pennsylvania
DEPARTMENT OF TRANSPORTATION

# Highway Incident Detection Timeline

FINAL REPORT

October 16, 2017

By: Joseph Coe (PI), Bechara Abboud (co-PI), Joseph Picone (co-PI), Siavash Mahvelati, and Aaron M. Gross

Temple University

COMMONWEALTH OF PENNSYLVANIA
DEPARTMENT OF TRANSPORTATION

CONTRACT # 4400011166
WORK ORDER # TEM 009

*Page Intentionally Left Blank*

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| FHWA-PA-2017-003-TEM WO 009 | | |

| 4. Title and Subtitle | 5. Report Date |
|---|---|
| Highway Incident Detection Timeline | October 16, 2017 |
| | **6. Performing Organization Code**<br>If Applicable |

| 7. Author(s) | 8. Performing Organization Report No. |
|---|---|
| **Joseph T. Coe (ORCID: 0000-0002-5404-2447), Bechara Abboud, Joseph Picone, Siavash Mahvelati, and Aaron M. Gross** | TEM-WO-009 |

| 9. Performing Organization Name and Address | 10. Work Unit No. (TRAIS) |
|---|---|
| Temple University<br>1947 N. 12th St.<br>Philadelphia, PA 19122 | **11. Contract or Grant No.**<br>**4400011166, TEM WO 009** |

| 12. Sponsoring Agency Name and Address | 13. Type of Report and Period Covered |
|---|---|
| The Pennsylvania Department of Transportation<br>Bureau of Planning and Research<br>Commonwealth Keystone Building<br>400 North Street, 6th Floor<br>Harrisburg, PA 17120-0064 | Final Report; 10/17/16 – 10/16/17 |
| | **14. Sponsoring Agency Code** |

**15. Supplementary Notes**
Technical Advisor: Jonathan Fleming - PennDOT, Section Chief, Emergency and Incident Management, (717)772-1771, jonfleming@pa.gov

**16. Abstract**
The Pennsylvania Department of Transportation (PennDOT) maintains information on road conditions through its Road Condition Reporting System (RCRS). PennDOT uses this technology for planning purposes and to report road closures, weather-related road conditions, and similar information to the general public via its official travel information services (511PA) and to outside agencies. The effectiveness of RCRS is dependent on the quality of information and the timeframe in which it is provided. PennDOT is consequently interested in the average time for notification of highway incidents from emergency dispatch centers. This information would support PennDOT in a number of ways: (1) Reduce the overall time to clear incidents and the time gap between when a highway closure occurs and when the public is informed; (2) Provide information to aid PennDOT in policy and decision making process related to traffic incident management; (3) Identify potential key elements and any critical missing information related to traffic incident management; and (4) Improve operation at statewide, regional, and district traffic management centers.

Based on the preceding discussion, the general objective of this project was to evaluate the highway incident detection timeline along a number of major highways in the Commonwealth of Pennsylvania, including Primary Interstate Highway I-76, I-80, I-81, and I-95, and Auxiliary Interstate Highway I-78 and I-83. This was accomplished by comparing emergency dispatch records and the PennDOT RCRS for incidents along the aforementioned highways over several years. These databases were normalized, pre-processed, and then examined using an integrated approach. Based on established data pairings, statistical analyses were performed to evaluate the notification time difference across a number of counties in Pennsylvania.

| 17. Key Words | 18. Distribution Statement |
|---|---|
| Road Condition Reporting System, RCRS, Highway Incident, Incident Detection, Emergency Dispatch Centers, 911 Dispatch, Traffic Incident Management, Road Closure, Lane Restrictions | No restrictions. This document is available from the National Technical Information Service, Springfield, VA 22161 |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 57 | **N/A** |

*Page Intentionally Left Blank*

## Disclaimer

This work was sponsored by the Pennsylvania Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration, U.S. Department of Transportation, or the Commonwealth of Pennsylvania at the time of publication. This report does not constitute a standard, specification, or regulation.

## Acknowledgements

The financial and technical support provided by the Pennsylvania Department of Transportation (PennDOT) for this project is gratefully acknowledged. The authors would like to extend their sincere gratitude to the project manager Mr. Michael Bindie for all his support throughout the project. The authors also appreciate all the time, helpful comments, and technical oversight offered by the technical advisors from the PennDOT Bureau of Maintenance and Operations, Mr. Jonathan Fleming, Mr. John Smyser, Mr. William Davenport, and Mr. Daniel Whetzel. The authors would also like to thank Noah M. Capp, Tameem Muhammad Ahsan, and Vineetha Mathew from Temple University for their efforts in processing the emergency call center data.

## Table of Contents

## List of Figures

## List of Tables

# PennDOT Research Executive Summary

## Highway Incident Detection Timeline

**QUICK INFO:**

**Start Date:**
October 17, 2016

**End Date:**
October 16, 2017

**Funding:**
$126,000.00

**Report Date:**
October 16, 2017

**Performing Organization:**
Temple University

*Bureau of Planning & Research*

(p) 717-787-5796
(f) 717-783-9152

Email: RA-penndotlibrary@pa.gov

### BACKGROUND

The Pennsylvania Department of Transportation (PennDOT) maintains information on road conditions through its Road Condition Reporting System (RCRS). The effectiveness of RCRS is dependent on the quality of information and the timeframe in which it is provided. PennDOT is consequently interested in the average time for notification of highway incidents from emergency dispatch centers. This information would support PennDOT in a number of ways: (1) Reduce the overall time to clear incidents and the time gap between when a highway closure occurs and when the public is informed; (2) Provide information to aid PennDOT in policy and decision making process related to traffic incident management; (3) Identify potential key elements and any critical missing information related to traffic incident management; and (4) Improve operation at statewide, regional, and district traffic management centers. Based on the preceding discussion, the general objective of this project was to evaluate the highway incident detection timeline along a number of major highways in the Commonwealth of Pennsylvania, including Primary Interstate Highway I-76, I-80, I-81, and I-95, and Auxiliary Interstate Highway I-78 and I-83. This was accomplished by comparing emergency dispatch records and the PennDOT RCRS for incidents along the aforementioned highways over several years. Based on established data pairings, statistical analyses were performed to evaluate the notification time difference across a number of counties in Pennsylvania.

### FINDINGS

An integrated framework using a graphical user interface and auomated data pairing was developed to match records between the PennDOT RCRS and county 911 databases. Using this framework, the statistical variability in noitifcation time was evaluated for counties that provided 911 call center data.

### RESULTS

The major results from this research project included the following:

- The time difference between RCRS entries and 911 call records is not normally distributed, necessitating the use of the median value to accurately describe the central tendency in the statistical data.
- The overall median time difference between matched RCRS and 911 entries was 12 minutes. Based on an upper quartile estimate, 75% of all matched records have a time difference less than approximately 28 minutes.
- Montgomery, Delaware, Cumberland, and Bucks counties all exhibit time differences shorter than the median computed for all counties.
- Generally, counties with smaller time differences tended to exhibit smaller variability in their distributions. Overall, there was large scatter in the results when comparing extremes in county-level results.
- Areas of the Commonwealth traversed by I-95, I-83, the southernmost section of I-81, and I-76 generally exhibited smaller time differences. Conversely, the sections of I-81 heading north from Lebanon County (and the only counties for which 911 data was procured along I-80 in this study) exhibited larger time differences.

### COUNCLUSIONS AND RECOMMENDATIONS

Since the RCRS system is a major tool with which PennDOT monitors highways, the statistical results from this study can aid PennDOT in developing best practices for policy and procedural decisions related to traffic incident management, which can improve operation at the statewide, regional, and district traffic management centers. Estimates of the time necessary for PennDOT to receive notification of highway incidents across the Commonwealth of Pennsylvania can allow PennDOT to better allocate resources and is the first step in minimizing the time gaps for highway closures in response to emergencies. This information would also allow PennDOT to identify potential key elements and any critical missing information related to traffic incident management across the Commonwealth of Pennsylvania. For example, one major lesson from this study is that there is significant difference between various 911 communication center operations, reporting procedures, and CAD systems across the Commonwealth. In many ways, this decentralization increases the difficulty of establishing links between 911 call data and existing PennDOT RCRS records. Increased integration of datasets among the various stakeholders involved with highway incidents can begin to address some of these issues and improve operational emergency management of highways in the Commonwealth of Pennsylvania. PennDOT can also implement the results of study by periodically deploying the integrated framework documented in this report to catalog any statistical changes in traffic incident timeline.

# 1 Introduction & Project Objectives

The Pennsylvania Department of Transportation (PennDOT) maintains information on road conditions through its Road Condition Reporting System (RCRS). PennDOT uses this technology for planning purposes and to report roadway and bridge closures, weather-related road conditions, lane restrictions, highway conditions, and construction activities to outside agencies and the general public. RCRS is available from any PennDOT intranet location and various locations in other agencies, which allows various personnel and bureaus to enter data and make it available for district engineering offices, the Pennsylvania Emergency Management Agency, and the Pennsylvania State Police. RCRS also forms the backbone for PennDOT's official travel information services (i.e., 511PA), which allows travelers to plan their routes and law enforcement and emergency response agencies rely to establish the quickest and most direct route when responding to an incident or event.

The effectiveness of RCRS is dependent on the quality of information and the timeframe in which it is provided. PennDOT is consequently interested in understanding how long it takes on average time to be notified of highway incidents with respect to when emergency dispatch centers (i.e., 911 call centers) are notified. This information would support PennDOT in a number of ways: (1) Reduce the overall time to clear incidents and reduce the time gap between when a highway closure occurs and when the public is informed; (2) Provide information to aid PennDOT in policy and decision making process related to all aspects of traffic incident management; (3) Identify potential key elements and any critical missing information related to traffic incident management across the Commonwealth of Pennsylvania; and (4) Improve operation at statewide, regional, and district traffic management centers.

Based on the preceding discussion, the general objective of this project was therefore to evaluate the highway incident detection timeline along a number of major highways in the Commonwealth of Pennsylvania. More specifically, efforts on this project focused on determining the average timeline for when the Pennsylvania Department of Transportation (PennDOT) is notified of incidents requiring highway closures along Primary Interstate Highway I-76, I-80, I-81, and I-95, and Auxiliary Interstate Highway I-78 and I-83. This was accomplished by comparing emergency dispatch (911) records and the PennDOT Road Condition Reporting System (RCRS) for incidents along the aforementioned highways over the previous several years. These databases were normalized, pre-processed, and then examined using an integrated approach. This resulted in the establishment of linkages between the two databases for those entries that described the same highway incident. Based on these pairings, statistical analyses were performed to evaluate the time latency across a number of counties in Pennsylvania. The following sections present a review of data acquisition and post-processing, an in-depth discussion of efforts involved to develop statistical estimates of notification latency, analysis of the latency results, and efforts to curate the data.

# 2 Project Data

The first step to achieve the objectives of this project was to identify and procure the data required to perform statistical analysis. As noted previously, two sources of information were considered simultaneously in order to determine the time delay for a given highway incident along Primary Interstate Highway I-76, I-80, I-81, and I-95, and Auxiliary Interstate Highway I-78 and I-83: (1) PennDOT RCRS records; and (2) time response logs kept at the 911 call center of the county where a highway incident occurred. Data collection efforts therefore consisted of requesting RCRS logs for the highways of interest from the PennDOT project team and contacting county 911 call centers throughout the Commonwealth of Pennsylvania.

## 2.1 Data Acquisition

PennDOT RCRS data was provided via email to the Temple research team on Tuesday, November 22, 2016 by the PennDOT project team. The data consisted of a zip file with a Microsoft Excel spreadsheet generated by the PennDOT RCRS software that detailed all reported incidents along the highways of interest in this study. Therefore, acquisition of all RCRS data necessary for successful completion of the project objective was completed as of Tuesday, November 22, 2016.

Based on the original RCRS data provided by PennDOT, a total of 20,950 entries occurred along Primary Interstate Highway I-76, I-80, I-81, and I-95, and Auxiliary Interstate Highway I-78 and I-83 between the dates of 01/01/2013 and 11/22/2016. These entries corresponded to multiple event "types", including Lane Restriction, Shoulder Closed, Traffic Disruption, Ramp Restriction, Ramp Closure, Closed, Open, and Residual Delays. Given this initial database from PennDOT and a review of the routes of the project highways, a total of 37 counties (of the 67 counties throughout Pennsylvania) were identified as pertinent to this project (Table 1). However, based on conversations between the Temple research team and the PennDOT project team at an in-person meeting (Friday, February 3rd, 2017), filter criteria was established to refine the RCRS data. This allowed the project to focus on the RCRS incident types most important to PennDOT (i.e., entries with status of "Closed", "Lane Restriction", "Ramp Closure", and "Ramp Restriction"). This changed the total number of pertinent PennDOT RCRS entries from 20,950 entries to 8,984 entries (20 of which were duplicate entries with different start and ending counties) and reduced the number of pertinent counties to 29 (Table 1).

Since all RCRS data was acquired early within the project, all remaining efforts related to data acquisition were concentrated on the 911 calls. This necessitated contact with all counties of interest, which the Temple research team initiated prior to refining the RCRS data. Therefore, the Temple research team reached out to all 37 counties initially identified at the start of the project. Many of the counties contained more than one of the highways of interest in this study. Contact with each of the counties was attempted in three different manners:

(1) Direct emails/phone calls to county dispatch centers by the Temple research team. The PennDOT project team provided a list of contacts at all counties of interest for the project. These contacts typically included the Director of Emergency Communications, or a similar title, that coordinated 911 communications for a given county. The Temple research team initiated contact with these individuals via email on Wednesday, November 23, 2016 and continued regularly as various counties responded or failed to respond.

(2) Direct emails/phone calls to county dispatch centers by the PennDOT project team. At the onset of the project during the kick-off meeting, the Temple research team was informed that the PennDOT project team had initiated contact via email with the counties to request data be provided. This email was submitted to the counties on Tuesday, October 18, 2016. A copy of the email and a list of counties and contact information were provided to the research team after the kick-off meeting. The PennDOT project team continued to contact various counties periodically as the project progressed based on any difficulties encountered by the Temple research team.

(3) Submission of official right-to-know requests (RTK). Generally speaking, the 911 time response logs are considered public records and can be accessed by submitting an official RTK. A number of counties had interfaces through their official county websites to perform this task. A form was submitted directly via the website, emailed to a contact person at the county, or printed out and mailed to the county. The nature of the form differed slightly depending on the county (e.g., Figure 1), but generally detailed the nature of the data requested. A total of 25 county RTK requests were submitted throughout the course of the project.

In some cases (Philadelphia, Delaware, Columbia, Franklin, and Monroe counties), data acquisition efforts included interfacing with contacts at the Pennsylvania State Police (PSP) and/or submitting RTK requests through PSP since their agency handles 911 calls regarding highways in the aforementioned counties. Additionally, in at least one county (Northumberland), the Temple research team was informed that a nearby county (Union) communication center handled their 911 dispatching for the pertinent highway along that stretch of the Commonwealth. Generally, the format of the provided 911 data varied substantially both in terms of file format/structure (e.g., Excel files, pdf files, etc.) and range of information included in the entry fields (Figure 1). Some counties were able to provide a lot more details [e.g., multiple time stamps, (latitude, longitude) coordinates] for each incident, while other datasets proved to be missing critical information (e.g., descriptive address). This reflected the wide range of computer aided dispatch (CAD) systems (in terms of both data input and export capabilities) and record-keeping policies in place across the Commonwealth. As noted by a few of the county 911 communication center directors, a number of counties are still using older CAD systems while some others have recently updated their systems, and a number of them are transitioning to different systems.

# COUNTY OF BERKS, PENNSYLVANIA

## Office of Open Records


Services Center, 13th Floor
633 Court Street
Reading, PA 19601

Phone: 610.478.6136
Fax: 610.478.6293
E-mail: chiefclerk@countyofberks.com

Christian Y. Leinbach, Commissioner Chair
Kevin S. Barnhardt, Commissioner
Mark C. Scott, Commissioner

Maryjo Gibson, Chief Clerk / Open Records Officer
Christine M. Sadler, Esq., County Solicitor


**UNIFORM RIGHT-TO-KNOW REQUEST FORM**

**DATE REQUESTED:** 09/09/2016

**REQUEST SUBMITTED BY:** [E-MAIL] U.S. MAIL FAX IN-PERSON

**NAME OF REQUESTOR (Optional):** Siavash Mahvelati Shams Abadi

**STREET ADDRESS (Optional):** 1947 N 12th St

**CITY/STATE/ZIPCODE/COUNTY (Required):** Philadelphia/PA/19122/Philadelphia

**TELEPHONE (Optional):** (215)-992-4180

**E-MAIL (Optional):** smahvelati@temple.edu

**RECORDS REQUESTED:**
**Provide as much specific detail as possible so the agency can identify the information.*

As part of a Temple University research project regarding highway accidents, dispatch information from 911 calls over the previous five (5) years (starting from 2011) is requested. Ideally, this information would be provided in a searchable electronic format such as a spreadsheet.

**DO YOU WANT COPIES?** [YES] or NO

**DO YOU WANT TO INSPECT THE RECORDS?** [YES] or NO

**DO YOU WANT CERTIFIED COPIES OF RECORDS?** YES or [NO]

**RIGHT TO KNOW OFFICER:**

**DATE RECEIVED BY THE AGENCY:**

****Public bodies may fill anonymous verbal or written requests. If the requestor wishes to pursue the relief and remedies provided for in this Act, the request must be in writing and the requestor's name and address become required. (Section 702.)*

*****Written requests need not include an explanation why information is sought or the intended use of the information unless otherwise required by law. (Section 703.)*



**Figure 1. Example of right-to-know (RTK) request submitted to Berks County.**

| Date and Tim | Address Nur | Dir Pr | Feature Name | Dir St | Feature Typ | Mu | Type Coc | Eng Descript | Subtype Eng |
|---|---|---|---|---|---|---|---|---|---|
| 20110101014914ES | NULL | NULL | NULL | NULL | NULL | SWT | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | NO INJURIES |
| 20110101021901ES | 212 | NULL | PINE | NULL | ST | STL | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | NO INJURIES |
| 20110101025741ES | NULL | NULL | NULL | NULL | NULL | SWT | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | NO INJURIES |
| 20110101031200ES | NULL | NULL | GRAYSON | NULL | RD | SWT | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | NO INJURIES |
| 20110101081722ES | NULL | NULL | NULL | NULL | NULL | LDT | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | NO INJURIES |
| 20110101115351ES | NULL | W | 322 | NULL | ROUTE | DRY | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | NO INJURIES |
| 20110101170242ES | 2300 | NULL | COLONIAL | NULL | RD | LPT | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | NO INJURIES |
| 20110101183913ES | NULL | NULL | NULL | NULL | NULL | SUS | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | NO INJURIES |
| 20110101191731ES | NULL | NULL | NULL | NULL | NULL | DRY | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | ENTRAPMENT / CONFINEMENT |
| 20110101193558ES | NULL | NULL | NULL | NULL | NULL | HBG | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | UNKNOWN INJURIES |
| 20110102083223ES | NULL | NULL | NULL | NULL | NULL | SUS | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | NO INJURIES |
| 20110102103500ES | NULL | NULL | NULL | NULL | NULL | STL | P131 | TRAFFIC / TRANSPORTATION ACCIDENT | ENTRAPMENT / CONFINEMENT |

**Montgomery County Department of Public Safety**

**Incident Summary Report for: 01/01/2016 to 12/31/2016**

| Date | Time | Event Type | Location | Latitude | Longitude | Municipality |
|---|---|---|---|---|---|---|
| 01/01/2016 | 23:41:56 | VEHICLE ACCIDENT | 35 RT309 EXPY NB | 40.118834 | -75.203594 | Springfield Township |
| 01/02/2016 | 16:25:52 | VEHICLE ACCIDENT / UNKNOWN INJURIES | 3290 SCHUYLKILL EXPY WB | 40.076246 | -75.363832 | Upper Merion Township |
| 01/02/2016 | 16:47:11 | VEHICLE ACCIDENT | 3388 SCHUYLKILL EXPY WB | 40.017512 | -75.217020 | Lower Merion Township |
| 01/02/2016 | 16:56:57 | VEHICLE ACCIDENT | 3370 SCHUYLKILL EXPY WB | 40.034470 | -75.242139 | Lower Merion Township |
| 01/02/2016 | 19:20:57 | VEHICLE ACCIDENT / UNKNOWN INJURIES | 1814 RT422 WB | 40.186206 | -75.518104 | Upper Providence Township |
| 01/02/2016 | 19:24:16 | VEHICLE ACCIDENT / UNKNOWN INJURIES | 1806 RT422 WB | 40.196003 | -75.526027 | Limerick Township |

