Journal of Archaeological Science 63 (2015) 203–205

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Review

Layers of assumptions: A reply to Timpson, Manning, and Shennan

Tobias Torfing

Human Development in Landscapes, Institut für Ur- und Frühgeschichte, Christian-Albrechts-Universität, Johanna-Mestorf-Straße 2-6, D-24118 Kiel, Germany

ARTICLE INFO

Article history:
Received 7 July 2015
Accepted 23 August 2015
Available online 25 August 2015

Keywords:
Population
Neolithic
Proxy
Sampling
Radiocarbon
C14

ABSTRACT

In a response to my paper "Neolithic population and summed probability distribution of ^{14}C-dates", Timpson, Shennan, and Manning accuse me of making a series of inferential mistakes. Their argument is based on the opinion that if advanced statistical treatment of data is performed and an explicit null hypothesis is tested, then the argument is well founded. I argue that they ignore a series of underlying assumptions that connect the object of interest - prehistoric population - to the data they utilize - radiocarbon dates. In my article, I explored some of the issues regarding these assumptions and demonstrated that it is difficult to argue that these assumptions are true. If the underlying assumptions are not true, or it cannot be established whether they are true or not, further statistical analysis of the data will not provide a reliable result.

Contents

1. Introduction

First I would like to thank Timpson, Manning, and Shennan for their response to my article "Neolithic population and summed probability distribution of ^{14}C-dates". It was interesting to read, and I always welcome a thorough debate. I would like to use the opportunity to A. clarify some points in the article which have been missed or misunderstood, and B. delve into some important aspects of the inferences made on the archaeological record and how these relate to the radiocarbon record. Especially I wish to take up two core aspects which the respondents have commented upon: the use of proxies and the nature of sampling in archaeology.

Before I continue with these aspects, I would like to state that I *do not* believe, as the respondents claim, that "*unless we have complete knowledge of all the factors that might possibly affect the record available to us, which of course we never will, then we cannot say anything at all*", nor do I believe that archaeology should limit itself to be concerned with catalogues and lists. I am very much in favour of both statistical inferences about past societies, interpretation of the archaeological record, and the development of models of social and economic character, even based on the record as we have it.

On one point I will concede, I *do* believe that science progresses through the accumulation of doubts, but I firmly believe that the doubts and the critical evaluation of these are exactly what will allow us to avoid a bog of ignorance.

A second misunderstanding is related to what is perceived as my argument in favour of excluding certain sites:

"*Indeed following this approach would logically result in the exclusion of all data. When we suspect a bias in some data (or that it is otherwise unreliable/erroneous) it is tempting to* ***assume that the exclusion of that subset must improve the overall quality****, and therefore the reliability of the inferences drawn*" (emphasis mine).

E-mail address: tdanborg-torfing@gshdl.uni-kiel.de.

http://dx.doi.org/10.1016/j.jas.2015.08.017
0305-4403/

This was not an assumption made in the article, as clearly stated: "*the result is not a more "accurate" population model, as the sites clearly exist and were used in the past.*" (Torfing, 2015, section 2., see also section 3.1 and 3.2 of the article for parallel quotes). Instead, I pose some questions (or doubts), regarding the results "*How do we compare the population "signature" of very diverse types of sites? Is estimation on all sites better or worse than those made on settlements alone?*" (Torfing, 2015, section 3.1). I do not answer these questions, since we cannot decide this beforehand. My point relates to some fundamental statistical considerations that need to be evaluated *before* sampling and statistical calculations are processed. The most important is the simple question: How does our proposed data-set relate to the questions we wish to answer? This is related to the nature of proxies and sampling, discussed below. I would never argue against the fact that shell middens, megaliths, or single burials, as well as settlements, are all reflections of past activities and thus related to population. The point is that we cannot assume that they are comparably *equal* traces of past populations across time and space, and thus calculating a combined input is only correct if it can be demonstrated that there is no major variation between the different site categories during the period investigated.

