

# On Cleaning Methods and the Raw Radiocarbon Data from the Shroud of Turin

Larry Schwalbe\*, Bryan Walsh

Shroud of Turin Center, Richmond, Virginia, USA

**Email address:**

lschwalbe@tampabay.rr.com (L. Schwalbe), Shroud\_Center@comcast.net (B. Walsh)

\*Corresponding author

**To cite this article:**

Larry Schwalbe, Bryan Walsh. On Cleaning Methods and the Raw Radiocarbon Data from the Shroud of Turin. *International Journal of Archaeology*. Vol. 9, No. 1, 2021, pp. 10-16. doi: 10.11648/j.ija.20210901.12

**Received:** February 19, 2021; **Accepted:** March 5, 2021; **Published:** March 12, 2021

---

**Abstract:** The Shroud of Turin is a long, narrow strip of linen cloth believed by many to be the burial cloth of Jesus. The Shroud is unique because faint images of a crucified man are clearly visible on one surface. These body images along with accompanying blood stains have been the subject of scientific inquiry for over a hundred years, yet the process of the image formation has been and remains unknown. Among the more recent of coordinated studies of the Shroud was a radiocarbon dating of excised samples. The results, published in 1989, place the origin of the cloth to sometime in or around the 14th century. The objective of the present study is to survey the cleaning methods (or pretreatments) that were applied to the samples removed for the radiocarbon study. Specifically, we explore the extent to which these methods may have given rise to a peculiar structure in the "raw" radiocarbon data published in 2019. The data from two of the participating laboratories, Zurich and Arizona, appear to bifurcate into groups separated by roughly 100 radiocarbon years. By comparing the pretreatment for each subsample and its group membership, we conclude that these pretreatments do not account for the bifurcation effect. As all subsamples represent portions excised from an originally intact and continuous sample of Shroud material, we assume they are all the same calendar age. Granted this assumption and given the results of the present study, two hypotheses remain to account for the curious anomaly: either 1) the carbon isotope ratios  $^{14}\text{C}/^{12}\text{C}$  of the fabric itself were altered by some currently unknown process, or 2) a non-isotropic distribution of contamination remained after the samples underwent the documented pretreatments. A resolution of the question is important for deciding whether future radiocarbon studies are called for and, if so, how the testing protocols should be structured.

**Keywords:** Shroud, Radiocarbon, Inter-laboratory, Statistics, Pretreatment

---

## 1. Introduction

The Shroud of Turin is a rectangular strip of linen cloth (4.4 m x 1.1 m) bearing on one surface the head-to-head, dorsal and ventral images of a man apparently crucified. Some revere the Shroud as the burial cloth of Jesus; others consider it a manufactured article with only historical significance. Curious persons from both groups have pondered the nature of the images, how they came to be, and the origin of the Shroud and its subsequent history.

In 1898, Secondo Pia, a Turinese photographer, produced the first high-quality photographic image of the face. This event marked the beginning of a series of largely uncoordinated observational and empirical studies that followed over the next 70 years. Wilson [1] published a

readable overview of most of these studies, which included additional photographs, textile and pollen analyses, medical forensics, and studies comparing features of the image with those expected from descriptions of the crucifixion in the Gospel accounts. The results of a few of these studies became available in *Sindon*, the official journal of the *Centro Internazionale di Sindonologia* in Turin, others through numerous privately published accounts. To our knowledge none of this work appeared in the more widely available scientific literature.

The first coordinated scientific inquiry occurred in 1978 by a group of volunteer scientists, engineers, and photographers organized as the Shroud of Turin Research Project (STURP). These researchers transported a collection of modern scientific equipment to Turin and carried out their

investigations in the cathedral where the cloth is kept. The results of these studies appear in a number of refereed scientific journals. Citations of 20 publications documenting the direct results of the investigations are found on the website [shroud.com/78papers.htm](http://shroud.com/78papers.htm). The listing also contains references to an additional 10 articles including non-refereed publications and results of follow-on technical studies.