**Highway Accidents on Interstate 81, Interstate 83 and Interstate 76**

10/24/2016
1:44:27 PM

| Received Date/Time | Location |
|---|---|
| 01/02/2013 12:47:12 | HARRISBURG PK/I 81 S ONRAMP EXIT 52B ,MX CU |
| 01/02/2013 13:33:24 | I 81 N 567 EXIT 57 MECHBG RT 114 ,SS CU |
| 01/02/2013 13:57:20 | I 81 N 619 TO 650 ,EP CU;MM 64 |
| 01/02/2013 15:28:47 | I 81 N 567 TO 587 ,SS CU;MM 57.5 |
| 01/02/2013 15:38:55 | I 81 N 619 TO 650 ,EP CU ;MM63 |
| 01/03/2013 01:08:41 | I 81 S 590 TO 607 ,HM CU; 60.4 |
| 01/03/2013 02:45:35 | I 81 N 608 EXIT 61 WERTZVILLE RD ,HM CU |
| 01/03/2013 15:40:28 | I 81 S 282 EXIT 29 KING ST RT 174 ,ST CU |

**Figure 2. Examples of 911 communication center logs provided by (top) Dauphin County (middle) Montgomery County, and (bottom) Cumberland County.**

Table 1 provides a summary of the final outcomes of the data acquisition efforts in this study. The Temple research team was able to acquire pertinent 911 call data from 17 counties out of the 29 counties with entries in the filtered RCRS database (Table 1). As a whole, these counties combined to provide a total of 1,015,743 entries in the 911 call database over a time period ranging from the start of 2013 to late 2016. These 17 counties also represented 50.8% of the total number of filtered RCRS entries (i.e., 4,561 of 8,984). The bulk of the remaining data was composed of 911 calls from Philadelphia county (3,834 entries for a total of 42.7% of the RCRS data), which was incapable of directly providing the data because highway-related calls are dispatched by PSP in that county. Efforts to acquire the data from PSP for Philadelphia County proved unsuccessful. Apart from Philadelphia County, the Temple research team was able to acquire 911 data from what amounts to 88.6% of the pertinent RCRS entries (i.e., 4,561 out of 5,150, excluding RCRS entries from Philadelphia County). Philadelphia County has a high population density relative to other Pennsylvania counties and its urban setting provides a more

extensive traffic camera network. However, nearby counties where the Temple research team was able to acquire data (e.g., Delaware and Montgomery) share some similarities to Philadelphia County regarding population and traffic camera density. So despite the lack of data from Philadelphia County, the Temple research team expects that the overall results from this study will be generally representative of the study area in this project and conditions across the Commonwealth of Pennsylvania as a whole.

| No. | RMDS | County | Provided Data | Explanation | 911 Entries (#) | RCRS Entries (#) | RCRS Entries (%) | Provided GPS | GPS Comments | Method of Pairing |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 006 | Berks | Y | - | 756901 | 88 | 1.0% | N | No explanation provided | Manual & GUI |
| 2 | 009 | Bucks | Y | - | 5012 | 492 | 5.5% | Y | GPS provided for a very small portion of data | GUI |
| 3 | 013 | Carbon | N | Data not readily available | - | 23 | 0.3% | - | - | - |
| 4 | 014 | Centre | Y | - | 18831 | 86 | 1.0% | N | No explanation provided | GUI |
| 5 | 016 | Clarion | N | Data not readily available | - | 51 | 0.6% | - | - | - |
| 6 | 017 | Clearfield | N | Not responsive | - | 151 | 1.7% | - | - | - |
| 7 | 018 | Clinton | N | Not responsive | - | 89 | 1.0% | - | - | - |
| 8 | 019 | Columbia | N | Data not readily available | - | 41 | 0.5% | - | - | - |
| 9 | 021 | Cumberland | Y | - | 4291 | 311 | 3.5% | N | GPS unavailable | Manual & GUI |
| 10 | 022 | Dauphin | Y | - | 48644 | 492 | 5.5% | N | No explanation provided | - |
| 11 | 023 | Delaware | Y | - | 47415 | 926 | 10.3% | Y | - | GUI |
| 12 | 028 | Franklin | Y | - | 4230 | 83 | 0.9% | N | No explanation provided | GUI |
| 13 | 033 | Jefferson | N | Not responsive | - | 49 | 0.5% | - | - | - |
| 14 | 035 | Lackawanna | Y | - | 6331 | 89 | 1.0% | Y | - | Manual & GUI; Automatic w/ GPS |
| 15 | 038 | Lebanon | Y | - | 23644 | 105 | 1.2% | N | Not responsive | Manual & GUI |
| 16 | 039 | Lehigh | Y | - | 4857 | 148 | 1.6% | N | GPS unavailable | Manual & GUI |
| 17 | 040 | Luzerne | Y | - | 26611 | 122 | 1.4% | N | GPS unavailable | Manual & GUI |
| 18 | 043 | Mercer | N | Not responsive | - | 43 | 0.5% | - | - | - |
| 19 | 045 | Monroe | N | RTK rejected | - | 71 | 0.8% | - | - | - |
| 20 | 046 | Montgomery | Y | - | 7446 | 890 | 9.9% | Y | - | Manual & GUI |
| 21 | 047 | Montour | N | RTK rejected; Referred to PSP | - | 18 | 0.2% | - | - | - |
| 22 | 048 | Northampton | Y | - | 2743 | 60 | 0.7% | N | GPS unavailable | Manual & GUI |
| 23 | 049 | Northumberland | N | Referred to Union County | - | 26 | 0.3% | - | - | - |
| 24 | 051 | Philadelphia | N | RTK rejected; Referred to PSP; PSP RTK also rejected | - | 3834 | 42.7% | - | - | - |
| 25 | 054 | Schuylkill | Y | - | 5635 | 75 | 0.8% | N | GPS unavailable | Manual & GUI |
| 26 | 058 | Susquehanna | Y | - | 685 | 52 | 0.6% | Y | GPS manually determined | Manual & GUI; Automatic w/ GPS |
| 27 | 060 | Union | N | Not responsive | - | 27 | 0.3% | - | - | - |
| 28 | 061 | Venango | Y | - | 5052 | 30 | 0.3% | N | No explanation provided | GUI |
| 29 | 067 | York | Y | - | 47415 | 512 | 5.7% | Y | - | GUI |

**Table 1. Final status of data acquisition for all counties.**

## 2.2 Data Pre-Processing and Normalization

Given that the goal of the study was to develop an integrated framework that matched 911 data to PennDOT RCRS entries, it was important to understand what information in the 911 data logs most effectively paired with the PennDOT RCRS records. However, as anticipated, the 911 call

center data was provided in a number of different file formats (e.g., Excel files, pdf files, etc.) and with a wide range of information included in the entry fields (Figure 1). Therefore, extra care was necessary to normalize data in a manner that would allow the Temple research team the ability to definitively match entries between the two databases.

As part of initial efforts to pre-process and normalize the data, the research team went about manually pairing a small subset of PennDOT RCRS records. Initially, this consisted of matching the first 100 records in the PennDOT RCRS to the corresponding county 911 logs. Since the typical definitive match rate was approximately 1/3 of the examined records, this was expanded to approximately 300 records so that the first 100 matches were encountered. Data from the following counties was included in the first 100 matches: Lehigh, Luzerne, Lebanon, Berks, Lackawanna, Schuylkill, Northampton, and Susquehanna. This exercise was performed prior to the February in-person meeting (Friday, February 3rd, 2017) between the Temple research team and PennDOT project team. During this meeting, a number of pertinent questions were answered regarding filtering of the PennDOT RCRS data. As a result, the aforementioned pilot study was repeated with the new filter criteria (i.e., entries with status of "Closed", "Lane Restriction", "Ramp Closure", and "Ramp Restriction"), which changed the total number of pertinent PennDOT RCRS entries from 20,950 entries to 8,984 entries. This changed the entries that were examined to determine first 100 matches.

The goal of this initial data pairing pilot study was to compare the nature of the data provided by the RCRS in relationship to records provided by the different 911 call centers. This would allow the research team to identify what information was typically provided by most call centers and what data would prove the most useful in pairing between the 911 and RCRS databases. Data pre-processing efforts could then focus on ensuring the necessary data was in place as the integrated framework was developed for pairing efforts.

Based on the manual pairing efforts from the initial pilot study, the following parameters were noted as critical in being able to confidently identify matches between PennDOT RCRS and county 911 entries: Date/timestamp, GPS coordinates, location, and incident type. This set of information would allow filtering to occur initially on date/time by defining a threshold time period around each PennDOT RCRS entry. For manual pairing, the location field in the spreadsheet would then allow filtering to take place by mile marker along the highway of interest. This process proved accurate but labor intensive and inefficient for scale up to the large number of entries in the PennDOT RCRS data. As noted in later sections of the report, automated pairing efforts were considered as they would have the benefit of using GPS coordinates to trace the location of entries in the 911 county records. The PennDOT RCRS records already included this information but it was not included in the majority of 911 call center data (Table 1). However, in cases where it was available, this data was considered highly desirable and necessary to include in normalization efforts due to the time savings associated with an automated pairing algorithm. The final field in the normalization process was incident

type. This proved beneficial because for a number of counties, the 911 communication center data was not pre-filtered for the necessary incident types highlighted by PennDOT at the February in person meeting and March conference call (Thursday, March 16, 2017).

Once the specified normalized fields were determined, efforts focused on developing an efficient algorithm for developing normalized datasets for all county 911 records received. Figure 3 presents an example of a normalized data file for Susquehanna County. For the normalization process, the original county records were placed into a single Microsoft Excel® file. If the original records were spread across multiple files or sheets in a single spreadsheet file, all data was manually sorted into the same file. Any files in a .pdf or .txt format were read into a single Excel® file. These Excel® files were then parsed and all column headers were removed. The file was then revised using a custom developed script written in the programming language Python. The code utilizes the "Pandas" Python library to read and write Excel® files and manipulate the rows and columns of the data. Rules were developed for each individual county to format the pertinent columns within each spreadsheet in a consistent manner. For example, after each county Excel® spreadsheet was passed through the Python code, all normalized times were in the format YYY-MM-DD HH:MM:SS no matter what was provided in the original source format. When provided (or manually resolved), the latitude and longitude of the GPS coordinates were formatted to 6 decimal places (approximately 0.1 m resolution). Mile markers, if present in the location fields of the original data files, were identified automatically by the Python code and reported to the nearest 0.1 mile. The textual information (including any street addresses if applicable) was also passed along to the normalized data spreadsheet. Finally, if the county listed incident type, it was identified by the Python code and passed directly to the normalized data file with no further processing. In cases where some of the desired normalized information was not provided, the columns were still included in the revised normalized spreadsheet but were left blank for potential future updates with revised data. The remaining columns in the normalized data represent the raw records before normalization, and as such vary from county to county. This was included so that the raw, unmodified data would be conveniently available when reviewing the effects of the normalization process. The final step in the Python code was to write the newly formatted data file to an Excel® data file. This pre-processing ensured all the data was normalized into a consistent format to increase the efficiency and effectiveness of the pairing efforts.

| Time | Lat | Long | Location | Incident Type | Date/Time | Location | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|
| 2014-01-02 17:08:16 | 41.658700 | -75.667000 | 207.0 I81 S | | 1/2/2014 17:08:16 | 207 I81 S | 41.6587 | -75.667 |
| 2014-01-03 11:29:30 | 41.959570 | -75.736690 | 230.0 I81 N | | 1/3/2014 11:29:30 | 230 I81 N | 41.95957 | -75.73669 |
| 2014-01-03 13:16:41 | 41.917370 | -75.731290 | 227.0 I81 S | | 1/3/2014 13:16:41 | 227 I81 S | 41.91737 | -75.73129 |
| 2014-01-03 13:21:52 | 41.866380 | -75.707680 | 223.0 I81 S | | 1/3/2014 13:21:52 | 223 I81 S | 41.86638 | -75.70768 |
| 2014-01-03 14:48:50 | 41.691236 | -75.680276 | 209.8 I81 N | | 1/3/2014 14:48:50 | 209 .8 I81 N | 41.691236 | -75.680276 |

**Figure 3. Example of normalized logs for Susquehanna County.**

## 2.3 Integrated Framework for Pairing Databases

As noted in previous sections, significant efforts were involved in acquiring data from 911 call centers and pre-processing the resulting data into a consistent format. Once the data was normalized an integrated framework was developed to identify matches between incidents in the PennDOT RCRS database and those in a county 911 communication center dataset. This integrated framework was continually revised as the project progressed and the data available to the Temple research team evolved. Generally, all attempts at developing this integrated framework resulted in a similar overall approach: (1) Identify an RCRS entry of interest and identify the county in which the entry was located; (2) Load all records from county of interest; (3) Pre-filter county 911 records to remove any entries unrelated to highway incidents; (4) Identify a threshold timeframe to filter county 911 entries and remove unlikely matches based on time alone (e.g., any entries greater than three hours and earlier than one hour from the RCRS entry timestamp); (5) Use location information (e.g., GPS, mile marker information, location descriptors, etc.) to identify most likely matches to RCRS entry; (6) Develop final criterion to select among most likely matches (if more than a single county 911 entry is a likely candidate). The main difference between different iterations of the framework was the manner in which location information was integrated into the pairing process. The following sections describe the various iterations of the integrated framework, all of which were used in some manner to pair data between PennDOT RCRS and county 911 data.

### *2.3.1 Manual Efforts*

The previous section on data normalization discussed initial manual efforts to pair data. These efforts highlighted the level of inefficiency present in this overall framework when GPS coordinates were not available. Matching between the databases was largely slowed in this case by the need to manually interpret mile markers and/or location descriptor information to deduce the location of the 911 entry and compare to the location information from GPS coordinates in the RCRS data. This often necessitated conversion of the GPS coordinates into a more qualitative descriptor that could be manually matched to the mile marker and/or location description information field in the 911 data or vice-versa. Additionally, the RCRS and 911 data were still in different Excel® spreadsheets, which necessitated having to switch windows repeatedly during the manual pairing process.

### *2.3.2 Automated Efforts*

Given the inefficiency of manual pairing, the Temple research team pursued an integrated framework based on an automated algorithm using the Python programming language. This approach used scripts to parse through the county 911 and PennDOT RCRS for time, location,

and incident type information. The tool developed a list of potential matching county 911 entries for each RCRS entry. This list was then manually reviewed for selection of the most likely match between county 911 entries and RCRS entries. The algorithm implemented GPS coordinates to filter the potential county 911 matching entries based on location. Direct comparison of GPS coordinates allowed rapid consideration of a distance threshold for filtering potential matching entries in the 911 data for a given RCRS entry. PennDOT RCRS records provided GPS information and two counties (Susquehanna and Lackawanna) also included it with their initial responses to the 911 call data requests. Comparison of the resulting matches using this version of the automated framework with GPS coordinates matched well with efforts from manual pairing for the aforementioned counties. However, the majority of the 911 call centers were unable to provide GPS data, despite repeated efforts to procure such data. As a result, the GPS-based iteration of the integrated framework was not implemented on the remaining counties to match records between 911 and RCRS databases.

### *2.3.3 Graphical User Interface*

Since GPS data was unavailable for the vast majority of the county 911 records, the Temple research team focused their efforts on developing an iteration of the integrated framework that increased the efficiency of the manual pairing process. Several design constraints affected the development of this framework, all of which were intended to improve the efficiency with which manual pairing could take place. The resulting tool needed to address the following items: (1) be simple to use, with an intuitive user interface; (2) present the data on a single screen with a uniform format so that it is easy to compare records from dissimilar sources; (3) automate as much of the pairing process as possible, leaving the end user with little additional efforts after the tool is implemented for a given set of RCRS entries; and (4) store the paired results in a consistent and easily decipherable format. It was decided that the most effective approach for this version of the framework was a graphical user interface (GUI) developed using the PyQT4 Python bindings for the Qt cross-platform GUI/XML/SQL C++ framework. As before, the pairing tool actually used to match records was built in Python 2.7 using the library Pandas to load, store, and manipulate data. These technologies were chosen to keep the resulting tool cross platform.

A video is provided on the project website that demonstrates the functionality of the resulting GUI tool. Additionally, Figure 4 provides a series of screenshots during operation. The GUI features a toolbar at the top which can be used to specify time difference and distance settings for filtering records (Figure 4a). Generally, for optimization purposes these filtering criteria were adjusted on a county by county basis. However, a time difference of 2-4 hours and a distance difference of approximately 5 miles were often used as a starting point during the matching process. When multiple 911 entries were potential matches, the earliest record was typically selected to counteract the issue of multiple 911 entries corresponding to a single RCRS entry.

For example, a given highway incident may be reported by several motorists, which can result in multiple 911 entries. However, for the purposes of this study, determination of the notification latency was based on when a 911 entry was first entered for a given incident. Below the GUI toolbar are two scrolling lists of entries. The left column lists the source records and the right column lists potentially matching destination records. In the Fig. 3 example, source records are from the RCRS system and destination records are from the corresponding county 911 database. However, this GUI tool allows pairing to occur in the other direction as well. The currently selected source record is indicated by a thick outline (Figure 4a). The user can left click on a destination record to indicate that it is a match to the currently selected source record, turning both green (Figure 4b). Alternatively, the user can right click on any destination record to indicate that the currently selected source record has no matches in the county records, turning the source record red (Figure 4c). Based on the data present in Fig. 3, the first destination record is likely a match to the selected source record based on the mile marker posts specified in both location fields. This can be marked as a match (Figure 4d) and the user can proceed to the next source record in the left-hand column by left clicking it. This will bring up a new list of potential matches in the right-hand column (Figure 4e). Based on mile marker and time, the second record in the destination column is a match for the new source record and can be left clicked to indicate a match (Figure 4f). The user continues in this manner until all source records have been processed. The resulting pairings are saved automatically to a file specified by the user.

**(a)**

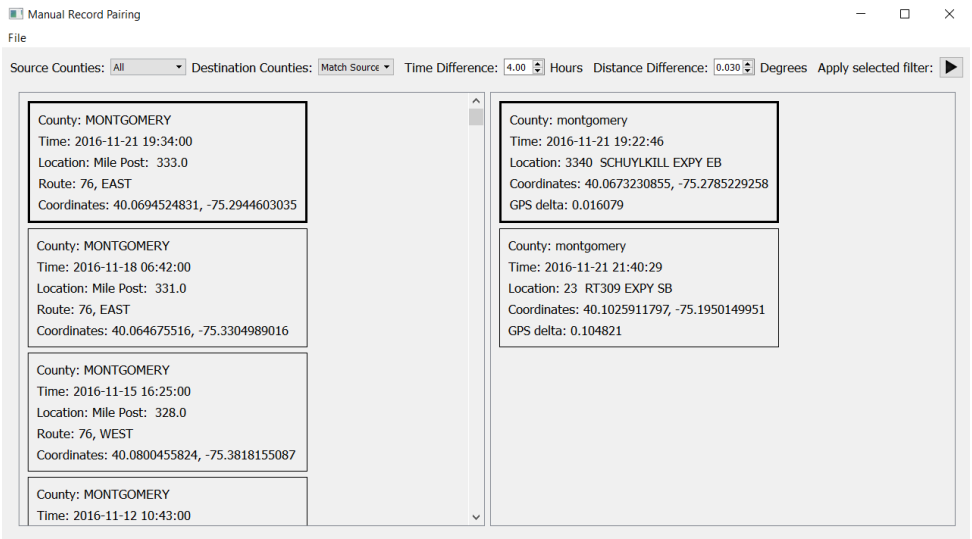
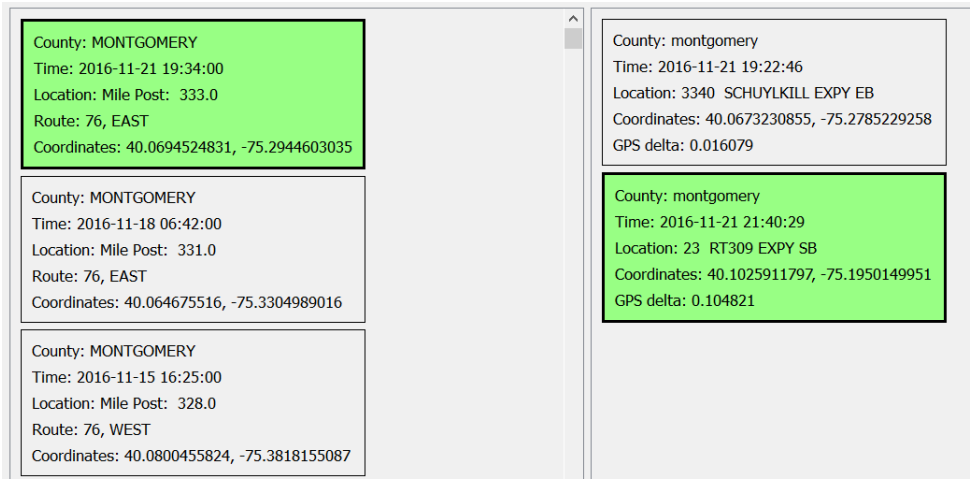
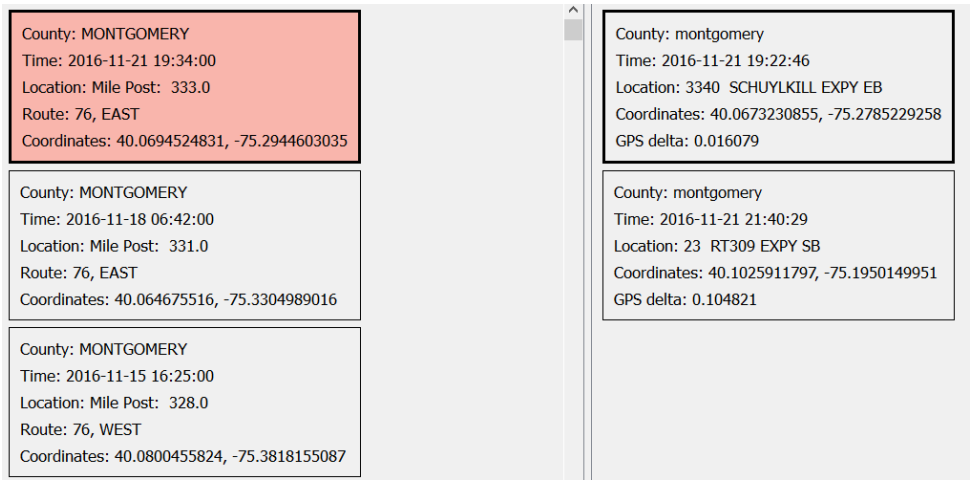