This comes back to the base premise of dates as data, proposed by Rick (1987), which the respondents refer to. While it is true that most proponents of the method applied recognise potential problems with using radiocarbon dates, they mostly try to overcome these issues at the final step, that of the ^{14}C-dates, and do not consider the formation of the record in the first place. The relevant problems lie outside the immediate calculations of the statistics and are of a more fundamental character of *how* to use dates as data. If we take a step back and investigate possible *other reasons* for the structure of the summed dataset, we will recognize that many other explanations need to be considered (Crombé and Robinson, 2014). These explanations might be of local, regional, or general character and need to be considered as such. In some cases it might be possible to investigate some of these biases, and some of the proponents of the method have done a very thorough job of this for their specific case, such as testing the influence of different types of dates. An example is how Williams compares detrital charcoal dates and dated hearths (Williams, 2012, 579–580). The issue is shown by the respondents' hypothetical example of a stable SPD over a long time with a 300 year peak related to monument building. I agree with them that we cannot be sure whether this signifies an increase in population, or if the population is overestimated or underestimated. That is the entire point: we cannot know. We have no access to independent data that tells us how to "weigh" different kinds of sites against each other. What remains is that a passage grave is preserved better than a single grave, and that it might already from the onset represent a very different population signature. Thus, in cases of changes in site use, we have no way of evaluating the input related to the output, no matter how sophisticated our statistics.

2. Proxies, for proxies of proxies of population

The first aspect I wish to discuss is proxies, which I find a useful, or even indispensible tool in our archaeological tool box. I agree with the respondents that a proxy cannot be categorised as either valid or invalid. However, we do need to continuously question the validity of any proxy. I did not conclude that radiocarbon dates were not valid as a population proxy, but that they were a *poor* proxy in situations with great changes in economic and social organisation. Why? Let us look at the underlying assumptions that are behind ^{14}C-dates as proxies for population. The first assumption is that all people leave a comparable amount of material. If this is true, then that material is a proxy for population. The second assumption is that all the material is deposited in a similar manner that allow for equal preservation. If this is true, the preserved material is a proxy for the material left, which is a proxy for population. The third assumption is that sites from all periods and all types of material are equally excavated and dated. If this is true, the dates can be taken as a proxy for the material preserved, which is a proxy for material left, which is a proxy for population. We have to make an explicit argument that all these assumptions are approximately "true" in the specific cases, before ^{14}C-dates can be considered a *good* proxy. If this argument can be made, then the method applied in the studies of Shennan et al., 2013 and Timpson et al., 2014 are indeed good at assessing biases in the calibration process and in the random variation in the samples. In my article, I investigated some of these assumptions. The first assumption is that all material left represented an equal population, but does the construction of a passage grave really reflect the same population as a single grave? Before moving on, the authors need to validate this assumption. The second assumption is that the material from different periods is preserved equally well, which is not true: The shell middens reflect a much better preservation environment than other sites (especially since the shells themselves are dated), and they are not equally used throughout the period. Similarly, there are no graves from the late Ertebølle phase, and this suggests that there is a bias due to the manner of deposition of the dead and the preservation of these remains. The final assumption is that all sites are equally excavated and dated. Here I point out that as much as 47% of the dates in one study (the Jutland region in Hinz et al., 2012) stem from one research project. As the respondents point out, it is a fact that at times a researcher will get unexpected dates, but it is also clear that a research project oriented at shell middens used in the Mesolithic and Early Neolithic will primarily yield Mesolithic and Early Neolithic dates.

That different social and economic settings will need different proxies is made very clear by the respondents' own ice cream example. They suggest ice cream sales can serve as a proxy for murder rate. This is true if a prior investigation shows that they have related in the past and a relation can be established (in this case through temperature). No such direct link has yet been made for the ^{14}C-dates across the Mesolithic–Neolithic transition. However, it shows very elegantly that we need to establish our proxies carefully and case specifically, and that it is difficult to expand the same proxy across radical shifts in society and culture: We cannot assume that since ice cream sales is a good proxy for murder in one case, this is also true for other periods or regions, like present day Greenland, medieval London, or indeed Neolithic Europe.

3. Sampling and re-sampling

Timpson, Manning, and Shennan in their response point out that the use of ^{14}C-dates as a proxy has grown out of an approach of counting sites per archaeologically defined phase. This is true, and therefore the method carries over most of the problems this method entails. However, it also raises the question of sampling in archaeology. In statistics, a sample is taken from a known or well defined population (whether people, objects or events), which allows the researcher to make inferences on a subset of the known population, and then with some level of certainty, based on the number of samples, extend the results to the whole population. In archaeology we do not know the original population from which our sample is taken (since this is the object of our inquiry). This entails that we do not know *a priori* if our sample is representative. It is also difficult to "test", since we are unable to compare our sample to the original population. I am not arguing this should make us stop doing statistics, but that we need to consider this in