Two summary papers [2, 3] detail the project's two major conclusions: 1) the carmine-colored "blood" images were derived from actual blood, and 2) the sepia-colored body images consist of fibers on the very outer surface of the cloth that were turned yellow by an oxidation/dehydration process that modified their chemical structure. No application method was identified that produces all of the chemical and physical characteristics of the body image.

The second coordinated inquiry was the radiocarbon study conducted in 1988 [4]. That work involved three internationally recognized testing facilities, those at the University of Oxford, the *Eidgenössische Technische Hochschule* in Zurich, and the University of Arizona in Tucson. Each laboratory employed accelerator mass spectrometry (AMS), a technology that required a minimal amount (roughly ~50 mg) of textile sample. In addition to the Shroud material, the study included three control samples with independently determined dates that spanned the range of reasonable estimates for the age of the Shroud.

The British Museum served as the program coordinator. Their involvement included the collection of each laboratory's data and the analysis and release of the final results. The conclusion of the study is that the extracted sample materials date to the 14<sup>th</sup> century AD.

## 2. Recent Studies

Casabianca and coworkers [5] initiated the latest interest in Shroud studies. As stated above, the protocol adopted for the radiocarbon program had the investigating laboratories submit their individual results – their "raw" data – separately to the British Museum for analysis and compilation. Differences between the "raw" data sets and the final version [4] consist of changes to seven of the "raw" data uncertainty terms for the measurements reported with six of the seven uncertainties increased. The changes to the three Oxford "raw" data all increase their uncertainties by anywhere from 20% to 50%. In the case of the Zurich measurements there are large and unexplained differences in two dates. These two modified data align closely with the lower pair of Arizona results and together become the focus of the present study. Until the later report [5], no details of the Museum's involvement were available, only the statistical analyses that Walsh and Schwalbe [6] challenge, but certainly nothing further about how and why changes in any of the data came about.

In response to a freedom-of-information request in 2017, all of the raw data held by the Museum were released and published [5]. In addition to making this information generally available, these researchers analyzed the raw data

and conclude that the results from the three laboratories are statistically heterogeneous, a condition that according to standard analytical procedure [7] precludes these dates from being combined to produce an accurate and unbiased average, or for that matter analyzed as a composite set in a program that applies a calibration curve to convert radiocarbon years to calendar years.

In addition, the Casabianca team built on some of the work that Walsh [8] undertook years earlier. Walsh identified the original locations of the laboratory samples in the precut fabric and, using the centroid positions of these samples as proxies for the locations of each subsample, he found a likely linear dependence for the dates on these locations. The Casabianca team followed the same prescription and came to the same conclusion for the raw data.

Casabianca and coworkers [5] attribute the variation of dates with position to a corresponding variation in the carbon isotope ratio  $^{14}\text{C}/^{12}\text{C}$  across the precut fabric. Without identifying a specific source, they argue that a non-uniform distribution of the carbon ratio of the observed magnitude over the limited spatial extent of the sample (~6 cm<sup>2</sup>) undermines the assumption that the test results reported in 1989 [4] are representative of the cloth as a whole (~4.8 m<sup>2</sup>), and it follows that a program of new testing is in order.

Shortly after the Casabianca paper appeared, Walsh and Schwalbe [6] published the results of their statistical analysis of the earlier published data [4]. Their results show the earlier data are also heterogeneous and well represented as a linear function of the samples' distancing. In addition to interpreting the effect as the consequence of a non-uniform distribution of radiocarbon in the textile, they allow for the possibility that the effect may instead be due to a residual sample contamination.

As an alternative to a linear relationship, Walsh and Schwalbe demonstrate that the laboratory data may be described equally well by a step function with the Oxford data offset from those of Zurich and Arizona by a statistically significant amount [6]. To suggest a mechanism for how an offset like this might have occurred, these authors note that the Oxford team used petroleum ether in their cleaning procedure which the other two laboratories did not. Petroleum ether is an efficient solvent for materials such as lipids and waxes that may have contaminated the samples and had not been removed in the treatments that Zurich and Arizona applied. Included with their analysis, Walsh and Schwalbe propose a limited set of experiments to test this hypothesis.