**Figure 4. Example of using GUI to locate matching county 911 and PennDOT RCRS records.**

**(b)**




**(c)**




**(d)**

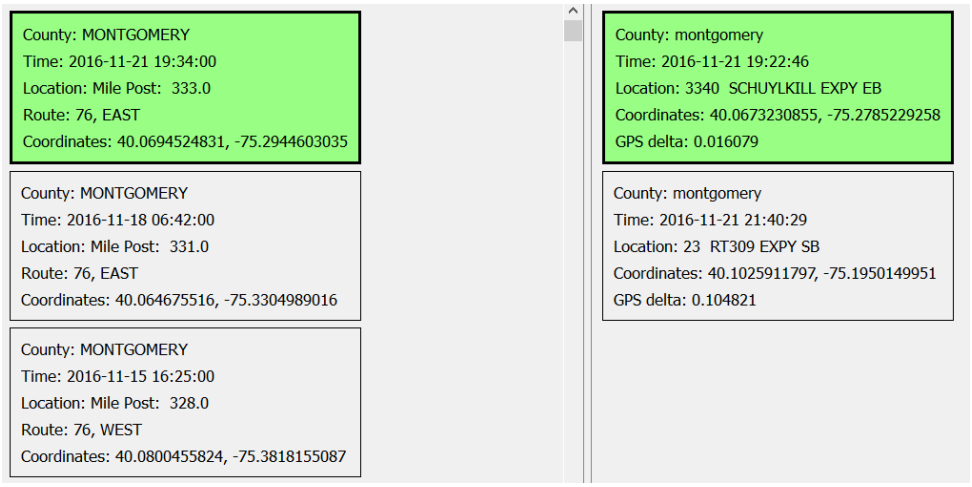
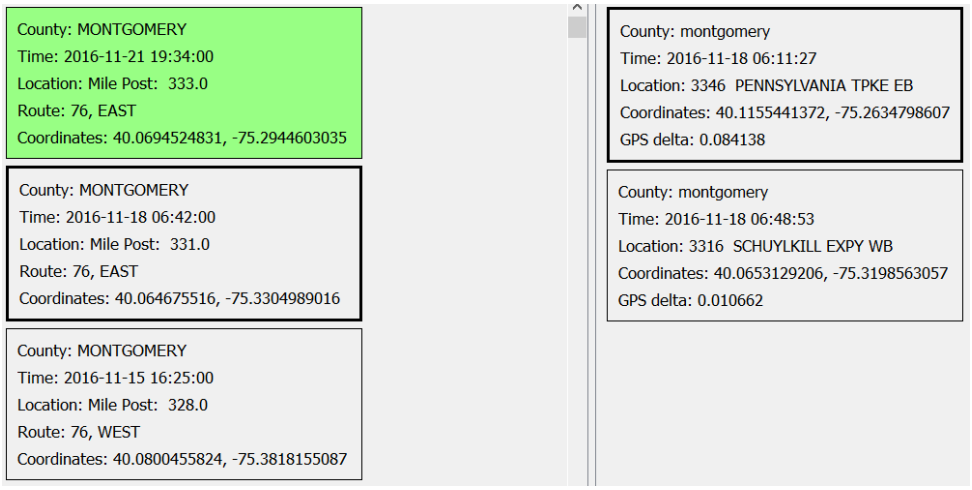



**Figure 4 (cont.). Example of using GUI to locate matching county 911 and PennDOT RCRS records.**

**(e)**




**(f)**

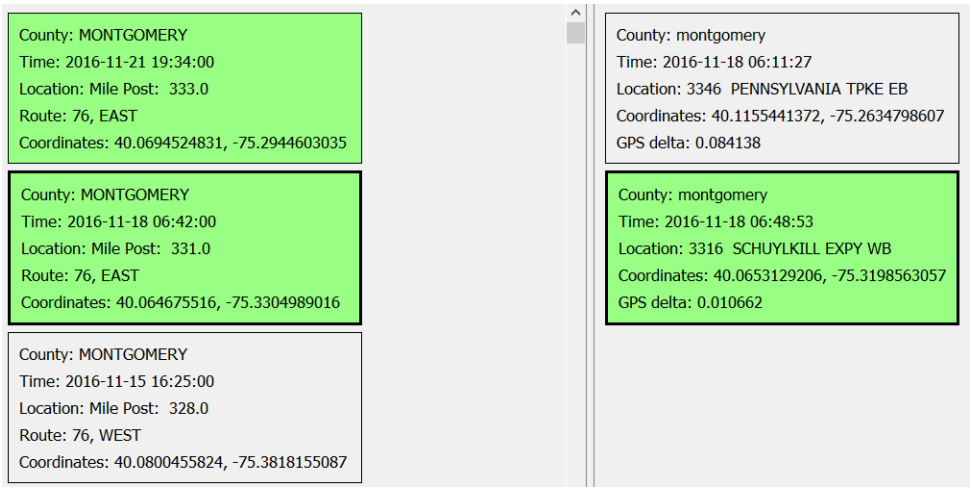



**Figure 4 (cont.). Example of using GUI to locate matching county 911 and PennDOT RCRS records.**

## 3 Data Analysis and Discussion

After development of the GUI-based tool, the Temple research team implemented it on all the datasets that had been previously paired (Table 1) to ensure compatibility of results and evaluate the improvements in the amount of time necessary to pair data. Though not as automated and rapid as the GPS-based framework, the GUI drastically improved the efficiency of the process over a fully manual approach. Based on previous discussion, the 17 counties for which 911 communication center data was obtained included 50.8% of the total number of filtered RCRS entries (i.e., 4,561 of 8,984). This percentage jumped to 88.6% (i.e., 4,561 out of 5,150) of the pertinent RCRS entries when excluding Philadelphia County, which singly contained nearly half of all the RCRS entries. Given the similarities between Philadelphia County and other nearby counties for which the Temple research team acquired data (i.e., Delaware and Montgomery

counties), the resulting statistical analysis is expected to be highly representative of conditions throughout the project study area. However, it should be noted that the Dauphin County data received did not allow manual pairing to RCRS entries (i.e., an overwhelming majority of location descriptors were labeled as "NULL"). Since Dauphin County contained 492 RCRS entries that were unmatchable based on the 911 county data provided, it was expected that a maximum of 45.3% of the total number of filtered RCRS entries would be matched after the integrated framework was implemented (79.0% if excluding Philadelphia County). The following sections discuss major aspects of the resulting statistical analyses performed after pairing was complete for each county.

## 3.1 Distribution of Data and Measure of Central Tendency

The results from statistical analysis are presented herein using a number of formats to facilitate discussion and interpretation. The main parameter of interest was a measurement of the time difference between when an incident was reported to 911 dispatch personnel and when PennDOT received notification of a highway incident. Figures 5 and 6 present stacked histograms and cumulative distribution plots for all matched county records. Versions of these figures are replicated in Appendix A for each county separately. All histograms and cumulative distribution plots have been normalized to the number of matched records for a given county (or all counties in the case where that result is plotted in the same figure). Figure 7 plots measurements of the central tendency of the time difference results based on county. Finally, Table 2 provides a tabular summary of the statistical results.

| County | Average (HH:MM:SS) | Std. Dev (HH:MM:SS) | Median (HH:MM:SS) | 25% Quartile (HH:MM:SS) | 75% Quartile (HH:MM:SS) | IQR (HH:MM:SS) | # of Records | # of Matches | # of No Matches | Match Rate (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| *All Counties* | *0:20:49* | *0:22:23* | *0:12:00* | *0:05:16* | *0:27:49* | *0:22:33* | *4069* | *2270* | *1800* | *55.8%* |
| Berks | 0:30:23 | 0:22:07 | 0:28:35 | 0:11:48 | 0:37:16 | 0:25:28 | 88 | 66 | 22 | 75.0% |
| Bucks | 0:17:19 | 0:19:09 | 0:10:00 | 0:05:00 | 0:21:30 | 0:16:30 | 492 | 395 | 97 | 80.3% |
| Centre | 0:41:42 | 0:24:08 | 0:35:03 | 0:27:40 | 0:49:26 | 0:21:45 | 86 | 39 | 47 | 45.3% |
| Cumberland | 0:16:17 | 0:19:55 | 0:09:38 | 0:04:46 | 0:18:31 | 0:13:45 | 311 | 227 | 84 | 73.0% |
| Delaware | 0:12:55 | 0:14:59 | 0:08:34 | 0:03:21 | 0:16:12 | 0:12:51 | 926 | 427 | 499 | 46.1% |
| Franklin | 0:19:31 | 0:16:09 | 0:12:25 | 0:07:28 | 0:34:05 | 0:26:37 | 83 | 14 | 69 | 16.9% |
| Lackawanna | 0:28:36 | 0:24:00 | 0:21:39 | 0:11:57 | 0:37:07 | 0:25:11 | 89 | 69 | 20 | 77.5% |
| Lebanon | 0:25:21 | 0:17:52 | 0:20:30 | 0:13:15 | 0:30:45 | 0:17:30 | 105 | 66 | 39 | 62.9% |
| Lehigh | 0:24:43 | 0:24:37 | 0:18:45 | 0:04:26 | 0:34:04 | 0:29:39 | 148 | 71 | 77 | 48.0% |
| Luzerne | 0:44:34 | 0:29:30 | 0:39:36 | 0:20:40 | 1:05:29 | 0:44:49 | 122 | 113 | 9 | 92.6% |
| Montgomery | 0:17:08 | 0:22:09 | 0:08:24 | 0:04:15 | 0:17:56 | 0:13:41 | 890 | 516 | 374 | 58.0% |
| Northampton | 0:25:09 | 0:24:55 | 0:16:49 | 0:04:22 | 0:38:16 | 0:33:54 | 60 | 38 | 22 | 63.3% |
| Schuylkill | 0:37:32 | 0:21:09 | 0:36:35 | 0:20:14 | 0:54:29 | 0:34:15 | 75 | 43 | 32 | 57.3% |
| Susquehanna | 0:37:11 | 0:24:28 | 0:34:27 | 0:18:20 | 0:49:55 | 0:31:35 | 52 | 36 | 16 | 69.2% |
| Venango | 0:37:38 | 0:23:49 | 0:33:00 | 0:19:07 | 0:48:08 | 0:29:01 | 30 | 21 | 9 | 70.0% |
| York | 0:19:44 | 0:23:23 | 0:12:25 | 0:08:08 | 0:21:06 | 0:12:58 | 512 | 129 | 383 | 25.2% |

