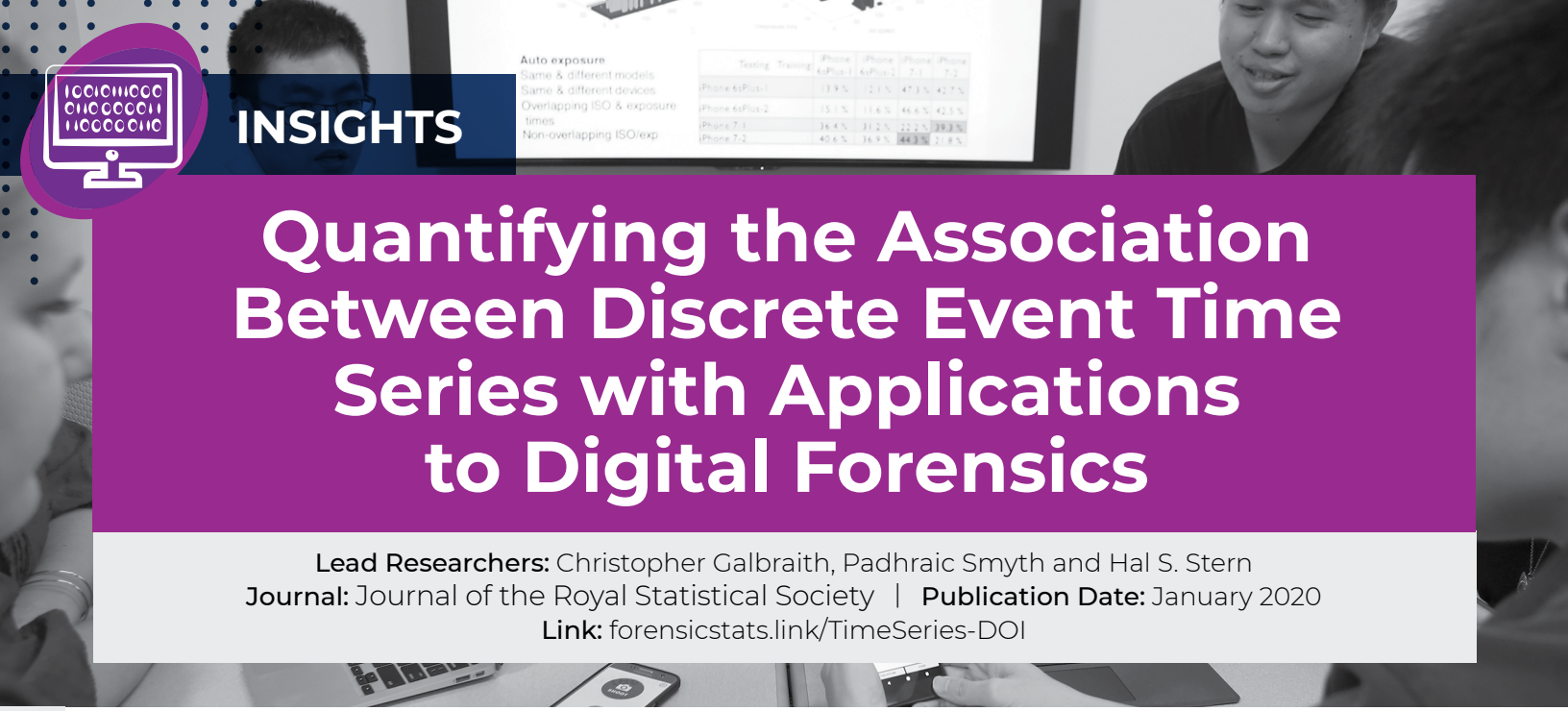

INSIGHTS

# Quantifying the Association Between Discrete Event Time Series with Applications to Digital Forensics

**Lead Researchers:** Christopher Galbraith, Padhraic Smyth and Hal S. Stern
**Journal:** Journal of the Royal Statistical Society | **Publication Date:** January 2020
**Link:** forensicstats.link/TimeSeries-DOI

## OVERVIEW

Digital devices provide a new opportunity to examiners because for every user event — like opening software, browsing online, or sending an email — an event time series is created, logging that data. Yet, using this type of user-generated event data can be difficult to correlate between two devices for examiners. The research team set out to quantify the degree of association between two event time series *both with and without population data.*

## THE GOALS

1. Investigate suitable measures to quantify the association between two event series on digital devices.
2. Determine the likelihood that the series were generated by the same source or by different sources — ultimately to assess the degree of association between the two event series.

## APPROACH AND METHODOLOGY

Researchers explored a variety of measures for quantifying the association between two discrete event time series. They used multiple score functions to determine the similarity between the series. These score functions were discriminative for same- and different-source pairs of event series.

The following methods for assessing the strength of association for a given pair of event series proved most accurate:

1. Constructing **score-based likelihood ratios (SLRs)** that assess the relative likelihood of observing a given degree of association when the series came from the same or different sources. This uses a population-based approach.
2. Calculating **coincidental match probabilities (CMPs)** to simulate a different-source score distribution via what the research team refers to as sessionized resampling when working with a single pair of event series. When a sample from a relevant population is not available, this method still produces accurate results.

**CONTACT**

**Website:** forensicstats.org **Email:** csafe@iastate.edu

## KEY TAKEAWAYS FOR PRACTITIONERS

1. The population-based approach of SLRs remains the preferred technique in terms of accuracy and interpretability.

2. The resampling technique using CMPs shows significant potential for quantifying the association between a pair of time event series, helping examiners determine the likelihood that two different time series were created by the same person, especially when no population sampling data is available.

3. With multiple-event series, combining these techniques could be valuable for pattern mining to determine which event series are associated with one another.

4. Developments in this area have the capacity to positively impact work in forensic and cybersecurity settings.

## NEXT STEPS

Both SLR and CMP techniques require more extensive study and testing before being used in practice by forensic examiners.

*Access the full research study to learn more:*

forensicstats.link/TimeSeries

All techniques that are described are implemented in the open-source R package **assocr**, available at forensicstats.link/AssocR.

## FUNDING

csafe
Center for Statistics and Applications in Forensic Evidence

*CSAFE is a publicly funded organization headquartered at Iowa State University. The National Institute of Standards and Technology (NIST) is one of the center's providers, supporting CSAFE as a nationally recognized Center of Excellence in Forensic Sciences, NIST Award # 70NANB15H176.*