## 3. Objectives and Scope

The present study accepts the hypothesis that a non-uniform distribution of the carbon-isotope ratio  $^{14}\text{C}/^{12}\text{C}$  in the cellulose fabric may have been the source of the statistical heterogeneity in the data and in the non-constant functional behavior as discussed above. However, another source may have been a residual contamination that the documented pretreatments failed to remove. In either case, we agree with

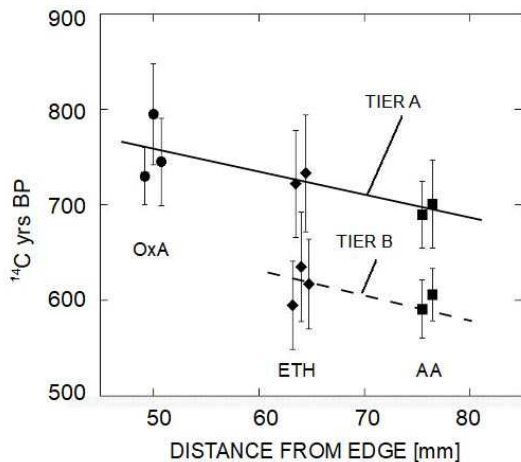
Casabianca that solid arguments follow for retesting the Shroud material using up-to-date AMS methods.

The objective of the present study is to examine the possible effect of the pretreatments documented in Damon [4] as they bear on a portion of the raw data sets that Casabianca [5] presents. Our study utilizes only the data published in these two sources. As both articles appear in respected, peer-reviewed journals, we have no cause to question the validity or the accuracy of the data, and we make no effort to do so. Our intent is solely to explore the heterogeneity in the data and develop testable hypotheses to identify the source of that heterogeneity. We do not assess nor challenge the general medieval dating conclusions reached in Damon [4]. Besides whatever implications our results have for the composition of the tested Shroud samples, we hope that our findings prove useful in the specifications of cleaning protocols should a future program of radiocarbon testing be undertaken.

## 4. The Raw Radiocarbon Data

Figure 1 is a plot of radiocarbon years before present ( $^{14}\text{C}$  yrs BP) as a function of distance from the left-hand edge of the original fabric sample as defined and represented in Walsh and Schwalbe [6]. All of the data are drawn from Table 1 of Casabianca [5]. The data sets in the figure are labeled according to the official laboratory code designations (OxA for Oxford, ETH for Zurich, and AA for Arizona).

The points labeled “OxA” and “ETH” are from the columns that Casabianca labels “Oxford Raw” and “Zurich Raw”, respectively. The points labeled “AA” are from the column labeled “Arizona Nature”. Unfortunately, the designation “Arizona Nature” has caused some confusion because it seems to imply that the “raw” data from Oxford and Zurich are being mixed with Arizona data that were released much later and published in *Nature* [4]. In fact, all of the data in the figure were reported to the British Museum at roughly the same time, before their analyses and compilations.



**Figure 1.** Plot of raw  $^{14}\text{C}$  data as a function of the sample positions. The data are presented at  $\pm 1$  sigma uncertainties.

Damon [4], Casabianca [5], and Van Haelst [9] document

the sequence of events surrounding the Arizona data. To begin, Damon reports that Arizona divided their sample into four subsamples. Subsequently, Van Haelst learned that Arizona produced eight “paired” dates which he terms “dependent” data. Each pair of measurements represents one of the four subsamples. According to both Casabianca and Van Haelst, each pair of data was generated individually on the same day with the same set of standards and blanks. These data, appearing in Casabianca under the column heading “Arizona Raw 1”, were those originally transmitted to the Museum.

Shortly after this transmittal, Arizona corrected two data from a single pair that would otherwise have prevented their combination. The corrected 8-point set appears in the Casabianca report as “Arizona Raw 2” [5]. After the correction, the Museum requested that the individual pairs of subsample data be combined into an “independent” set. The resulting 4-point data set is the final form that appears in the *Nature* article [4] and is reproduced in Table 1 of Casabianca under the column heading “Arizona Nature” [5].