**Table 2. Summary of statistical results based on pairing PennDOT RCRS and county 911 entries.**

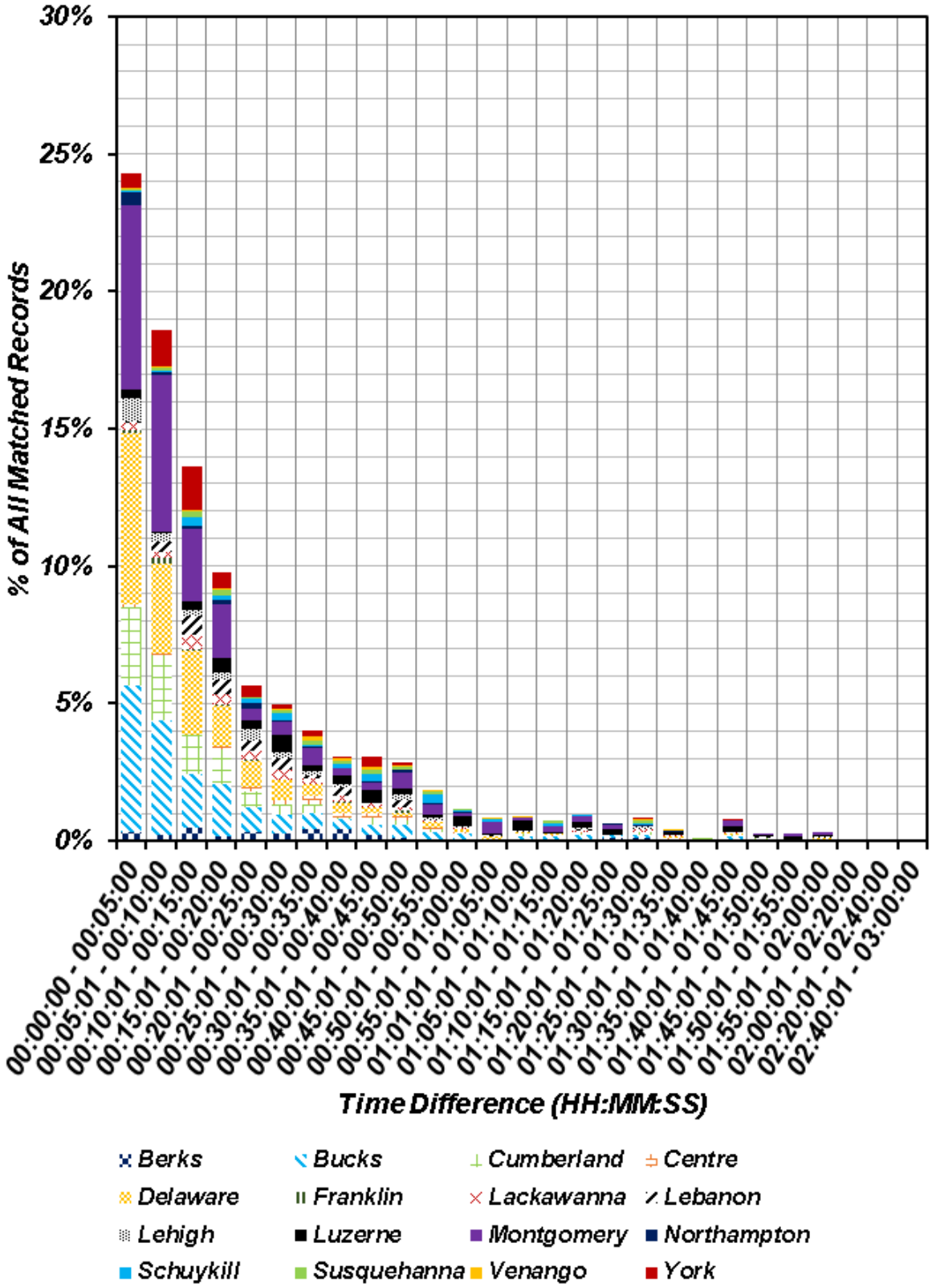



**Figure 5. Stacked histogram of time difference for all county matched records.**

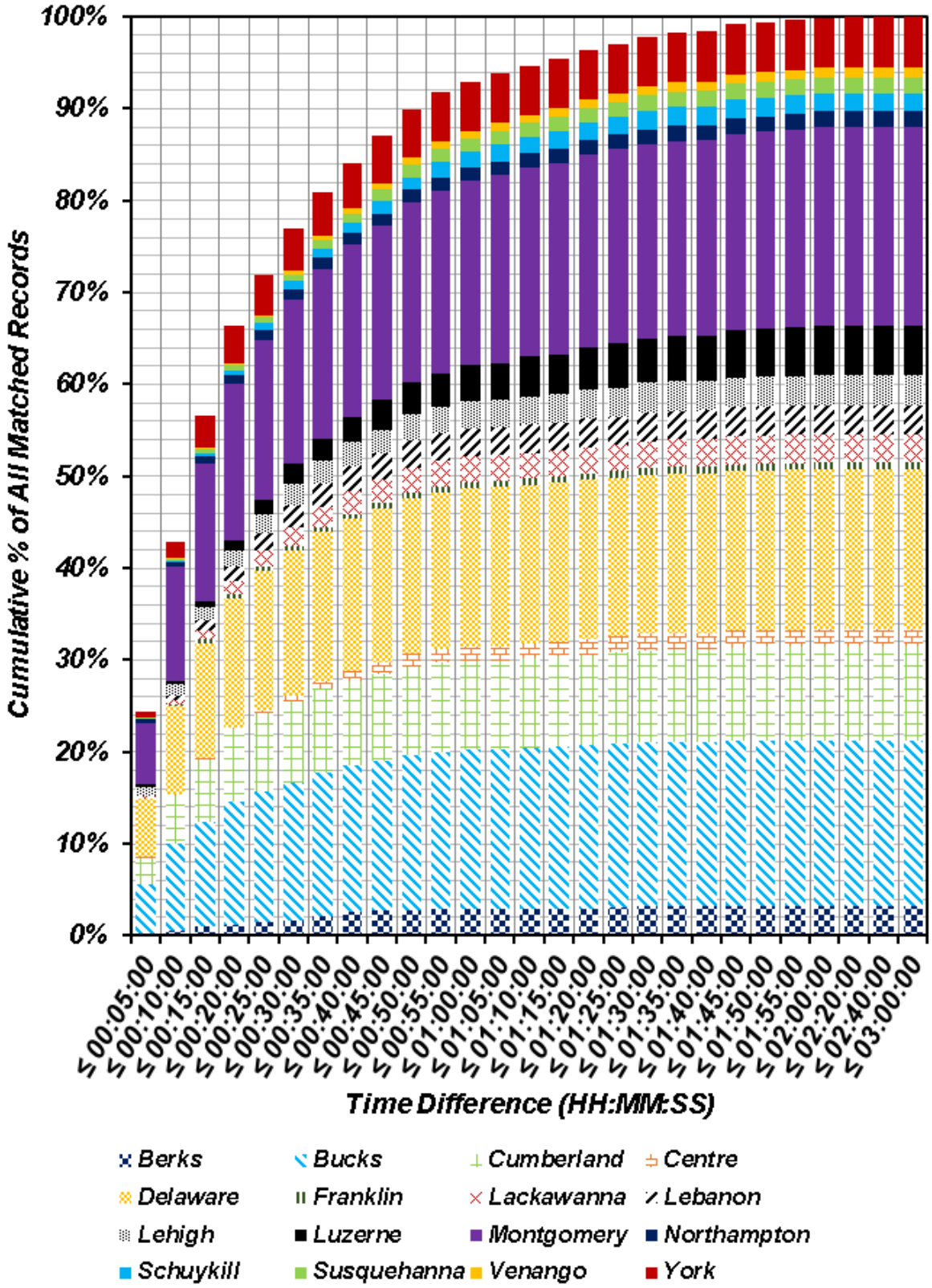



**Figure 6. Stacked cumulative distribution plot of time difference for all county matched records.**

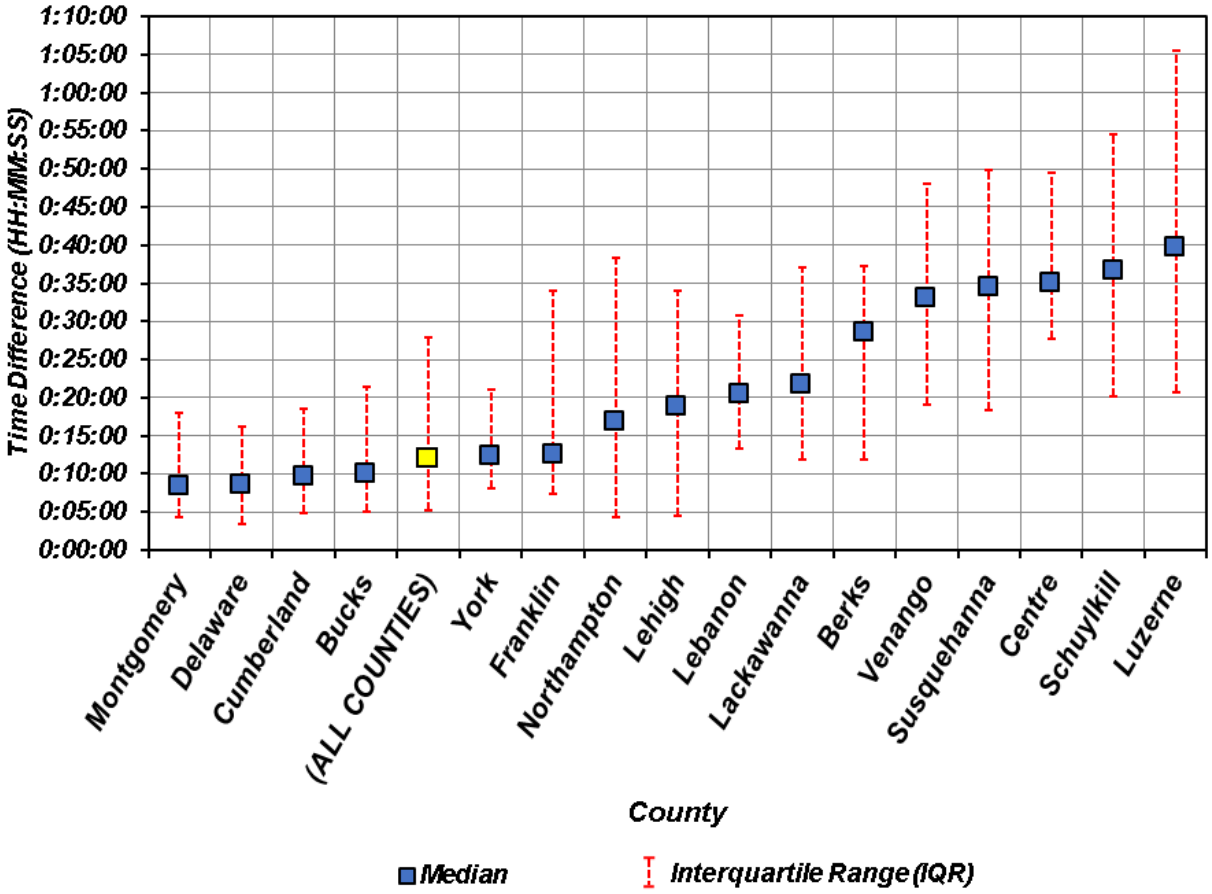



**Figure 7. Median time difference and IQR for matched records based on county.**

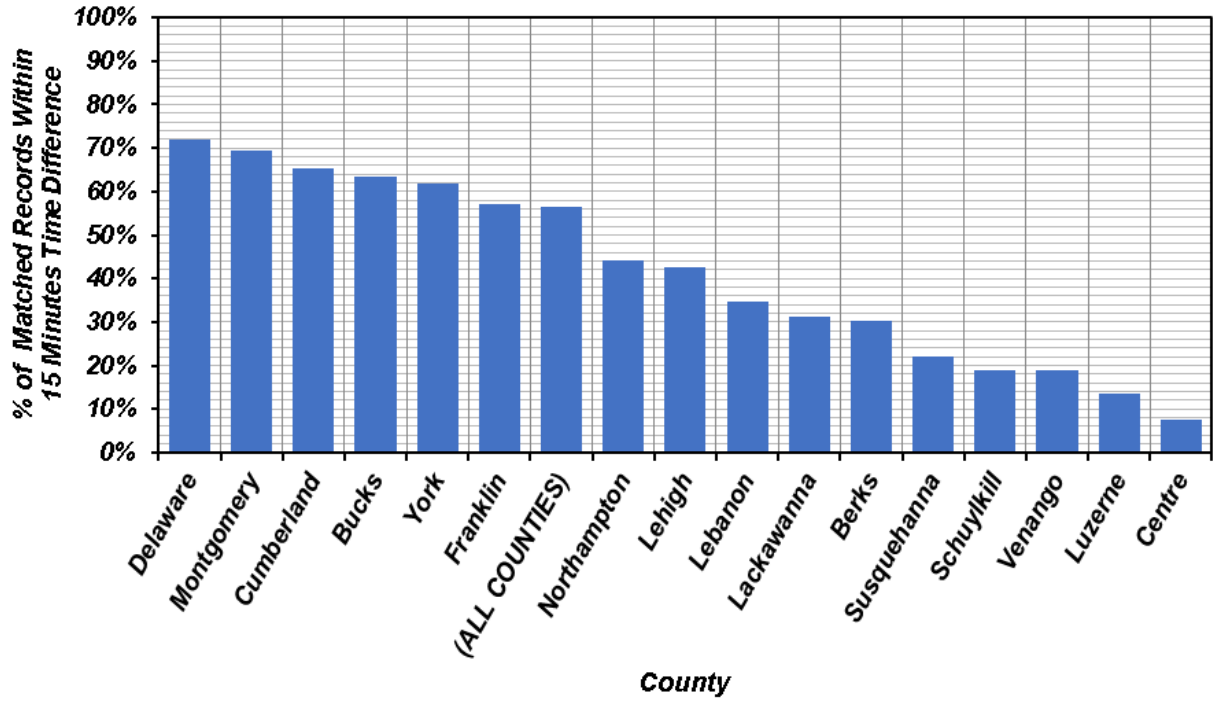



**Figure 8. Comparison of counties based on percentage of matched records within a 15 minute time difference.**

A number of observations are notable in the aforementioned figures and tables, particularly related to the central tendency of the time difference results. An examination of Figures 5 and 6 highlights that the resulting time difference is not normally distributed. The distribution is heavily skewed towards the shorter time differences and exhibits a significantly long tail. For example, nearly 70% of all matched records exhibit a time difference of less than 20 minutes. However, nearly 10% of all matched records push that latency to at least an hour or more. This represents a factor of three increase in the time difference. This overall pattern is similar to either an exponential distribution or a Pareto distribution, which is a skewed, heavy-tailed distribution that is sometimes used to model the distribution of incomes and other financial variables. The majority of the counties exhibit a similar distribution when viewed separately from each of the other counties (e.g., see figures in Appendix A). Given the non-normally distributed nature of the time difference results, it is important to use a suitable measurement to represent "average" time difference for all the counties so that overall performance can be evaluated. Generally, in descriptive statistics the "average" value of a parameter highlights the central tendency of its distribution. This central tendency is combined with some measurement of dispersion/variability of a parameter to provide a systematic manner with which to compare and contrast distributions. Typically, the arithmetic mean ($\mu$) is often used as a measure of the central tendency of a distribution and standard deviation ($\sigma$) is calculated to represent variability. However, these parameters are inherently tied to normal distributions of data and can be ineffective or

misleading in describing distributions with significant skew (as is the case in the time difference results from this study). The reason for that is that large outliers will tend to disproportionately affect $\mu$ and the resulting $\sigma$ may describe nonsensical or impossible outcomes. For example, an examination of the statistical descriptors in Table 2 demonstrates that the arithmetic mean of the time difference for all counties was approximately 21 minutes. The standard deviation was approximately 22.5 minutes. A confidence interval of $\pm 1\sigma$ leads to a negative time difference. Multiple counties suffer from this issue (e.g., Bucks, Cumberland, Delaware, etc.), particularly those that tend to exhibit shorter latency times.

Given the preceding discussion, a better measurement of the central tendency is the median value. The median of a distribution is determined by identifying the middle value when the data is ranked from largest to smallest. In that manner there are an equal number of data points larger and smaller than the median, and the median represents the middle quartile of the data (i.e., $50^{th}$ percentile). A typical confidence interval that is used to represent data variability with the median value is the interquartile range (IQR). IQR represents the difference between the upper quartile (i.e., $75^{th}$ percentile) and lower quartile ($25^{th}$ percentile) of the data. IQR is not affected by extreme values and is therefore often used together with the median when the distribution is skewed. Table 2 presents the median values and IQR for each of the counties in addition to $\mu$ and $\sigma$. The median values are generally quite different from $\mu$ as would be expected for distributions with significant skew. Figure 7 plots the median time difference and IQR as the confidence interval to allow a comparison of latency by county. Note that the use of IQR prevents the prediction of negative time difference in the confidence interval surrounding the median value. Based on the data in Table 2 and Figure 7, the overall median time difference between matched RCRS and 911 entries was 12 minutes. Based on the upper quartile estimate, 75% of all matched records have a time difference less than approximately 28 minutes. When viewed at an individual county level, Montgomery, Delaware, Cumberland, and Bucks counties all exhibit time differences shorter than the median computed for all counties. In fact, of those counties, all but Bucks County exhibit a time difference in single digit minutes (with Bucks recording a median of 10 minutes exactly). Generally, a trend could be noted where counties with smaller time differences tended to exhibit smaller variability in their distributions (i.e., smaller IQR). For example, Luzerne County, which exhibited the largest median time difference, also had the largest IQR. Overall, there was large scatter in the results when comparing extremes in county-level results. For example, Figure 8 highlights the county-level differences in how often PennDOT was informed of a highway incident within 15 minutes. A value of 15 minutes for the time difference is roughly between the arithmetic mean and the median for time difference across all counties, so it serves as a useful gauge of overall county statistical behavior. In this case, there is nearly a factor of ten difference between the county with the highest percentage of matched records with less than 15 minutes time difference and the county with lowest. The largest median time difference was also nearly five times slower than the smallest median time difference (Table 2 and Figure 7). Finally, the county-level difference in IQR (i.e., data variability) between

smallest and largest was nearly as high at approximately 3.5 times (Table 2 and Figure 7). These results suggest appreciable variability in the various factors affecting 911 reporting practices across the Commonwealth of Pennsylvania.

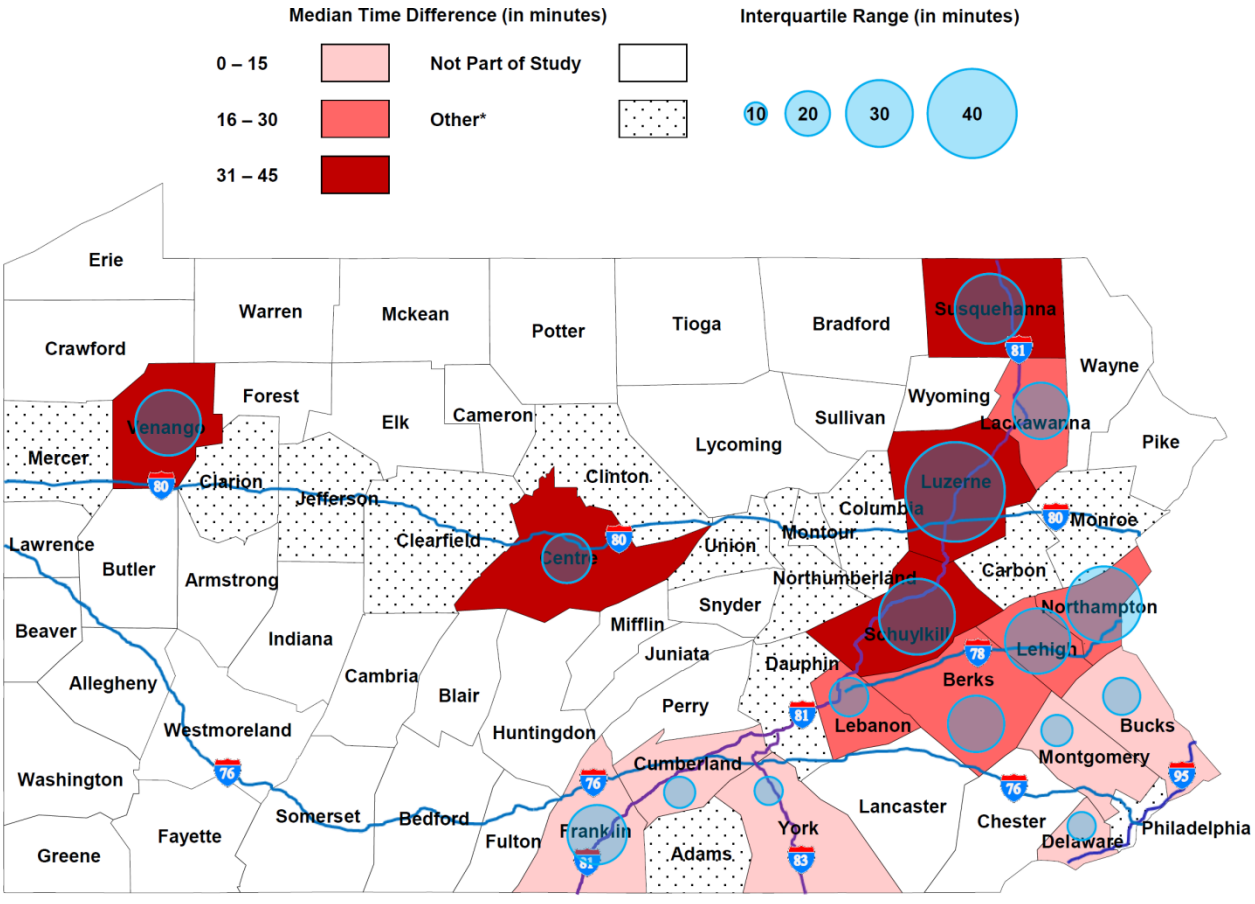
## 3.2 Geospatial Distribution of Results

In addition to examining the statistical descriptors of the time differences across counties, it is also useful to visually examine the geospatial distribution of these time differences across the Commonwealth of Pennsylvania. This may reveal patterns related to geography that may prove useful for PennDOT practice (e.g., one particular stretch of one of the highways in this study may be significantly different than nearby stretches). Figure 9 presents a choropleth map of counties across the Commonwealth of Pennsylvania with a single hue progression color scheme representing the median time difference computed in this study. Also included for each county is a circular representation of the IQR to highlight variability in the results. An examination of Figure 9 yields a number of items worthy of discussion. First, due to issues with data acquisition, much of the RCRS entries along the I-80 corridor across Pennsylvania were not matched to 911 county records. Areas of the Commonwealth traversed by I-95, I-83, the southernmost section of I-81 starting in Cumberland County, and (where data is available) I-76 generally exhibited smaller time differences between RCRS entries and 911 records. Conversely, the sections of I-81 heading north from Lebanon County (and the only counties for which 911 data was procured along I-80 in this study) exhibited larger time differences. There are a number of potential factors that may explain these observations. For example, differences in 911 reporting procedures, allocation of responsibility for incidents along highways, population density, density of the 511PA traffic camera network, and 911 call volumes may all play a role in the geospatial distribution highlighted in Figure 9.

## 3.3 Match Rate

One item of interest from Table 2 is the rate with which RCRS entries were matched to county 911 records. As noted in Table 2, the average match rate for all counties was nearly 60%. Some counties exhibited nearly complete matching rates (e.g., Luzerne County with nearly 93% of entries matched) while others had less favorable matching rates (e.g., Franklin County with only approximately 17% of records matched). There are a number of explanations for these discrepancies. In an ideal scenario, all RCRS entries correspond to at least one entry in databases maintained by 911 call centers. However, there may be some situations where highway incidents are not reported to 911 call centers or are otherwise unavailable within their CAD systems. For example, PennDOT personnel monitoring traffic cameras may respond to an incident and generate a lane closure without any 911 calls being generated. In some counties, highway

incidents may be handled by another nearby agency/call center (e.g., Northumberland County referred the Temple research team to Union County for records related to I-80 in Northumberland County) or by PSP. In many cases, the CAD systems used by the 911 call centers cannot maintain records longer than a specified time period (e.g., one year). In the case of this study where RCRS entries extended back three years, the 911 data provided to the Temple research team was incomplete for certain counties, which may explain the non-matching RCRS entries. Therefore the match rate column in Table 2 highlights the difficulties caused by differences in 911 call center operations, reporting procedures, and CAD systems across the Commonwealth. One lesson from this study is that increased integration of datasets among the various stakeholders involved with highways can begin to address some of the issues noted in this study and improve operational emergency management of highways in Pennsylvania.




* Other may include: data analysis in progress, rejection of RTK request, pending RTK request, poor data quality, or unresponsive county.

**Figure 9. Geospatial distribution of median time difference and IQR.**

# 4 Data Curation & Archiving Efforts

Given the large amount of data collected and the importance of this study to highway safety across the Commonwealth, it was important that the project data, analysis, and deliverables be archived for future access by PennDOT personnel. In addition to the raw data from the 911 call centers and PennDOT RCRS, the Temple research team generated a number of other outputs as the project progressed and estimates were developed for the latency between 911 call center entries and PennDOT RCRS entries: (1) copies of normalized data from counties investigated in study; (2) Python-based scripts that manipulate data and perform analysis to locate pairings between 911 and RCRS entries; (3) results from pairing efforts between the 911 and RCRS databases; and (4) task deliverable reports related to research efforts throughout project. The Temple research team developed a project website with a user friendly interface to help archive all of the aforementioned information related to research efforts. This website has and will continue to serve as the instrument by which data is curated for the project. The following sections describe the structure and content of the website.

## 4.1 Summary of Project Website

The TEM WO 009 project website is available via the following hyperlink: https://www.isip.piconepress.com/projects/penndot_response_time/. Most of the website is unprotected, but a username and password are required to download restricted information (e.g., data files, Python tools, etc.). The project website is organized using a simple structure. The home page provides a quick summary of the project objective and a list of recent project highlights. Included on the project home page are hyperlinks to the following items: (1) Overview; (2) Downloads; and (3) Docs. Item (1) is a relatively minor aspect of the website in relation to the project efforts. The "Overview" link sends the website viewer to a single webpage that provides a more thorough discussion of the project goals and research efforts. It is meant to augment the home page and present the project in a more detailed context so that users unfamiliar with the project (e.g., other PennDOT personnel with which the PennDOT project team wishes to share the website) can better understand the project prior to exploring other aspects of the site. The "Docs" link sends the website viewer to an index site with various documentation regarding the project. For example, from the "Docs" link, the user can download all relevant project reports, a demo of the pairing tool developed in this project, and the user manual for the PennDOT RCRS system. For ease of use, there is also a direct link to the project draft final report and pairing demo video on the project home page.

The most significant aspect of the project website is the location of all relevant project files, which is accessible from the "Downloads" hyperlink on the project website home page. Following this hyperlink leads the user to a landing page with additional hyperlinks to relevant project items. The most important of these is multiple iterations of the compressed archive of all

data files in a single downloadable GNU compressed tar file. Additionally, a directory of all data files by county is also available via hyperlink on this page. The directory of all data files includes the following list of folders (Figure 10):

- 000_penndot: This folder contains Excel® spreadsheets of all RCRS data provided by PennDOT for the purposes of this research project. There are multiple versions of the spreadsheet. The first "penndot_tidy.xlsx" presents all the filtered RCRS data for this project as originally provided in the master RCRS spreadsheet provided via email by the PennDOT project team on Tuesday, November 22, 2016. The other versions present modifications to the RCRS data as performed during data normalization. For example, "000_penndot.xlsx" presents the normalized copy of the filtered RCRS data based on the normalization techniques described earlier in this report. This data is also provided in a comma-separated values (CSV) file ("000_penndot.csv" and "000_penndot_norm.csv"). Appended to each entry is a unique hash identifier for simplified, rapid data lookup functionalities using the Python scripts developed for this project. The CSV files can be opened using either a simple text editor (e.g., the Notepad program typically provided within the Windows operating system) or Microsoft Excel® among other software. Also included in this folder is an archival subfolder ("_ARCHIVES") that contain original archives of the RCRS data prior to any normalization and filtering. The subfolder "_DOCS" in this particular case is empty but included to maintain consistency with the structure of the remaining data folders for the counties (see below).

- County folders: A folder is provided for each county for which data was acquired and processed in this study. The naming convention for these folders is "XXX_county name" where the XXX represents the county number. Inside each county-level directory folder are the following files (e.g., Figure 11):

    - CSV file with normalized 911 call center data. As with the normalized PennDOT RCRS data, a unique hash identifier has been affixed to each 911 data entry in the normalized CSV file. The naming convention is "XXX_county name.csv" where XXX is the county number.

    - DAT files containing pairing information between 911 call entries and RCRS data. Two DAT files are provided: (1) "XXX_county name_ppc.dat" represents the pairing information generated when matching PennDOT RCRS entries to 911 call center entries, starting with PennDOT RCRS as the source data and 911 data as destination data; and (2) "XXX_county name_pcp.dat" represents the pairing information generated when matching 911 call center entries to PennDOT RCRS, starting with the 911 records as the source data and RCRS entries as the destination data. In both cases, the pairing information is expressed as a two column table with source hash identifier paired with destination hash identifier. So in cases where the pairing goes

from PennDOT RCRS to 911 data, the first column of the table in the DAT file would contain the hash identified associated with a particular RCRS entry and the second column would contain the corresponding 911 call entry hash identifier. In cases where a credible match could not be located, "N/A" was inserted in the second column. This format was repeated in the DAT file for pairing data in the opposite direction. The DAT folder can be opened using a simple text editor as mentioned with CSV files.

- Archival folder ("_ARCHIVES") with original 911 data. Typically this folder contains Excel® files as provided by the county 911 call center (e.g., Figure 12). However, in some circumstances, these original data files were provided in PDF files or other formats. Additionally, the archival folder typically contains any revised versions of the data files that aided in the normalization process. For example, in some cases, the original Excel® spreadsheets contained merged cells or other distracting properties that hindered parsing of the data. The Temple research team manually revised such files to remove any of these issues prior to the pairing process. The revised spreadsheet is then also provided in the archival folder in addition to the original data files provided by the county 911 call center. Another common issue was when a 911 call center provided a PDF file, which necessitated conversion into a spreadsheet. Again, both files were included for the applicable counties in its archival folder.