our sampling and re-sampling strategies. The papers of Shennan et al., 2013 and Timpson et al., 2014 make a great effort of re-sampling and testing their samples, that is the radiocarbon dates. However, as with the proxies above, the radiocarbon date is only the last sample of a series of samples. The date is a sample from a site, which is a sample of all excavated sites, which is a sample of all sites not destroyed (including unexcavated sites), which is a sample of all sites created in prehistory, which is a reflection of past population patterns. At each level a non-random selection is occurring, and we have little information on how this "sampling" is influencing our results. I will not argue, as some might, that the statistics thus become too problematic to use, but rather that if we wish to continue a specific investigation, we need to test every step of the sampling as thoroughly as the mentioned studies did the last step of samples (the dates). Some of the later steps can be sampled the same way as the radiocarbon dates, that is we can treat the dated sites (instead of the dates) or the research projects/researchers as "units" and draw sub-samples from these. The subsamples or subsets of data need to adhere to the same strict statistical treatment as the dates themselves. If the results are still robust and uniform in such subsets, we can say that modern research does not affect the result. In my article such statistics were not used, which is because the number of sites and the number of research projects were too limited (in contrast to the dates) to be relevant as "population" regarding the Law of Large numbers. Thus, it was simply not possible to draw samples large enough. Therefore, I subjectively suggested that when one researcher is responsible for 47% of the dates, it is not possible to estimate the effect of this project on the results. This means we cannot estimate the margin of error of the result, and thus we cannot make conclusions based on the samples.

However, this is not the most difficult obstacle, which is instead that of sampling the prehistoric site formation and the subsequent preservation of these. Here we do not have access to the parent population, and re-sampling is thus impossible. Therefore, careful archaeological evaluation and critical discussion of these factors are necessary in each and every case (every region, every period etc). This makes ^{14}C-dates as a proxy for periods undergoing dramatic changes (such as the Mesolithic–Neolithic transition), where these parent populations change their behaviour *a hopeless endeavour*, especially without the addition of investigation into the subsistence economy, settlement pattern and the preservation state of the datable material (as also demonstrated by Crombé and Robinson, 2014). This is completely lacking in the Shennan et al., 2013 and Timpson et al., 2014 studies, as well as other studies relating to the Neolithic–Mesolithic transition.

4. Conclusions

The use of radiocarbon dates as a proxy has grown in popularity over the recent years. This is problematic in any study where the link between the primary population and the sample is not clearly established. The problem is particularly relevant in any case where there are major changes in the site formation and site preservation between the periods investigated. If it cannot be demonstrated that the site formation, the preservation and the excavation and dating intensity are constant in relation to the period of investigation, the link between population and radiocarbon dates is not made, and the sample possibly unrepresentative. If a sample is unrepresentative, and it is not known how much or in which direction, no statistical treatment, however refined, can solve this.

The respondents place the obligation on me to come up with ways to salvage the dataset and improve it as a population proxy (be constructive). I find this a bit unfair, since I argue, here and in my article, that it is not possible. They want me to repair the unrepairable before I am allowed to point out that it is broken in the first place. I think the responsibility should be on those who wish to utilize the method to substantiate the assumptions underlying their proxies and test/discuss every step of the sampling process. In other words, the authors might be better served handling the critique in a constructive way instead of shooting at the messenger.

References

Crombé, P., Robinson, E., 2014. 14C dates as demographic proxies in Neolithisation models of northwestern Europe: a critical assessment using Belgium and northeast France as a case-study. J. Archaeol. Sci. 52, 558–566. http://dx.doi.org/10.1016/j.jas.2014.02.001.

Hinz, M., Feeser, I., Sjögren, K.G., Müller, J., 2012. Demography and the intensity of cultural activities: an evaluation of Funnel Beaker Societies (4200-2800 cal BC). J. Archaeol. Sci. 39, 3331–3340. http://dx.doi.org/10.1016/j.jas.2012.05.028.

Rick, J.W., 1987. Dates as data: an examination of the Peruvian pre-ceramic radiocarbon record. Am. Antiq. 52, 55–73.

Shennan, S., Downey, S.S., Timpson, A., Edinborough, K., Colledge, S., Kerig, T., Manning, K., Thomas, M.G., 2013. Regional population collapse followed initial agriculture booms in mid-Holocene Europe. Nat. Commun. 4, 2486. http://dx.doi.org/10.1038/ncomms3486.

Timpson, A., Colledge, S., Crema, E., Edinborough, K., Kerig, T., Thomas, M.G., Shennan, S., 2014. Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. J. Archaeol. Sci. 52, 549–557. http://dx.doi.org/10.1016/j.jas.2014.08.011.

Torfing, T., 2015. Neolithic population and summed probability distribution of 14C-dates. J. Archaeol. Sci. 63, 193–198.

Williams, A.N., 2012. The use of summed radiocarbon probability distributions in archaeology: a review of methods. J. Archaeol. Sci. 39, 578–589. http://dx.doi.org/10.1016/j.jas.2011.07.014.