## 5. Observations and Assessments

Figure 1 displays a curious feature that is not readily apparent in corresponding plots of the data published by Damon [4]: the Zurich and Arizona dates show roughly matching bifurcations (or gaps) that are not easily explained. In this regard, it is important to stress that all of the data in the figure were derived from subsamples cut from a single, originally intact and continuous strip of linen fabric (see e.g., Figure 1 in Walsh and Schwalbe [6]). This information leads us to proceed with the working assumption that all subsamples share the same calendar age. If in addition, the original fabric sample contained a uniform distribution of residual contamination or no contamination at all, we would expect to see all of the data in the figure agreeing within experimental error, that is, all falling roughly along a single horizontal line in the figure. As we and others have shown, this is clearly not the case.

To support our continuing study, we present a statistical analysis of the bifurcation effect in Appendix 1. The results show that the data structure is statistically significant and may therefore derive from a non-uniform distribution either of the fabric’s isotopic composition, the  $^{14}\text{C}/^{12}\text{C}$  ratio, or of a residual contamination after the pretreatments detailed in [4] were performed.

Some have argued that natural variations in the atmospheric concentrations of  $^{14}\text{C}$  could account for the bifurcation, particularly since the calibration curve (which converts radiocarbon years to calendar years) appears to vary substantially during the 1200-1400 AD time period. However, it is clear the cause of the effect lies elsewhere. As stated above, all of the subsamples derive from a single strip of textile and should be the same age. An application of the calibration curve should therefore reflect identical concentrations of atmospheric  $^{14}\text{C}$  for each of the radiocarbon dates. If this information is properly incorporated into the

analysis, the entire data structure would then be transformed uniformly rather than as individual, uncorrelated points separately. The plot of calendar dates would show the same structure as seen in Figure 1 albeit with different dates and uncertainties.

Moreover, because the radiocarbon dates are dependent in this sense, it is not legitimate simply to plug them into a calibration program, as if they were statistically independent, and expand the individual error limits shown in Figure 1 based on the uncertainties of the calibration. Doing so admittedly expands the limits of the calendar dates to the extent that the data structure is rendered statistically insignificant. But the operation as stated is not legitimate. That is, an extension of the uncertainty limits of the individual calendar dates should each have the same sign and magnitude. The best that might be done is to present an average radiocarbon date and a corresponding calendar date with its uncertainty limits. However, as stated in a previous section, even this approach is problematic because Casabianca [5] demonstrates that the complete set is heterogeneous and should not be combined per the standard analytical procedure [7].

Thus far, we have only discussed the bifurcation effect seen in the Zurich and Arizona data sets. The Oxford data show no bifurcation, but as stated in Casabianca [5], Oxford performed five measurements and subsequently averaged several of these to provide the Museum with the three data that were ultimately published. Whatever the details of this reduction process, the “higher” clusters of two points each in the Zurich and Arizona sets appear to align with the complete Oxford set. The “lower” points can be roughly described by a line approximately parallel to that drawn through the “upper” set but displaced downward by roughly 100  $^{14}\text{C}$  yrs. To facilitate discussion of the bifurcation effect, we refer to the higher set of points as Tier A data and the lower set as Tier B.

Table 1 shows the results of an error-weighted regression analyses applied to the data in each of the separate tiers. We find the linear dependence exhibited by the upper tier to be statistically significant. That for the lower tier is not. The line drawn through the lower-tier data is therefore conjectural and is shown as dashed to indicate it as such. The linear dependence of the upper tier is similar to the functional relationship that [5] reported for the raw data and as [6] did for the data published in *Nature* [4]. The step-function relationship that [6] discussed may appear less compelling when the Oxford results are taken together with the upper tier data of Zurich and Arizona, but we believe the experiments that [6] suggest to resolve the differences in the functional relationships should be completed to ascertain definitively the true nature of the different date measurements.