- Documents folder ("_DOCS") that contains pertinent documentation regarding the county data (Figure 13). For example, in cases where an RTK request was formally made, the completed form is provided in this subfolder. Also included is also any other pertinent documentation that may be useful for understanding the acquired data (e.g., document with list of abbreviations, user manuals, etc.).

- Explanation text file: A simple text file ("_AAREADME.txt") is provided in this folder (and subsequent sub-folders) that explains a number of items related to the data and/or structure of the current folder. For example, the "_AAREADME.txt" in the main directory briefly discusses many of the items included in this deliverable report (e.g., file naming conventions, folder/sub-folder structures, etc.). It also provides a change log to keep track of changes over previous iterations of the directory.

- Documentation folder: This documentation folder ("_DOCS") contains the county codes used in this study as an Excel® spreadsheet. This file is also accessible when clicking on the "County Codes" hyperlink on the project website home page as previously described.

- Tools folder: This folder ("_TOOL) contains the various Python codes used to parse and match the data in this project. The raw coding and documentation related to usage of the codes are all provided.

# Index of /projects/penndot_response_time/downloads/v0.1

| Name | Last modified | Size | Description |
|---|---|---|---|
| Parent Directory | | - | |
| 000 penndot/ | 2017-07-03 12:53 | - | |
| 002 allegheny/ | 2017-07-03 12:53 | - | |
| 004 beaver/ | 2017-07-03 12:53 | - | |
| 006 berks/ | 2017-07-03 12:53 | - | |
| 009 bucks/ | 2017-07-03 12:53 | - | |
| 014 centre/ | 2017-07-07 12:51 | - | |
| 015 chester/ | 2017-07-05 20:31 | - | |
| 021 cumberland/ | 2017-07-03 12:53 | - | |
| 022 dauphin/ | 2017-07-03 15:40 | - | |
| 023 delaware/ | 2017-07-05 20:54 | - | |
| 028 franklin/ | 2017-07-03 12:53 | - | |
| 035 lackawanna/ | 2017-07-03 12:53 | - | |
| 036 lancaster/ | 2017-07-03 12:53 | - | |
| 038 lebanon/ | 2017-07-03 12:53 | - | |
| 039 lehigh/ | 2017-07-03 12:53 | - | |
| 040 luzerne/ | 2017-07-03 12:53 | - | |
| 046 montgomery/ | 2017-07-03 12:53 | - | |
| 048 northampton/ | 2017-07-03 12:53 | - | |
| 054 schuylkill/ | 2017-07-03 12:53 | - | |
| 058 susquehanna/ | 2017-07-05 20:43 | - | |
| 061 venango/ | 2017-07-05 20:43 | - | |
| 067 york/ | 2017-07-05 20:42 | - | |
| _AAREADME.txt | 2017-07-14 17:00 | 12K | |
| _DOCS/ | 2017-06-23 10:25 | - | |
| _TOOLS/ | 2017-06-23 10:36 | - | |

*Apache Server at www.isip.piconepress.com Port 443*

**Figure 10. Main downloads directory on project website.**

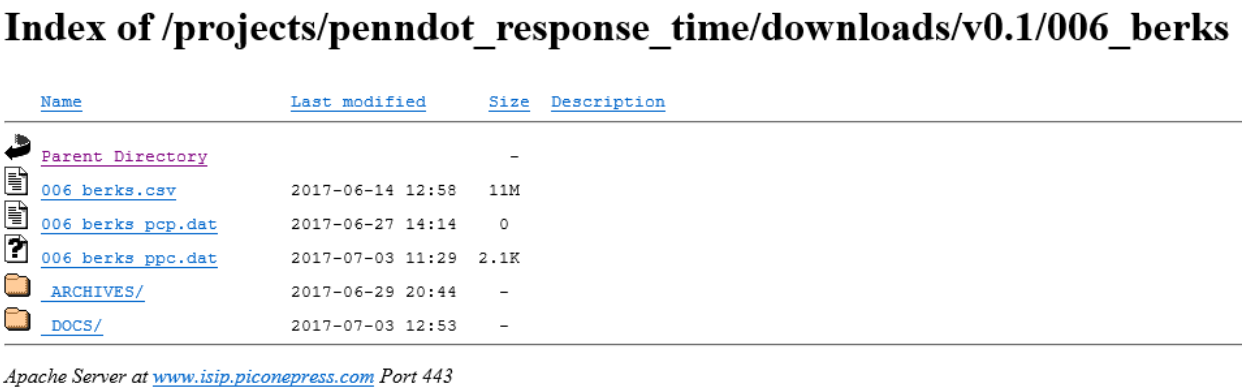
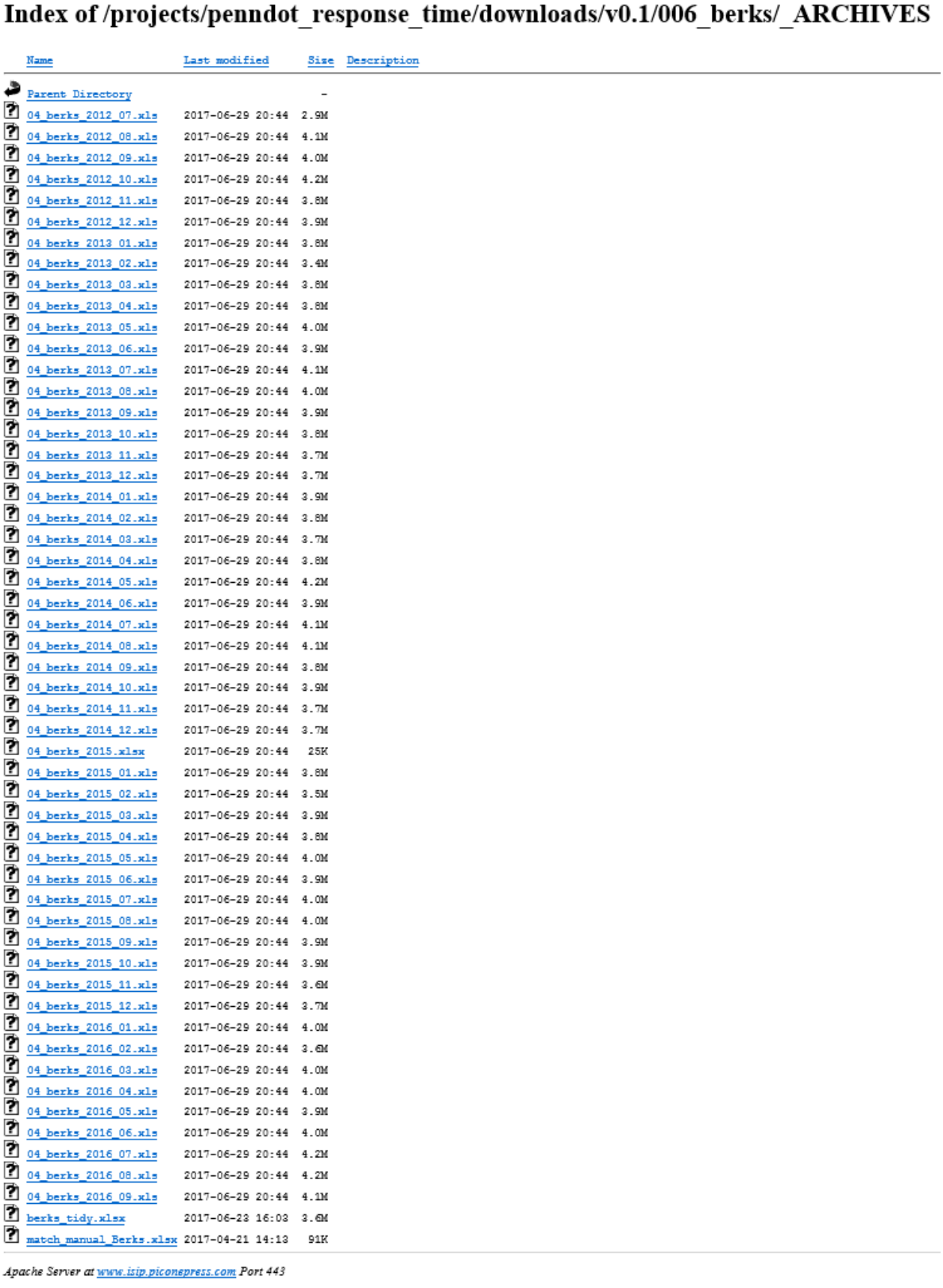



**Figure 11. Example county-level directory (Berks County).**



Index of /projects/penndot_response_time/downloads/v0.1/006_berks/_ARCHIVES

| Name | Last modified | Size | Description |
|---|---|---|---|
| Parent Directory | | - | |
| 04_berks_2012_07.xls | 2017-06-29 20:44 | 2.9M | |
| 04_berks_2012_08.xls | 2017-06-29 20:44 | 4.1M | |
| 04_berks_2012_09.xls | 2017-06-29 20:44 | 4.0M | |
| 04_berks_2012_10.xls | 2017-06-29 20:44 | 4.2M | |
| 04_berks_2012_11.xls | 2017-06-29 20:44 | 3.8M | |
| 04_berks_2012_12.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2013_01.xls | 2017-06-29 20:44 | 3.8M | |
| 04_berks_2013_02.xls | 2017-06-29 20:44 | 3.4M | |
| 04_berks_2013_03.xls | 2017-06-29 20:44 | 3.8M | |
| 04_berks_2013_04.xls | 2017-06-29 20:44 | 3.8M | |
| 04_berks_2013_05.xls | 2017-06-29 20:44 | 4.0M | |
| 04_berks_2013_06.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2013_07.xls | 2017-06-29 20:44 | 4.1M | |
| 04_berks_2013_08.xls | 2017-06-29 20:44 | 4.0M | |
| 04_berks_2013_09.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2013_10.xls | 2017-06-29 20:44 | 3.8M | |
| 04_berks_2013_11.xls | 2017-06-29 20:44 | 3.7M | |
| 04_berks_2013_12.xls | 2017-06-29 20:44 | 3.7M | |
| 04_berks_2014_01.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2014_02.xls | 2017-06-29 20:44 | 3.8M | |
| 04_berks_2014_03.xls | 2017-06-29 20:44 | 3.7M | |
| 04_berks_2014_04.xls | 2017-06-29 20:44 | 3.8M | |
| 04_berks_2014_05.xls | 2017-06-29 20:44 | 4.2M | |
| 04_berks_2014_06.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2014_07.xls | 2017-06-29 20:44 | 4.1M | |
| 04_berks_2014_08.xls | 2017-06-29 20:44 | 4.1M | |
| 04_berks_2014_09.xls | 2017-06-29 20:44 | 3.8M | |
| 04_berks_2014_10.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2014_11.xls | 2017-06-29 20:44 | 3.7M | |
| 04_berks_2014_12.xls | 2017-06-29 20:44 | 3.7M | |
| 04_berks_2015.xlsx | 2017-06-29 20:44 | 25K | |
| 04_berks_2015_01.xls | 2017-06-29 20:44 | 3.8M | |
| 04_berks_2015_02.xls | 2017-06-29 20:44 | 3.5M | |
| 04_berks_2015_03.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2015_04.xls | 2017-06-29 20:44 | 3.8M | |
| 04_berks_2015_05.xls | 2017-06-29 20:44 | 4.0M | |
| 04_berks_2015_06.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2015_07.xls | 2017-06-29 20:44 | 4.0M | |
| 04_berks_2015_08.xls | 2017-06-29 20:44 | 4.0M | |
| 04_berks_2015_09.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2015_10.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2015_11.xls | 2017-06-29 20:44 | 3.6M | |
| 04_berks_2015_12.xls | 2017-06-29 20:44 | 3.7M | |
| 04_berks_2016_01.xls | 2017-06-29 20:44 | 4.0M | |
| 04_berks_2016_02.xls | 2017-06-29 20:44 | 3.6M | |
| 04_berks_2016_03.xls | 2017-06-29 20:44 | 4.0M | |
| 04_berks_2016_04.xls | 2017-06-29 20:44 | 4.0M | |
| 04_berks_2016_05.xls | 2017-06-29 20:44 | 3.9M | |
| 04_berks_2016_06.xls | 2017-06-29 20:44 | 4.0M | |
| 04_berks_2016_07.xls | 2017-06-29 20:44 | 4.2M | |
| 04_berks_2016_08.xls | 2017-06-29 20:44 | 4.2M | |
| 04_berks_2016_09.xls | 2017-06-29 20:44 | 4.1M | |
| berks_tidy.xlsx | 2017-06-23 16:03 | 3.6M | |
| match_manual_Berks.xlsx | 2017-04-21 14:13 | 91K | |

Apache Server at www.isip.piconepress.com Port 443

**Figure 12. Example county-level archival folder (Berks County).**


**Figure 13. Example county-level documents folder (Berks County).**

# 5 Conclusions

The effectiveness of the PennDOT RCRS system is dependent on the quality of information and the timeframe in which it is provided. Since the RCRS system is a major tool with which PennDOT monitors highways, the statistical results from this study can aid PennDOT in developing best practices for policy and procedural decisions related to traffic incident management, which can improve operation at the statewide, regional, and district traffic management centers. The follow section describes a proposed implementation plan to guide PennDOT efforts in effectively using the results from this study.

## 5.1 Implementation of Research Results

The primary results of this study include statistical evaluation of time difference between incidents recorded in the PennDOT RCRS and those recorded by emergency dispatch personnel at county 911 call centers. These research results present a wealth of information regarding traffic incident management and can be implemented in a number of ways to improve traffic operations across the Commonwealth of Pennsylvania:

- Estimates of the time necessary for PennDOT to receive notification of highway incidents across the Commonwealth of Pennsylvania can allow PennDOT to better allocate resources and is the first step in minimizing the time gaps for highway closures in response to emergencies. The spatial distribution of time difference is particularly useful as PennDOT can use the research findings to identify which stretches of highways should be targeted for improvements in notification timeline. PennDOT can then collaborate with other agencies to evaluate potential mechanisms by which to decrease the time difference.

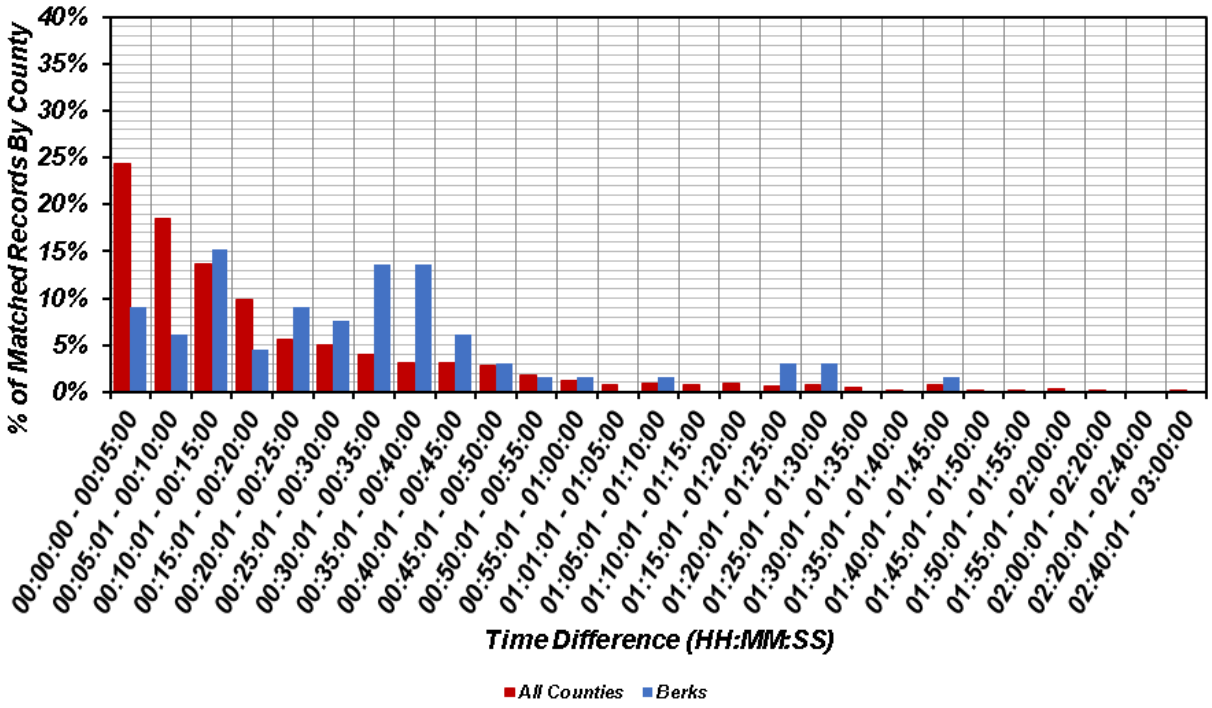
- Statistical information on notification timeline allows PennDOT to identify potential key elements and any critical missing information related to traffic incident management across the Commonwealth of Pennsylvania. The findings in this research study can be used by PennDOT to establish baseline statistical responses for counties where significant data was provided. PennDOT can then continue to use the integrated framework developed in this study to augment the existing dataset with critical missing data and examine other highways and/or areas within the Commonwealth of Pennsylvania.

- One major lesson from this study is that there is significant difference between various 911 communication center operations, reporting procedures, and CAD systems across the Commonwealth. In many ways, this decentralization increases the difficulty of establishing links between 911 call data and existing PennDOT RCRS records. Increased integration of datasets among the various stakeholders involved with highway incidents can begin to address some of these issues and improve operational emergency management of highways

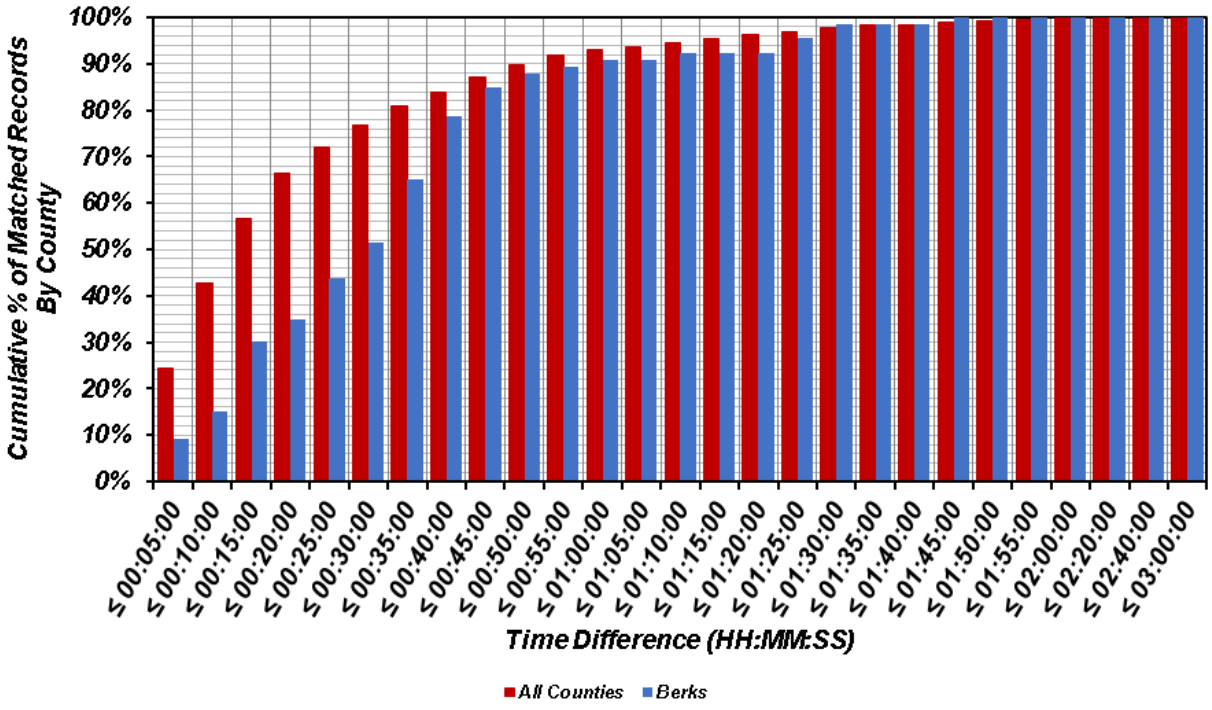
in the Commonwealth of Pennsylvania. PennDOT can use the results from this research study to explore how increased collaboration and integrated databases can improve notification latency.

- The statistical data examined in this study presents a snapshot of traffic incidents and the corresponding response timeline within the Commonwealth of Pennsylvania. PennDOT can implement the results of study by periodically deploying the integrated framework documented in this report to catalog any statistical changes in traffic incident timeline. This is particularly warranted after PennDOT implements any efforts to affect the timeline and improve notification time. This will allow PennDOT to evaluate the efficacy of traffic operations improvement efforts as well as to model any changes in county activities.

