable determinations generated as normal random numbers

Figure 2 shows the distribution of $r_{(n)j}$ and clearly indicates that irrespective of the distribution adopted, majority of the probable maximal determinations fall between 600-1000 BP, roughly 950 AD - 1350 AD. Table 3 also shows that it is highly unlikely to see any determination exceeding 1588 BP (~ 372 AD) within the variability observed in the Shroud sample used for radiocarbon dating. Thus, it is not reasonable to claim that had the localised issues had not been there, an earlier date falling closer to 33 AD would have been probable.

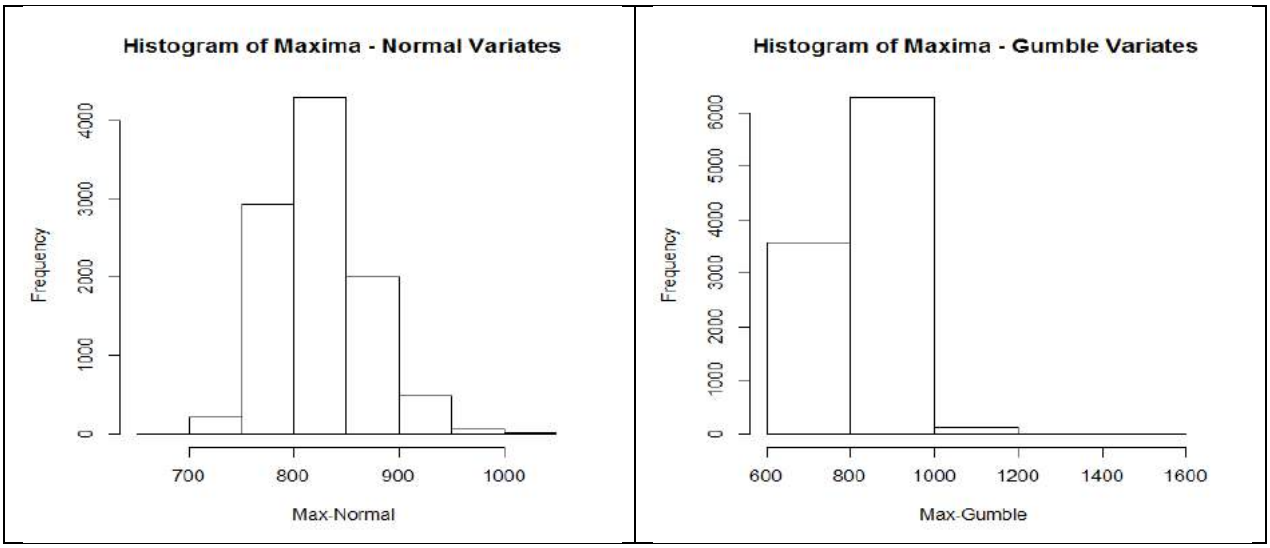


Fig 2. Histograms showing the distribution of probable maximal determinations ($r_{(n)j}$) generated as normal and Gumbel random numbers

Table 3. Maxima of determinations $r_{(n)(m)}$ generated as normal and Gumbel random numbers

Trial of 10,000 simulations->	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	Maximum Age obtained in terms of BP (Shroud Sample)									
Normal Data	1031	1081	1028	998	1043	1077	1077	1077	1038	1028
Gumbel Data	1253	1277	1366	1211	1588	1347	1353	1249	1286	1313

CONCLUSION

The above analysis showed that the radiocarbon determinations published by Damon et al in 1989 had also captured the variability coming from localised mending and contamination within the Shroud sample as the differences between laboratories in addition to intra-laboratory variability. If the heterogeneity due to these localised issues had been big enough to mask the assumed age popularly attributed to the Shroud, i.e. ~33 AD, the onus on the critics of the 1988 radiocarbon dating should now be to show the very extreme probability of producing a determination closer to 1900 BP. The analysis discussed above failed to show even a very unlikely determination of the age of the Shroud going beyond 1588 BP making such a claim likely to be untenable. However, this sets the future direction for the researchers who believe that Shroud is much older than the medieval period. These researchers, instead of focusing on local variability within the Shroud sample to explain away the medieval dates from radiocarbon dating, should now find strong evidence to prove that the sample taken for the radiocarbon dating in 1988 was completely different to the rest of the Shroud.

ACKNOWLEDGMENTS

The authour sincerely acknowledges all the people who read the versions of the manuscript and provided feedback.

References

Damon, P.E., Donahue, D.J., Gore, B.H., Hatheway, A.L., Jull, A.J.T., Linick, T.W., *et al., Radiocarbon dating of the Shroud of Turin*. Nature, 1989. 337, p. 611–615.

Riani, M., Atkinson, A.C., Fanti, G. and Crossilla, F., *Regression analysis with partially labelled regressors: carbon dating of the shroud of Turin*. Statistics and Computing, 2013. 23 (4): p. 551–561.

Christen, J.A., *Summarizing a Set of Radiocarbon Determinations: A Robust Approach.* Applied Statistics, 1994. 43 (3): p. 489-503.

Marino, J.G., and Prior, E.J., Chronological *History of the Evidence for the Anomalous Nature of the C-14 Sample Area of the Shroud of Turin*, 2008 [cited 2017 May 21]. Available from www.shroud.com/pdfs/chronology.pdf

Benford, M.S. and Marino, J., *New Historical Evidence Explaining the "Invisible Patch" in the 1988 C-14 Sample Area of the Turin Shroud*, 2005 [cited 2017 May 21]. Available from http://www.shroud.com/pdfs/histsupt.pdf

Ward, G.K. and Wilson, S.R., *Procedures for comparing and combining radiocarbon age determinations: a critique*. Archaeometry, 1978. 20 (1):p. 19-31

R Core Team., R: *A language and environment for statistical computing. R Foundation for Statistical Computing*, 2013 [cited 2017 May 22]. Available from URL http://www.R-project.org/.

Arachige, D., *An Analysis of Hominin Cranial Capacity Data Using Simulations* [Internet]. Anthropol, 2015 Aug [cited 2016 Dec 26]. 3 (2) :p157. Available from https://www.omicsonline.org/open-access/an-analysis-of-hominin-cranial-capacity-data-using-simulations-2332-0915-1000157. DOI:10.4172/2332-0915.1000157

Tanizaki, H., *Computational Methods in Statistics and Econometric, 2004,* New York: Marcel Dekker. p. 130

Coles, S., *An Introduction to Statistical Modeling of Extreme Values*, 2001, London: Springer. p. 47