**Table 1.** Results of error-weighted linear regression performed on separate data tiers.

	intercept	slope	F-stat	p-value
	[ $^{14}\text{C}$ yrs]	[ $^{14}\text{C}$ yrs/mm]		
Tier A	843.7	-1.9284	8.6531	0.0322
Tier B	705.0	-1.4295	1.5795	0.2978

We list all of the raw radiocarbon data that Casabianca [5]

published in Tables 2, 3, and 4, grouping the data into their respective tiers and assigning to each the identification code used by Damon [4] in their Table 1 “Basic Data (individual measurements)”. As stated above, the apparent bifurcation is puzzling, but it is reasonable to consider the pretreatments as a possible source for the effect since the various methods were not all applied uniformly to the subsamples (see Appendix 2). Therefore, along with the tier assignments and the individual dates, we include the pretreatments that Damon [4] describes for each subsample.

Beginning with the initial steps in the pretreatment program, we note that only Oxford and Zurich reported any specifics, but these were applied uniformly to each of their respective sets of subsamples. Apart from the pet-ether treatment that Oxford used and that Walsh and Schwalbe [6] proposed as a possible cause for a step-function behavior in the published data [4], the initial pretreatment steps listed in the present tables look to be an unlikely source for the bifurcation feature.

**Table 2.** Oxford reported raw radiocarbon data with pretreatments.

Tier	ID Code	OXFORD		
		Raw ( $^{14}\text{C}$ yrs BP)	Initial Cleaning	Pretreatments
A	O1.1u	795 $\pm$ 53	pet ether	strong A-A-A unbleached
	O1.1b	745 $\pm$ 46	pet ether	strong A-A-A bleached
	O1.2b	730 $\pm$ 30	pet ether	strong A-A-A bleached

**Table 3.** Zurich reported raw radiocarbon data with pretreatments.

Tier	ID Code	ZURICH		
		Raw ( $^{14}\text{C}$ yrs BP)	Initial Cleaning	Pretreatments
A	Z1.1u	733 $\pm$ 61	ultrasonic	none
	Z1.1w	722 $\pm$ 56	ultrasonic	weak A-A-A
	Z1.1s	635 $\pm$ 57	ultrasonic	strong A-A-A
B	Z1.2w	617 $\pm$ 47	ultrasonic	weak A-A-A
	Z1.2s	595 $\pm$ 46	ultrasonic	strong A-A-A

**Table 4.** Arizona reported raw radiocarbon data with pretreatments.

Tier	ID Code	ARIZONA (4-POINT)		
		Nature ( $^{14}\text{C}$ yrs BP)	Initial Cleaning	Pretreatments
A	A1.4a	701 $\pm$ 33	none reported	method a (weak A-A-A)
	A1.2b	690 $\pm$ 35	none reported	method b (detergent)
B	A1.3a	606 $\pm$ 41	none reported	method a (weak A-A-A)
	A1.1b	591 $\pm$ 30	none reported	method b (detergent)

Regarding the pretreatment procedures, we look first at the Oxford data (Table 2) where each sample received a relatively strong acid-alkali-acid (A-A-A) treatment followed in two cases by bleaching. The bleaching seems to have had little cleansing effect because the three data agree within experimental error. Similarly, the strong A-A-A treatment is likely not to have been a factor either because of reasons to be discussed shortly.

The Zurich and Arizona data paint a similar picture. Subsample Zu1.1u received no further pretreatment after its ultrasonic cleaning, yet it produced the oldest  $^{14}\text{C}$  yrs BP value, suggesting it may either have had a greater amount of

earlier contamination or a lesser amount of the more modern. Subsamples Z1.1w and Z1.2w both received weak A-A-A treatments as did the Arizona subsamples A1.4a and A1.3a, which they describe as their "method a" treatment (see Appendix Table 5), yet the individual subsamples in these pairs appear in different tiers thereby suggesting that the weak A-A-A treatment is not the cause of the bifurcation. We draw the same conclusion from the subsample pair A1.1b and A1.2b, both of which were subjected to the Arizona "method b" treatment (see Appendix A2).

To probe the Arizona data a bit further, we compute the correlation coefficient between the individual data in the 8-point set "Arizona Raw 2" and their respective pretreatments ("method a" versus "method b" per Damon [4]). The result yields a coefficient of  $0.098 \pm 0.224$ , which is likewise consistent with no correlation.