# Appendix A: Histograms and Cumulative Distribution Plots By County

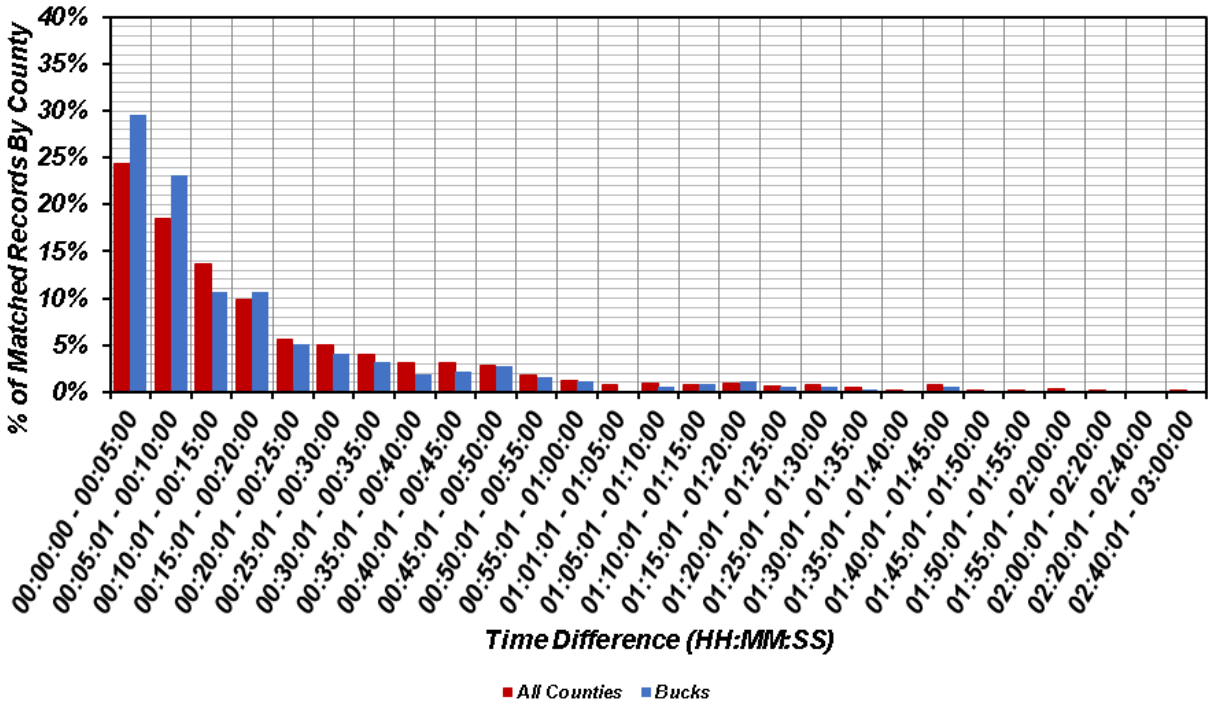



**(a)**

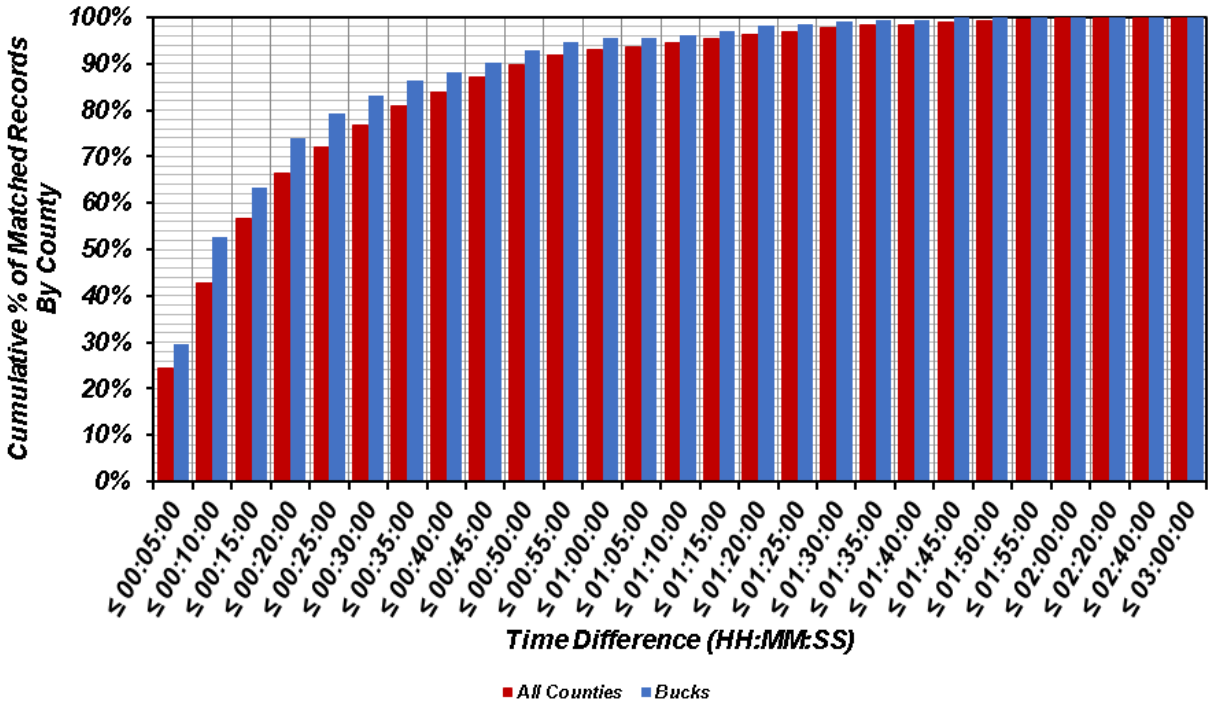



**(b)**

**Figure A.1. Comparison of time difference for matched records in Berks County versus all counties: (a) histogram; and (b) cumulative distribution.**

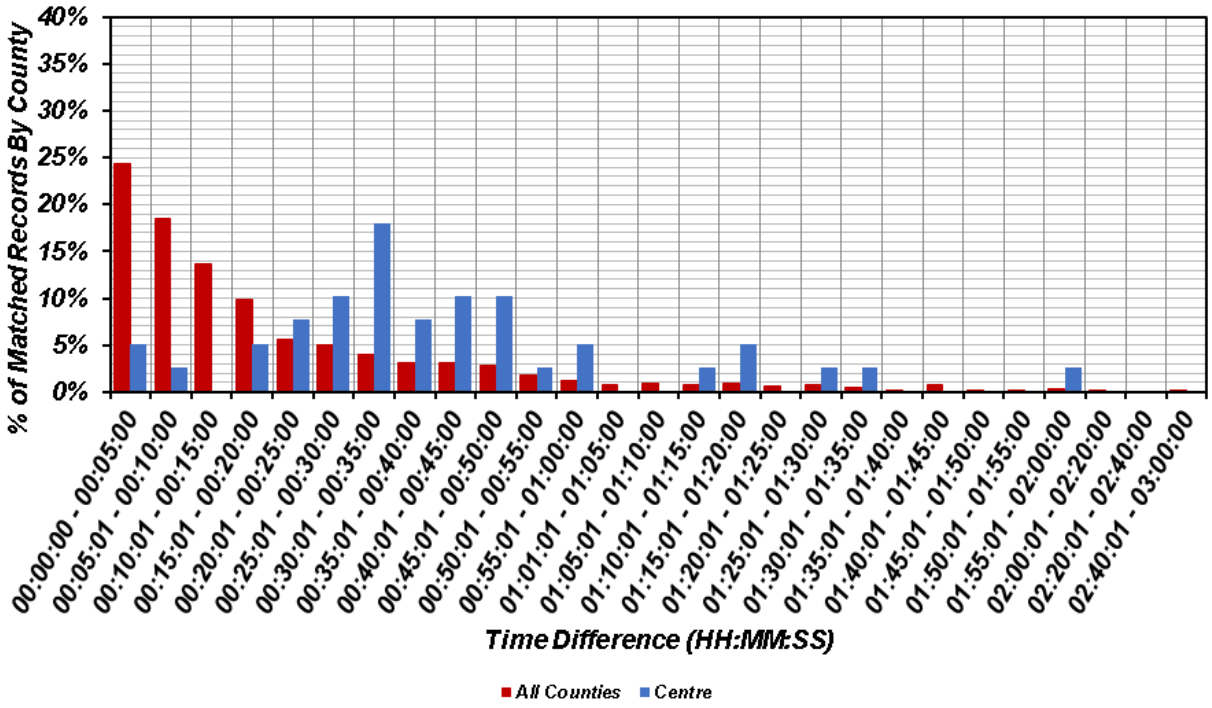



**(a)**

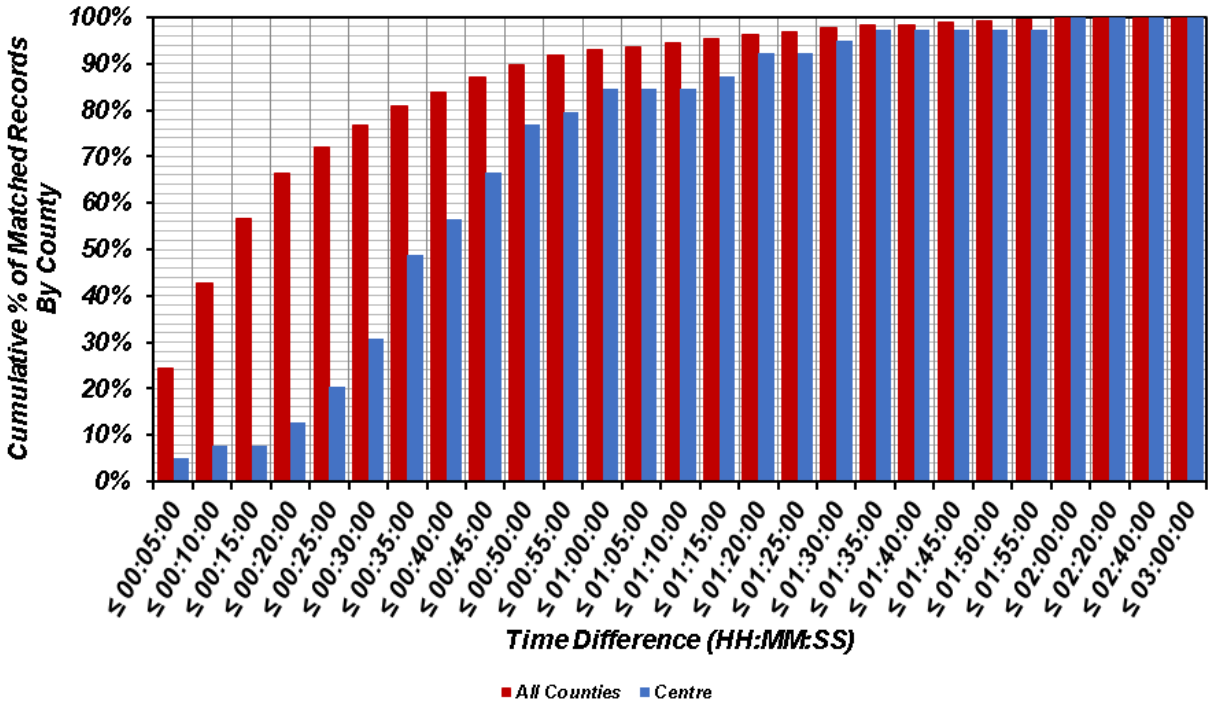



**(b)**

**Figure A.2. Comparison of time difference for matched records in Bucks County versus all counties: (a) histogram; and (b) cumulative distribution.**

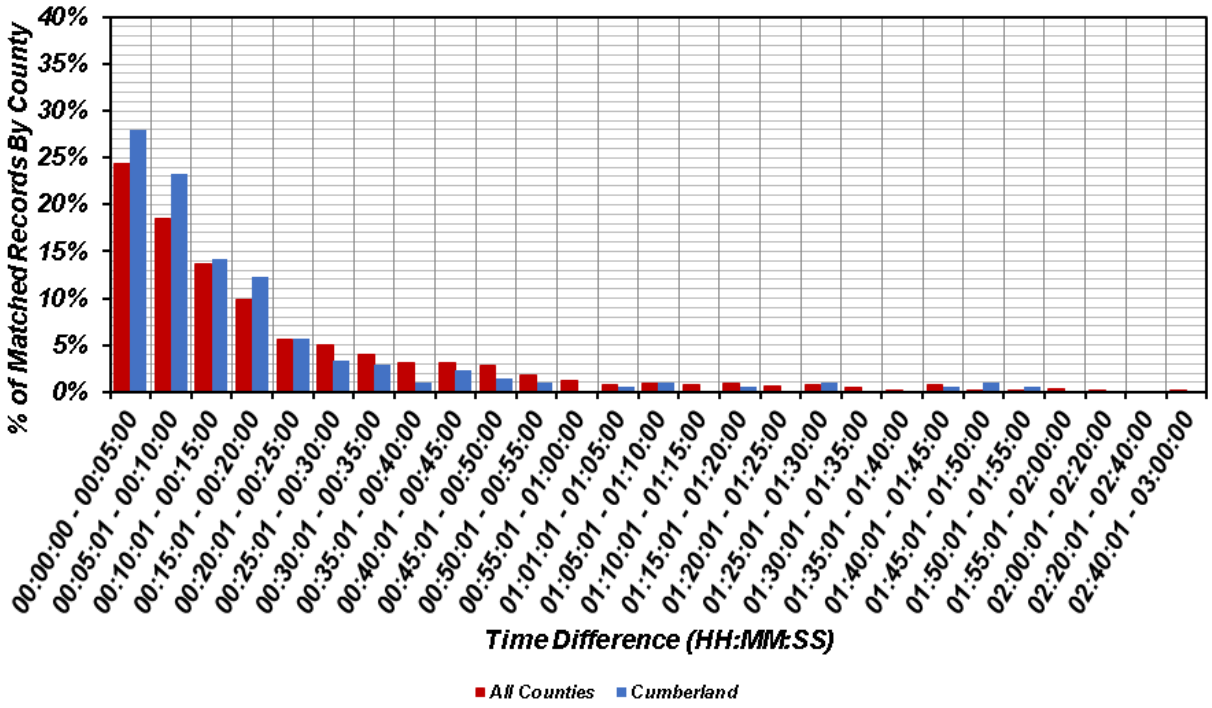



**(a)**

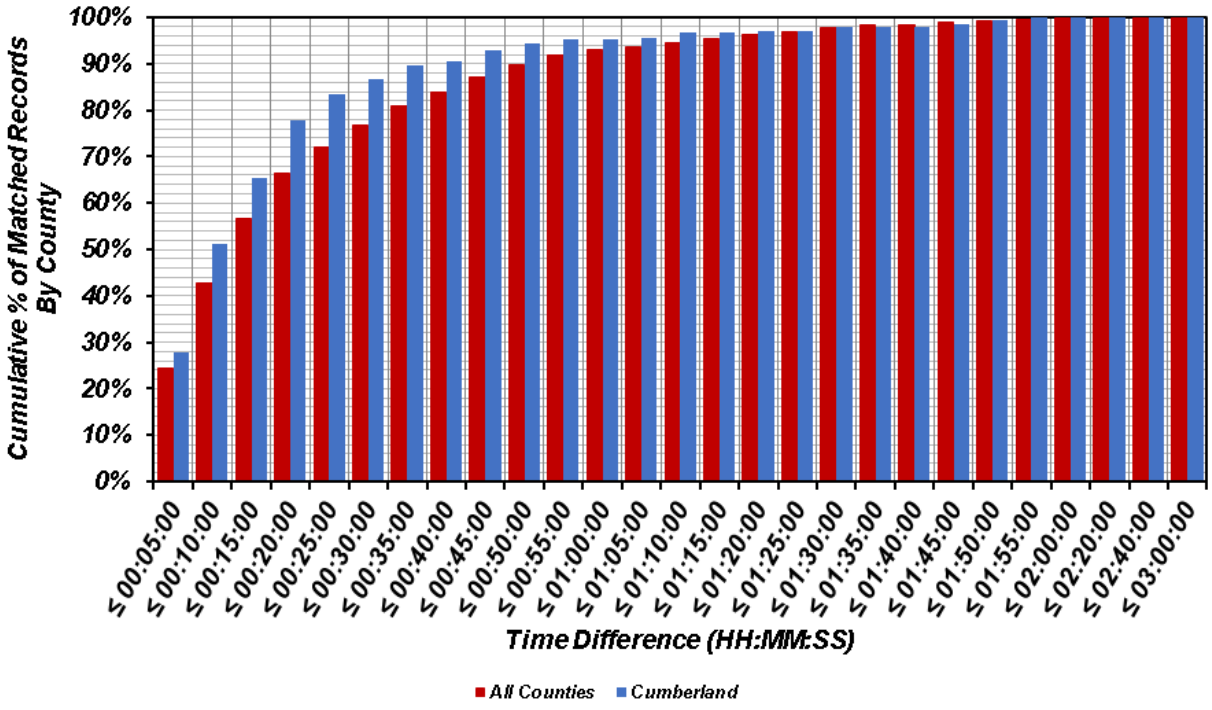



**(b)**

**Figure A.3. Comparison of time difference for matched records in Centre County versus all counties: (a) histogram; and (b) cumulative distribution.**

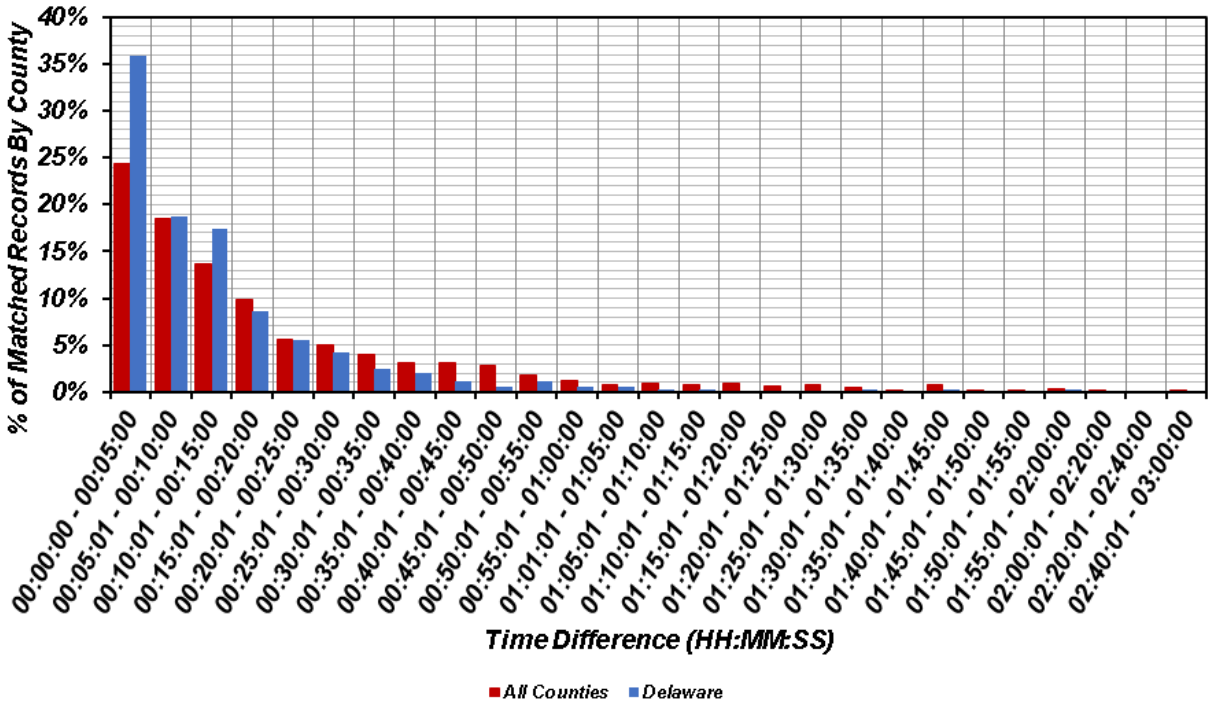



**(a)**

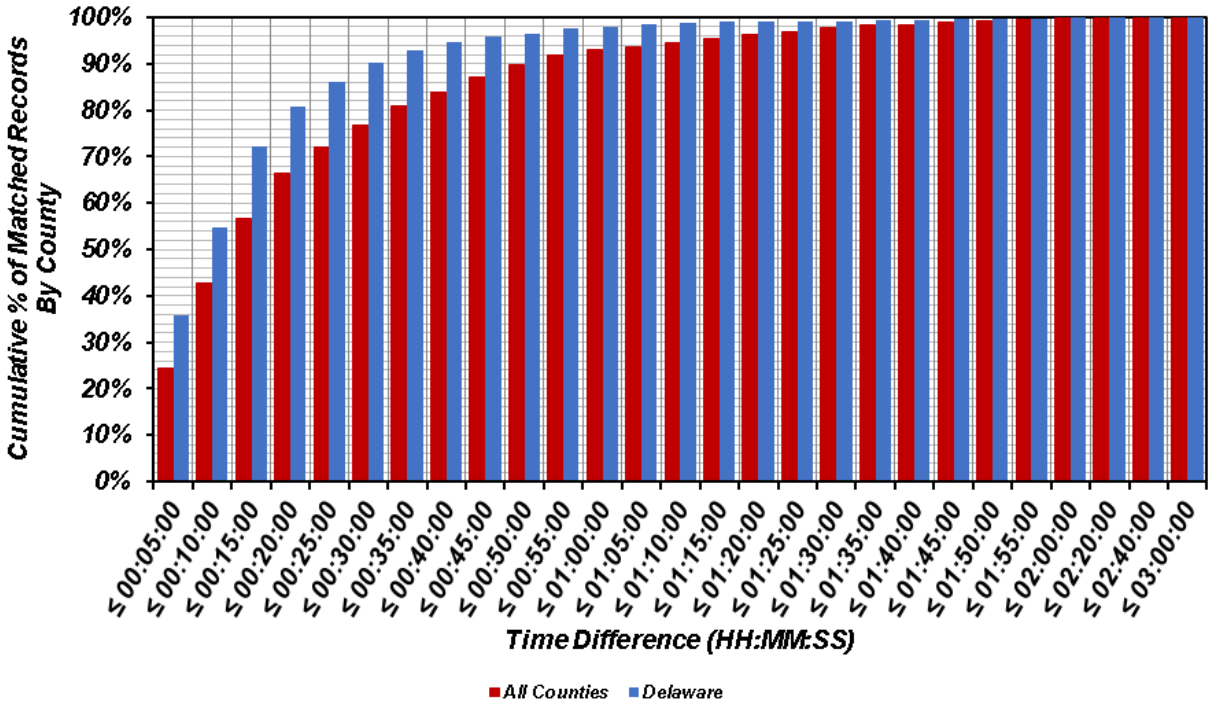



**(b)**

**Figure A.4. Comparison of time difference for matched records in Cumberland County versus all counties: (a) histogram; and (b) cumulative distribution.**

