

Research rebuttal paper uncovers misuse of Holocaust datasets

June 29 2021, by Debra Levey Larson



One of over 7,000 lists of names from concentration camps in the U.S. Holocaust Memorial Museum. This one is a handwritten list of Serbian and Croatian women who were deported to the Jasenovac concentration camp. Credit: United States Holocaust Memorial Museum

Aerospace engineering faculty member Melkior Ornik is also a mathematician, a history buff, and a strong believer in integrity when it comes to using hard science in public discussions. So, when a story

popped up in his news feed about a pair of researchers who developed a statistical method to analyze datasets and used it to purportedly refute the number of Holocaust victims from a concentration camp in Croatia, it naturally caught his attention.

Ornik is a professor in the Department of Aerospace Engineering at the University of Illinois Urbana-Champaign. He proceeded to study the research in depth and used the method to re-analyze the same data from the United States Holocaust Memorial Museum. Then he wrote a rebuttal paper debunking the researchers' findings.

Ornik's rebuttal is published in the same journal as the original article. He said the editor asked him to include a list of answers to some of the potential questions other scientists may have when they read his paper. A few weeks later, the journal placed a note on the original article stating that they do not endorse or share the authors' views, and recommended reading Ornik's paper.

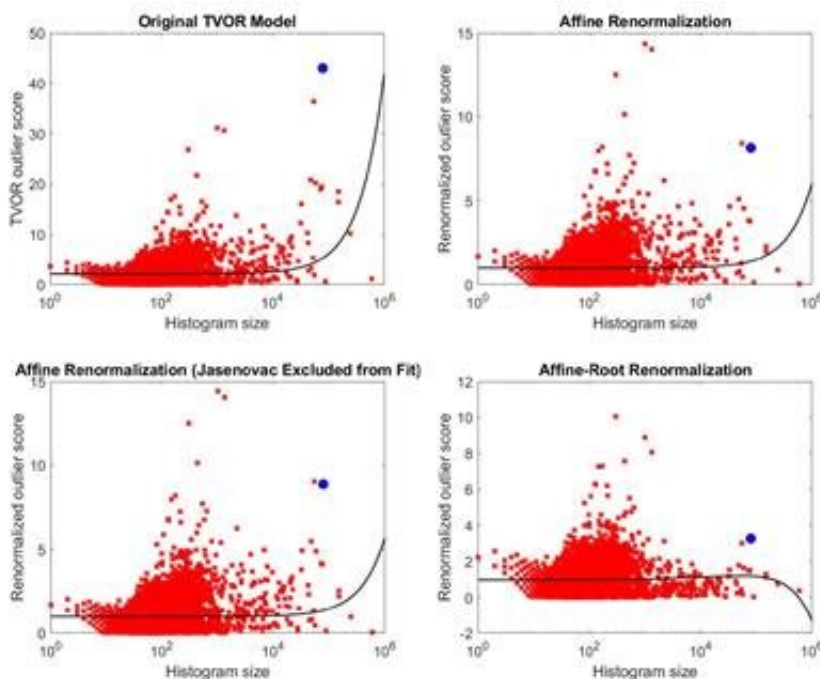
"As scientists, as engineers, I think it's our duty to correct flawed and faulty science," Ornik said. "There is so much effort to get the public and policymakers to believe in the science, that when a math expert says they have proof, it brings credence to the argument. But when their claims are demonstrably not true, it's not good for science and it's not good for society. That's why it's especially important for scientists to challenge false findings when we discover them."

According to Ornik, some individuals promote the view that concentration camps either didn't exist or weren't used to kill people, or that the currently widely accepted numbers of victims have been substantially inflated. Most historians do not take the claims seriously in light of vast available data and evidence.

"For the authors of the original paper to claim that they have found

mathematical proof that the list of victims of that camp was fabricated has obvious historical implications," Ornik said. "I think, to some extent the damage has already been done, but I felt the need to go on record with the assumptions, inaccuracies, and misuse of the raw museum data I found in the original research."

The paper Ornik responded to presents a novel method to identify anomalies across a set of histograms. Ornik said he does not dispute the merits of the method presented in the original paper, just its application to the Jasenovac [concentration camp](#).



Comparison of the original outlier identification model and three models derived from it. Due to the inapplicability of its assumptions to the considered dataset, the original model has no theoretical foundation. Three alternative models are less biased to size than the original model and produce opposing results. Credit: Melkior Ornik

Ornik became suspicious of the paper's conclusions because the researchers implied in one case that a smaller list naturally has a smaller outlier score, but they compared scores across victim list sizes to claim that the one related to Jasenovac, one of the biggest ones, was problematic.

"I started looking to see if there was some sort of a bias for the size and whether they were actually more likely to assign the flag of being problematic to a larger list or not. And it turns out, despite the authors' claims, they were," Ornik said. "The bigger lists are more likely to be computed to be problematic than the smaller lists when their method is applied to the data."

Ornik, who commonly uses similar statistical analysis in aerospace applications, explained another reason that their statistical argument doesn't work.

"When you look at data, a collection of anything, and you want to figure out an outlier—something that's different—you need to assume that all of the pieces of data come from the same source, the same distribution. Take a list of victims by birth year. It would yield a graph of the ages of each person. Say 10 percent are older than 70 years old. Now, that distribution wouldn't be true for a list of deported children, for example, because that list, by definition, is structurally different. It is also different from a list of everyone who has an identity card. Identity cards are issued only to people who are not children. Yet, the lists that these researchers worked with came from a multitude of sources and include lists of children, lists of people getting married, lists of prisoners of war—things that by definition cannot have come from the same distribution."

Another major error in the original paper, Ornik said, is that some duplicate lists were treated as two separate lists. This meant that

approximately 67 percent of their entire database was actually sub-lists of the larger list.

"The 7,000-plus lists published online by the Holocaust Museum are not curated," Ornik said. "For instance, there are two lists that contain exactly the same data; one is in Cyrillic and the other one uses the Latin alphabet. But they treated them as two separate lists. There are other lists that contain the same name, but there is no way of knowing if they are the same person or two different people born on the same day with identical names. They could have removed the very egregious errors in which a list is clearly duplicated but the rest, you would need access to the original historic data."

Both the original paper and Ornik's [paper](#), "Comment on 'TVOR: Finding Discrete Total Variation Outliers Among Histograms,'" are published in *IEEE Access*.

More information: Melkior Ornik, Comment on "TVOR: Finding Discrete Total Variation Outliers Among Histograms," *IEEE Access* (2021). [DOI: 10.1109/ACCESS.2021.3082900](https://doi.org/10.1109/ACCESS.2021.3082900)

Provided by University of Illinois Grainger College of Engineering

Citation: Research rebuttal paper uncovers misuse of Holocaust datasets (2021, June 29) retrieved 11 July 2024 from <https://phys.org/news/2021-06-rebuttal-paper-uncovers-misuse-holocaust.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.