Finally, in the Zurich data, the strong A-A-A treatment only appears in Tier B as an application to subsamples Z1.1s and Z1.2s; however, these data bracket Z1.2w, which from the preceding observation received an apparently ineffective A-A-A treatment. Together, these observations imply that the strong A-A-A treatment is likewise ineffective and unable to explain the bifurcation.

Regarding the linear trends as functions of the inter-sample locations (see Figure 1), as shown on Table 1 the combined data seem to be too sparse to warrant any firm conclusions. The Oxford grouping showing the oldest  $^{14}\text{C}$  yrs BP values underwent strong A-A-A treatments, but the Zurich samples Z1.1s and Z1.2s having undergone similar treatments are found in the lower Tier B, some 140  $^{14}\text{C}$  yrs younger than the Oxford grouping. Therefore, almost all of the pretreatments appear to be insignificant factors in the linear trends or in the standing of the Oxford results, although as Walsh and Schwalbe [6] suggests the pet-ether pretreatment may have played a role in the latter.

## 6. Discussion

The results in Figure 1 indicate a bifurcation in the Zurich and Arizona data that cannot be explained by any of the documented pretreatments applied to them. Interestingly, we observed above that the untreated sample Z1.1u and the unbleached sample O1.1u both have the oldest  $^{14}\text{C}$  yrs BP values in their respective sample sets. We noted that this standing may have resulted from a minimal amount of more modern contamination, but it could also result from the residual presence of early contamination that the pretreatments on the other subsamples were able to remove. The bifurcation effect may therefore result either from a non-uniform distribution of contamination that did not respond to the documented pretreatments or from a non-uniform distribution of  $^{14}\text{C}/^{12}\text{C}$  in the cellulose fabric. However, the upper-tier data do seem to suggest a uniform linear dependence similar to those described by Walsh and Schwalbe [6], Riani and coworkers [10], and others for the earlier published data [4].

## 7. Summary and Conclusions

We review the radiocarbon data that were collected by the laboratories participating in the 1989 study of the Shroud of Turin and originally submitted to the British Museum for analysis and compilation. The raw data cited in the Casabianca report [5] and plotted against the distances between the centroid locations of the laboratory samples show a bifurcation in the data sets that Zurich and Arizona reported. The results of statistical analysis support the contention that the bifurcation is a real effect. The radiocarbon dates in Tier A also appear to show a linear dependence on the original sample locations corresponding to  $\sim 20$   $^{14}\text{C}$  yrs/cm. An examination of the pretreatments applied to the individual Arizona and Zurich subsamples indicates that none of these procedures produced the bifurcation effect. However, it is possible the effect results from a non-uniform distribution of a contaminant that does not respond to the cleaning techniques applied in the radiocarbon study.

As the present findings join with observations of other unique aspects of the Shroud's makeup (see e.g. [11-13]), it appears the composition of the relatively small sample removed for the 1988 study is proving to be surprisingly complex. Indeed, the collection of evidence should encourage researchers to begin reconsidering the validity of the assumption that this sample adequately represents the composition of the Shroud as a whole. Should these concerns prompt follow-on radiocarbon studies, their test plans should include at a minimum 1) careful deliberations about sample locations, 2) a set of narrowly targeted non-destructive tests including optical microscopy, Fourier Transform Infrared Spectroscopy (FTIR) and UV fluorescence studies [14, 15], and 3) complete documentation not only of the sample locations on the main body of the cloth but also the locations of the subsamples, their respective  $\delta^{13}\text{C}$  values, %C content, pre-treatment yield etc.

## Appendix

### Appendix 1. Statistical Analysis

To assess whether or not the perceived bifurcation is a statistically significant feature of the data, we first apply the chi-square ( $\chi^2$ ) test for homogeneity (see e.g. Eq. 5 from [6]) among the full set of 12 raw data. The set under consideration are those from Casabianca [5] (3 Oxford, 5 Zurich, and 4 Arizona) as depicted in Figure 1 and described in the accompanying text. The null hypothesis  $H_0$  for this test is that the data are homogeneous and may be treated as an internally consistent set subject to follow-on analyses including combining the data to compute mean values, variances, etc. The results of the test are  $\chi^2 = 29.180$  with  $\text{dF} = 11$  yielding a  $p\text{-value} = 0.002 < 0.05$ . We find  $H_0$  rejected, an outcome consistent with the similar finding in [5].