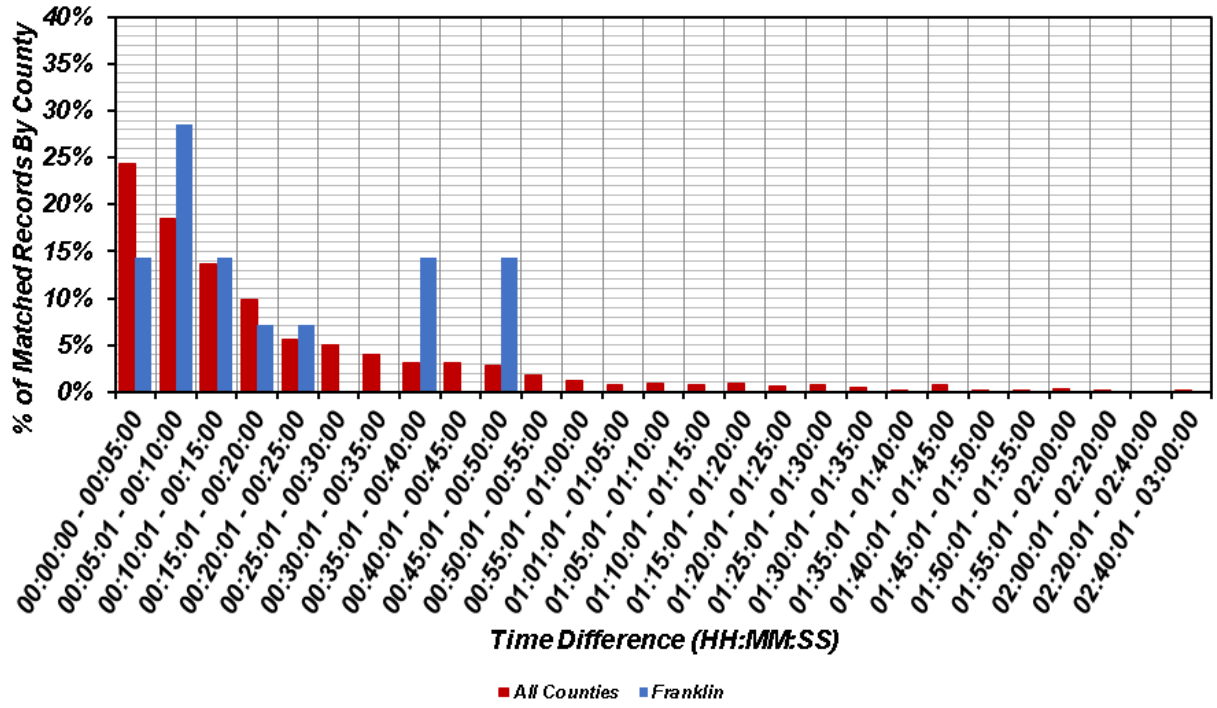



**(a)**

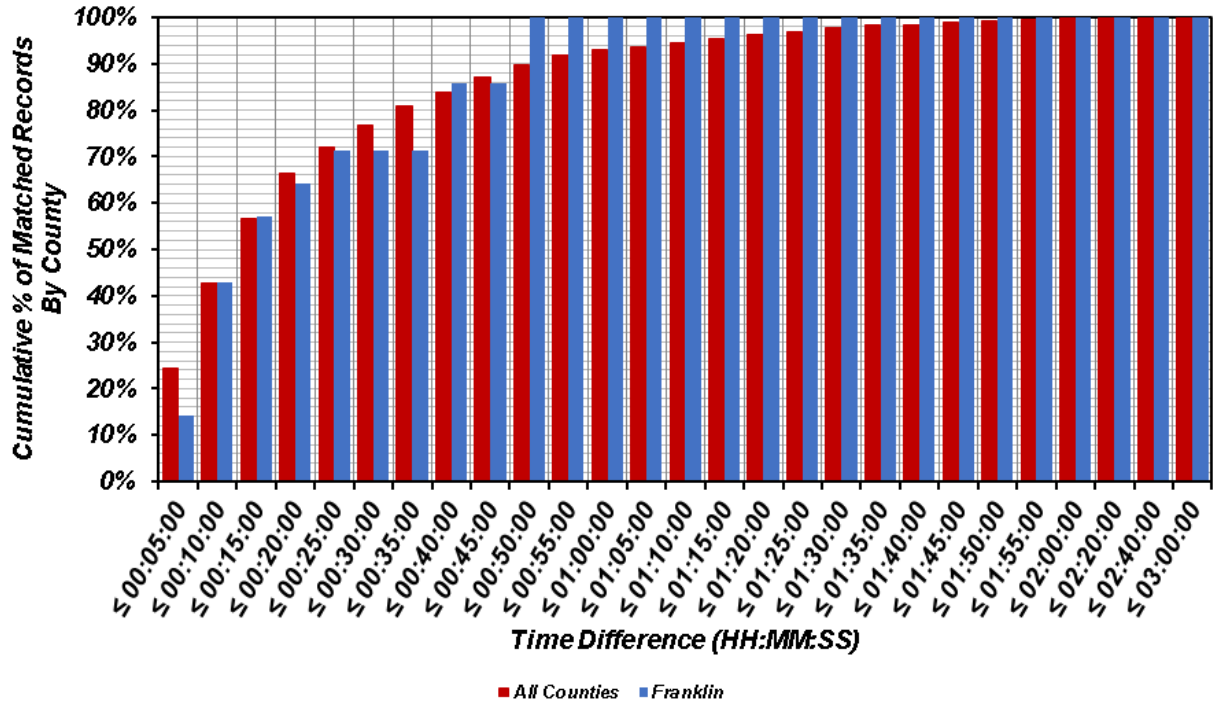



**(b)**

**Figure A.5. Comparison of time difference for matched records in Delaware County versus all counties: (a) histogram; and (b) cumulative distribution.**

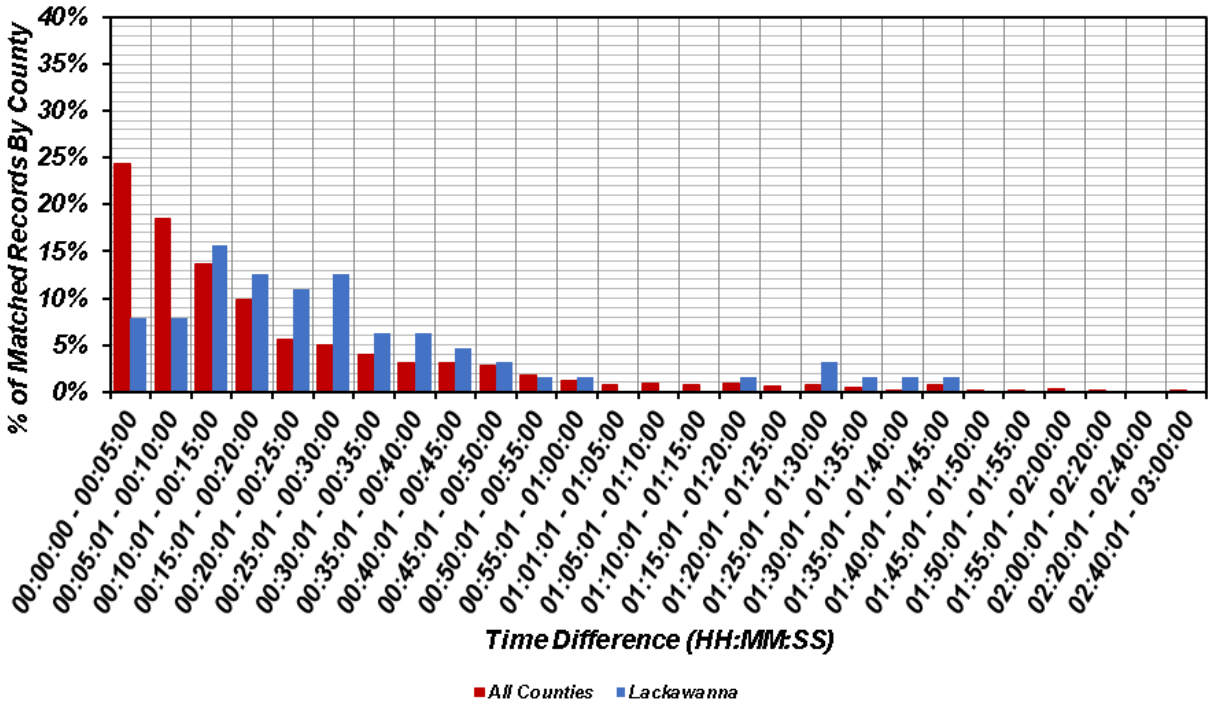



**(a)**

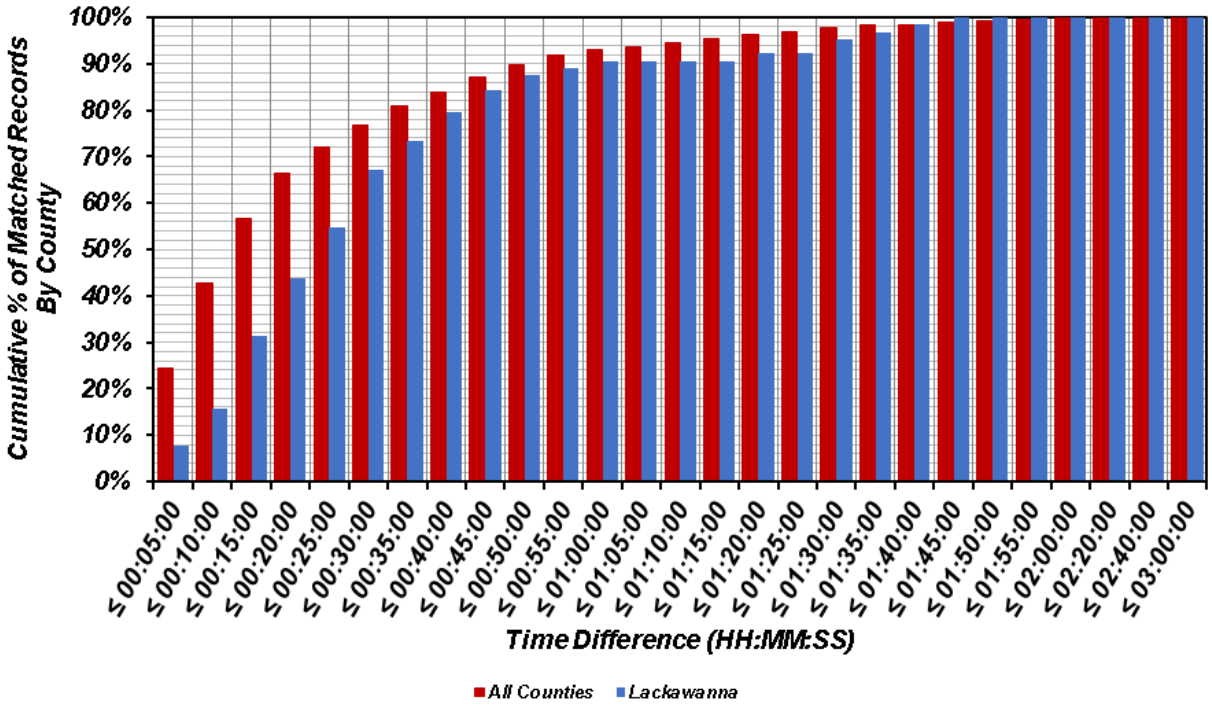



**(b)**

**Figure A.6. Comparison of time difference for matched records in Franklin County versus all counties: (a) histogram; and (b) cumulative distribution.**

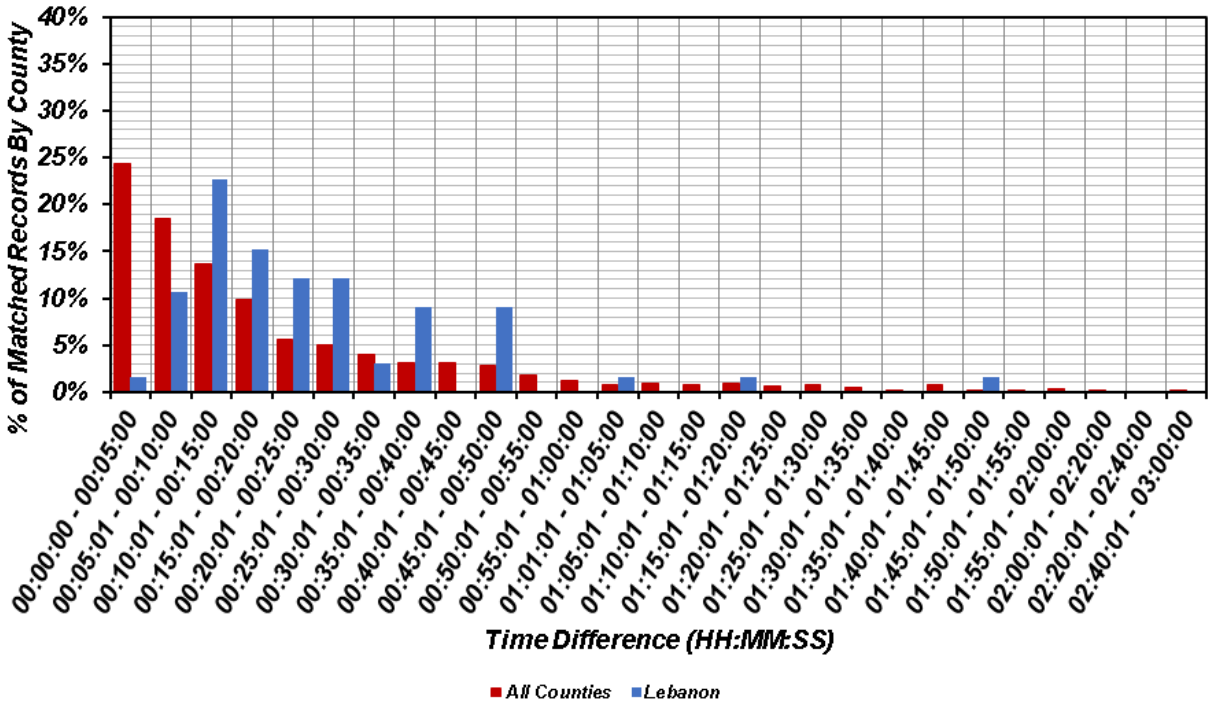



**(a)**

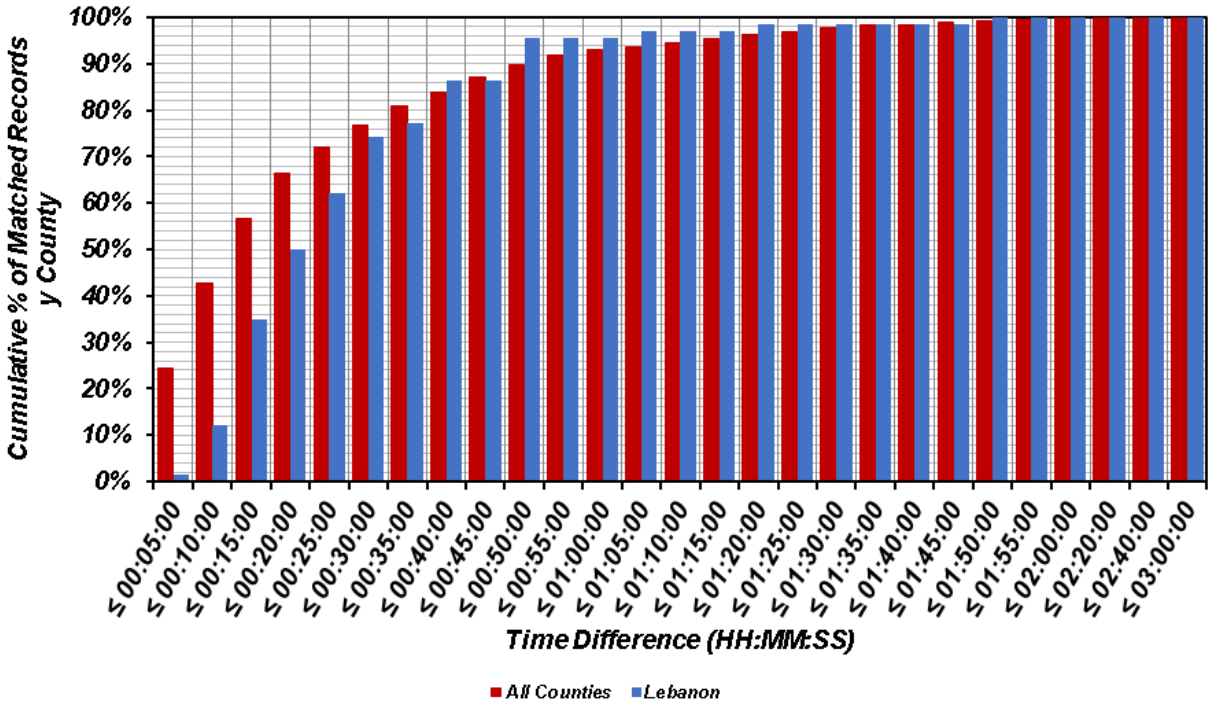



**(b)**

**Figure A.7. Comparison of time difference for matched records in Lackawanna County versus all counties: (a) histogram; and (b) cumulative distribution.**