Given the lack of justification to further treat the full 12-point set as a homogeneous composite, we restrict our continuing analysis to the reduced set of 9-points comprised only of those from Zurich and Arizona where the bifurcation is present. An

application of the  $\chi^2$  to this reduced set gives a  $\chi^2 = 14.426$  with  $dF = 8$  yielding a  $p$ -value =  $0.07 > 0.05$ . For this case, the null hypothesis is accepted thus allowing us to proceed with a working definition of the bifurcation or gap size.

We begin by computing the mean values and standard deviations of the Tier A and Tier B components of the 9-point set. The results are  $711.5 \pm 19.5$   $^{14}\text{C}$  yrs BP for Tier A and  $608.8 \pm 17.8$   $^{14}\text{C}$  yrs BP for Tier B. We take the difference,  $102.7 \pm 26.4$  radiocarbon years, to represent the gap size. It now remains to test whether the Tier A and B data clusters are measurably distinct by a statistically significant amount.

Two standard classical tests are available for this purpose: the Student  $t$  statistic which assumes the variances of the two distributions are equal and the Welch  $t$  that allows for unequal variances. We apply the Welch test because it permits us to evaluate all possible data sets regardless of variance. The null hypothesis  $H_0$  for this test is that the difference between the mean values of Tier A and Tier B is zero within the observed uncertainties. We obtain  $t = 8.148$  with a  $dF = 6$ .

Typically, we would calculate the associated  $p$ -value from the theoretical distribution and then compare that to some predetermined critical value – most often 0.05. This final step may be problematic, however. Since the Welch statistic, as well as many other classical statistics, requires independent data and since in this instance a single data set appears to

bifurcate as just described, there may be a concern for the actual independence of the two subsets. Questions arise because the null hypothesis going forward is that the data are all drawn from a common parent distribution, and this condition may disqualify the assumption of independence for the two data clusters.

To avoid this issue, we apply a Monte Carlo method to generate a distribution for the Welch statistical parameters. The model uses 9 normal distributions; each centered (arbitrarily) on zero and each having a standard deviation corresponding to each of the uncertainty values reported in the raw data set. The next steps are to: 1) generate a set of “data” by randomly selecting a value from each of the 9 normal distributions, 2) sort the resulting “data” into their order of increasing values, 3) compute the mean and variance of the lower 5 values (Tier B) and those of the higher 4 (Tier A), 3) compute the corresponding Welch  $t$  statistic, 4) increment a corresponding bin value to build a numerical probability density function, 5) repeat steps 1 through 4 a large number (200,000) of times, 6) compute the cumulative distribution  $P(t)$  using a running sum of the probability density histogram.

Using this method, we obtain a  $p$ -value =  $1 - P(8.148) = 0.0024 < 0.05$ . The test rejects the hypothesis that the mean values are equal within the uncertainties observed thus supporting the conclusion that the bifurcation or gap is a statistically significant feature of the data.

## Appendix 2. Sample Treatments

Table 5. Sample pretreatments and materials per Damon et al. (1989).

SAMPLE ID	PRETREATMENT TYPE	COMMENTS
OXFORD		
INITIAL CLEANING	Vacuum pipette removal of foreign material.	
	Petroleum ether (40°C for 1 hr)	Effective on complex hydrocarbons: lipids, waxes
All	Strong A-A-A: 1M HCl (80°C for 2 hr), 1M NaOH (80°C for 2 hr), repeat acid with intermittent rinsing.	Effective on polar compounds
O1.1u	One sample unbleached	Oldest measurement Oxford determined
O1.1b, O1.2b	Both samples bleached in NaOCl (2.5% at pH-3 for 30 min.)	
ZURICH		
INITIAL CLEANING	All samples in ultrasonic bath	
Z1.1u	No further treatment	Oldest measurement Zurich determined
Z1.1w, Z1.2w	Weak A-A-A: 0.5% HCl (room temperature), 0.25% NaOH (room temperature, repeat acid with rinsing in between.	Effective on polar compounds
Z1.1s, Z1.2s	Strong A-A-A: 5% HCl (80°C), 2.5% NaOH, repeat acid with rinsing in between	Effective on polar compounds
ARIZONA		
INITIAL CLEANING	None stated	
A1.3a, A1.4a	“method a” Weak A-A-A: dilute HCl, dilute NaOH, repeat acid with intermittent rinsing.	Effective on polar compounds
A1.1b, A1.2b	“method b” SDS/Triton X-100 followed by ethanol Soxhlet	Effective on polar compounds
Key:	A-A-A acid-alkalai-acid	
	SDS sodium dodecyl sulfate is an anionic surfactant used for denaturation of native proteins	
	Triton X-100 octyl phenol ethyloxylate where n is usually ~9 or 10 – nonionic surfactant	

## References

- [1] Wilson, I., The Shroud of Turin, Doubleday & Company, Inc., Garden City, New York (1978).
- [2] Schwalbe, L. A. & Rogers, R. N., Physics and chemistry of the Shroud of Turin, a summary of the 1978 investigation, *Anal. Chim. Acta* 135, 3-49. (1982).
- [3] Jumper, E. J., Adler, A. D., Jackson, J. P., Pellicori, S. F., Heller, J. H. & Druzik, J. R., A comprehensive examination of the various stains and images on the Shroud of Turin, *Archaeological Chemistry III*, ACS Advances in Chemistry No. 205, J. B. Lambert, Editor, Chapter 22, American Chemical Society, Washington D.C., pp. 447-476 (1984).

- [4] 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. & Tite, M. S., Radiocarbon dating of the Shroud of Turin, *Nature* 337 (6208), pp. 611-615 (1989).
- [5] Casabianca, T., Marinelli, E., Pernagallo, G. & Torrisi, B., Radiocarbon dating of the Turin Shroud: new evidence from raw data, *Archaeometry*, <https://doi.org/10.1111/arc.12467>.
- [6] Walsh, B., & Schwalbe, L., An instructive inter-laboratory comparison: the 1988 radiocarbon dating of the Shroud of Turin, *Journal of Archaeological Sciences-Reports* (2020).
- [7] Ward, G. K. & Wilson S. R., Procedures for comparing and combining radiocarbon determinations: a critique, *Archaeometry* 20, 19-31 (1978) <https://doi.org/10.1111/j.1475-4754.1978.tb00208.x>.
- [8] Walsh, B. J., The 1988 Shroud of Turin radiocarbon tests reconsidered, *Proceedings of the 1999 Shroud of Turin International Research Conference Richmond Virginia*, Magisterium Press, Glen Allen, Virginia (2000).
- [9] VanHaelst, R., Radiocarbon dating the Shroud a critical statistical analysis, <http://shroud.com/vanhels3.htm> (1997).
- [10] Riani, M., Atkinson, A. C., Fanti, G., & Crusilla, F., Regression analysis with partially labeled regressors: carbon dating of the Shroud of Turin, *J. Statistics and Computing* 23, 551-5561 (2012).
- [11] Adler, A. D., Updating recent studies on the Shroud of Turin, *Archaeological Chemistry: Organic, Inorganic, and Biochemical Analyses*, Mary Virginia Orno, ed., American Chemical Society, (1996).
- [12] Rogers, R. N., Studies on the radiocarbon sample from the Shroud of Turin, *Thermochimica Acta* 425 (1-2), 189-194 (2005).
- [13] Freer-Waters, R. & Jull, A. J. T., Investigating a Dated Piece of the Shroud of Turin, *Radiocarbon* 52 (4) 1521-1527 (2010).
- [14] Pellicori, S., UV fluorescence imagery of the Turin Shroud – digitally revisited, *International Journal of Archaeology*, 8 (2), 32-36 (2020) doi: 10.11648/j.ija.20200802.13.
- [15] Kearse, K. P., Unanticipated issues in serological analysis of blood species – the Shroud of Turin as a case example, *Forensic Science International – Reports* <http://doi.org/10.1016/j.fsir.2020.100073>