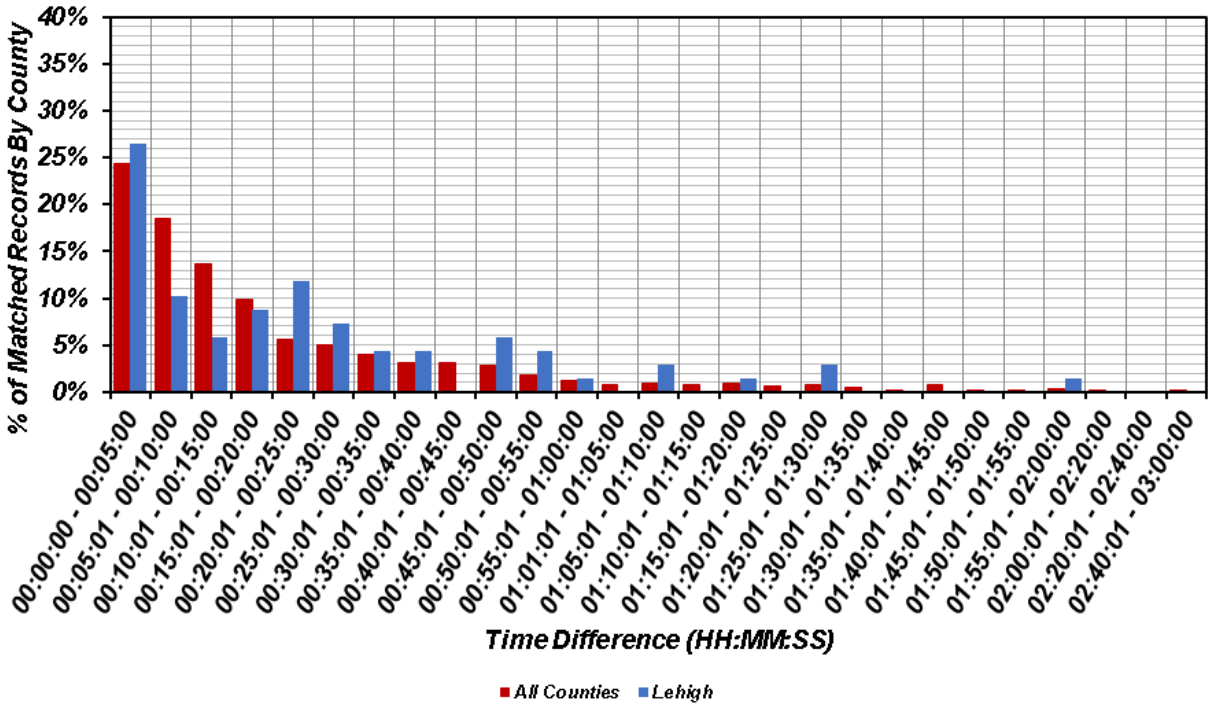



**(a)**

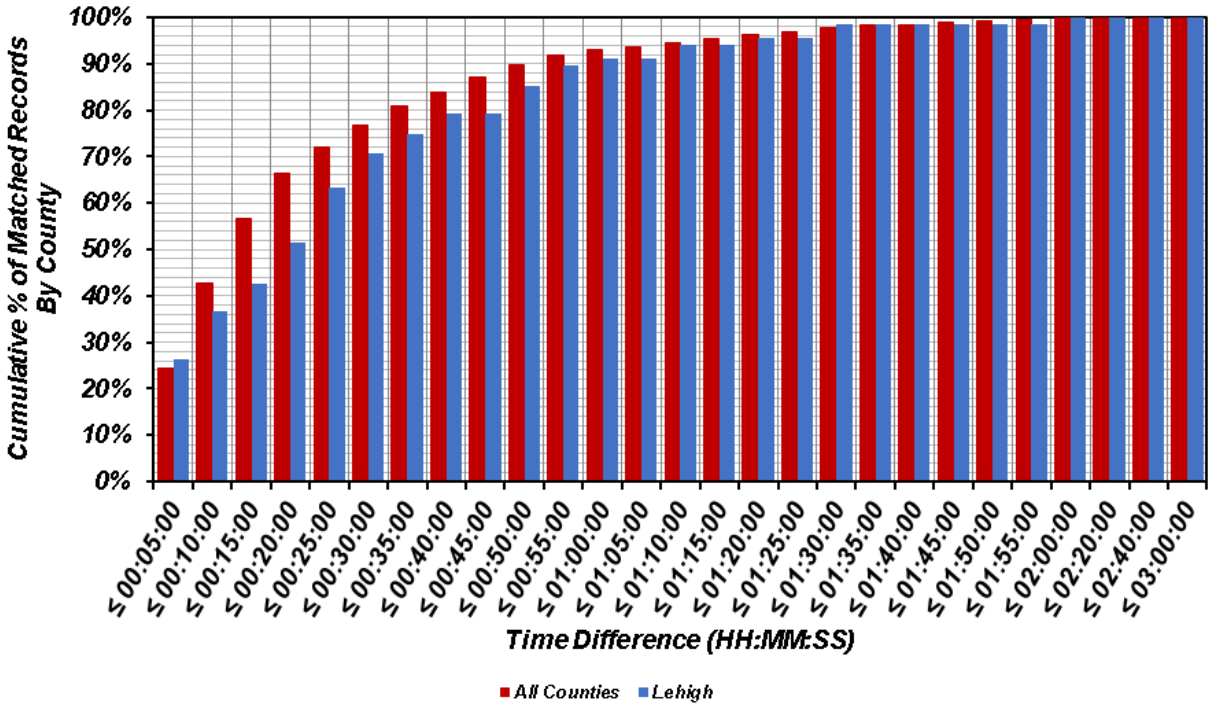



**(b)**

**Figure A.8. Comparison of time difference for matched records in Lebanon County versus all counties: (a) histogram; and (b) cumulative distribution.**

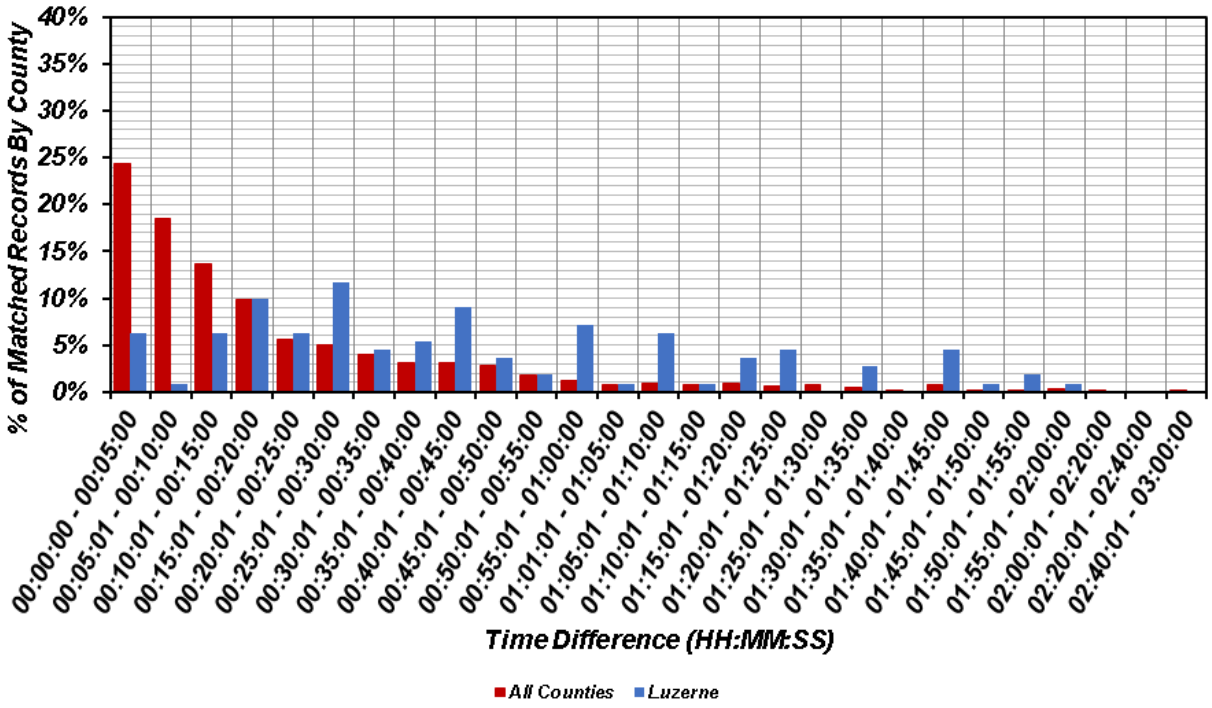



**(a)**

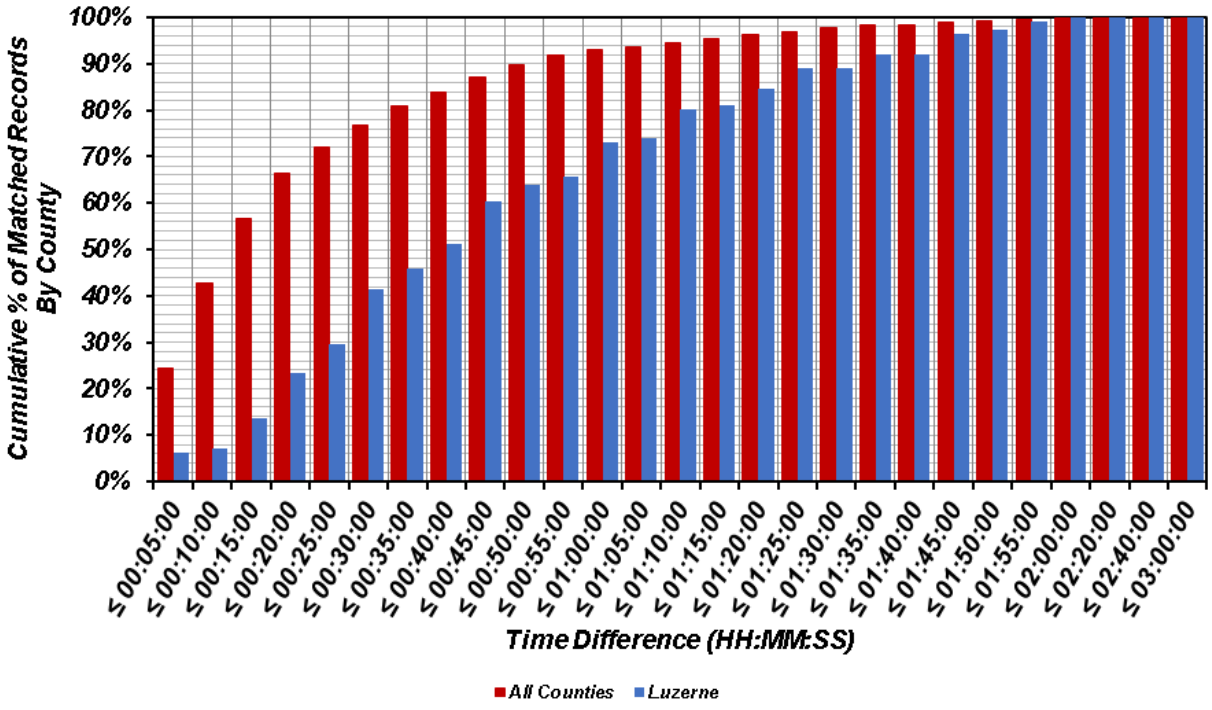



**(b)**

**Figure A.9. Comparison of time difference for matched records in Lehigh County versus all counties: (a) histogram; and (b) cumulative distribution.**

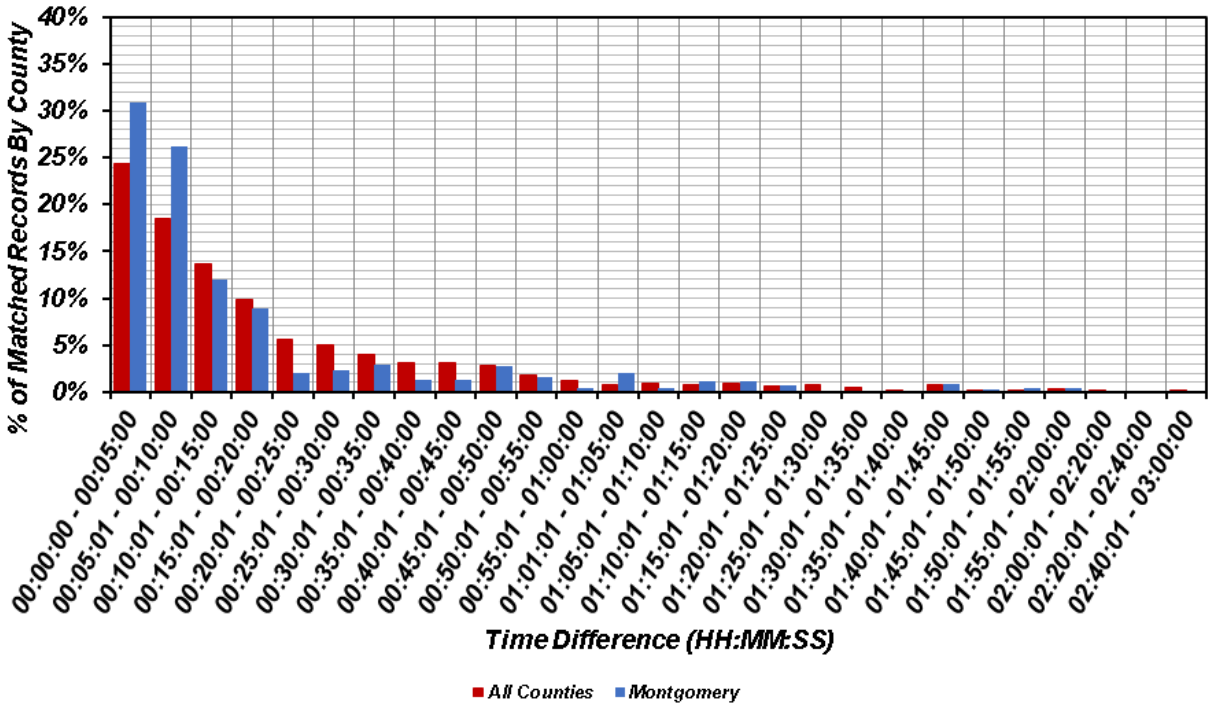



**(a)**

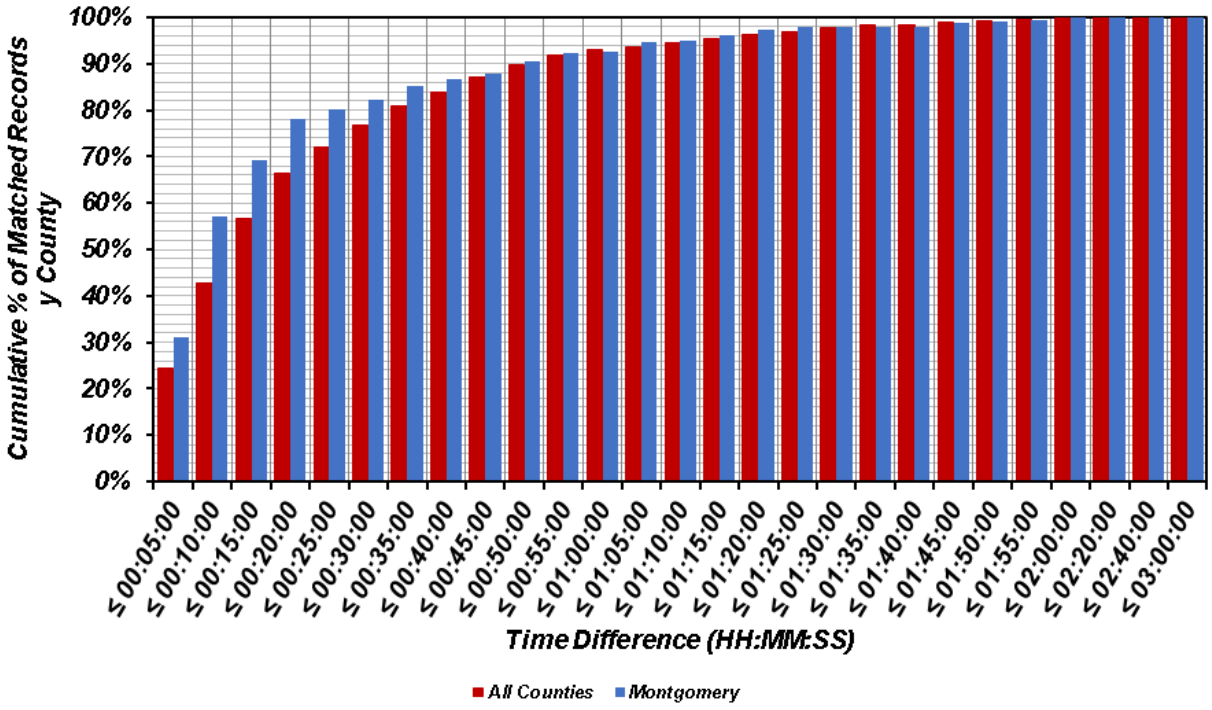



**(b)**

**Figure A.10. Comparison of time difference for matched records in Luzerne County versus all counties: (a) histogram; and (b) cumulative distribution.**

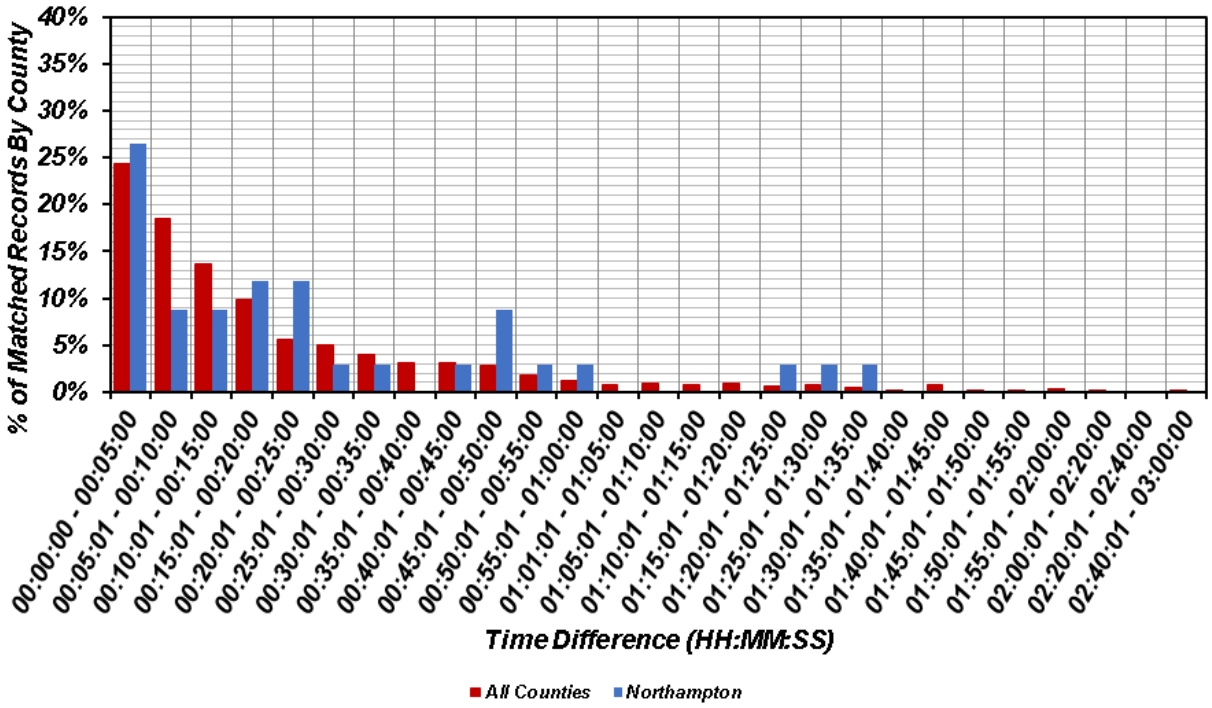



**(a)**

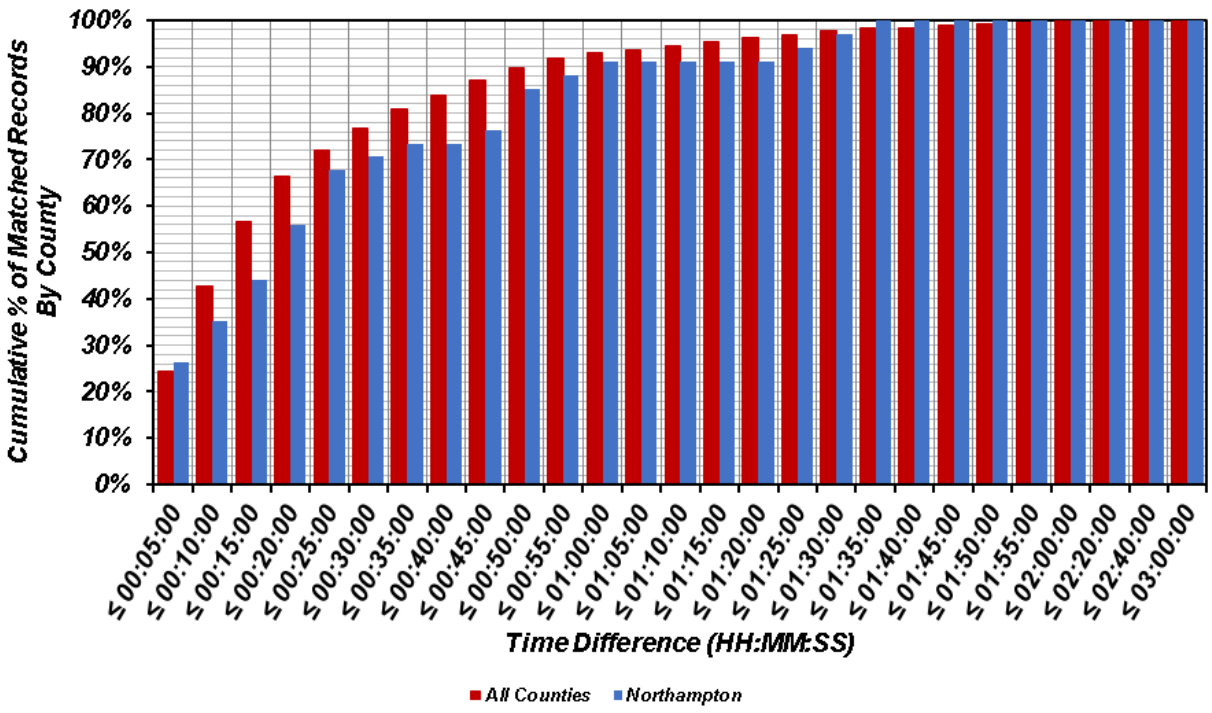



**(b)**

**Figure A.11. Comparison of time difference for matched records in Montgomery County versus all counties: (a) histogram; and (b) cumulative distribution.**

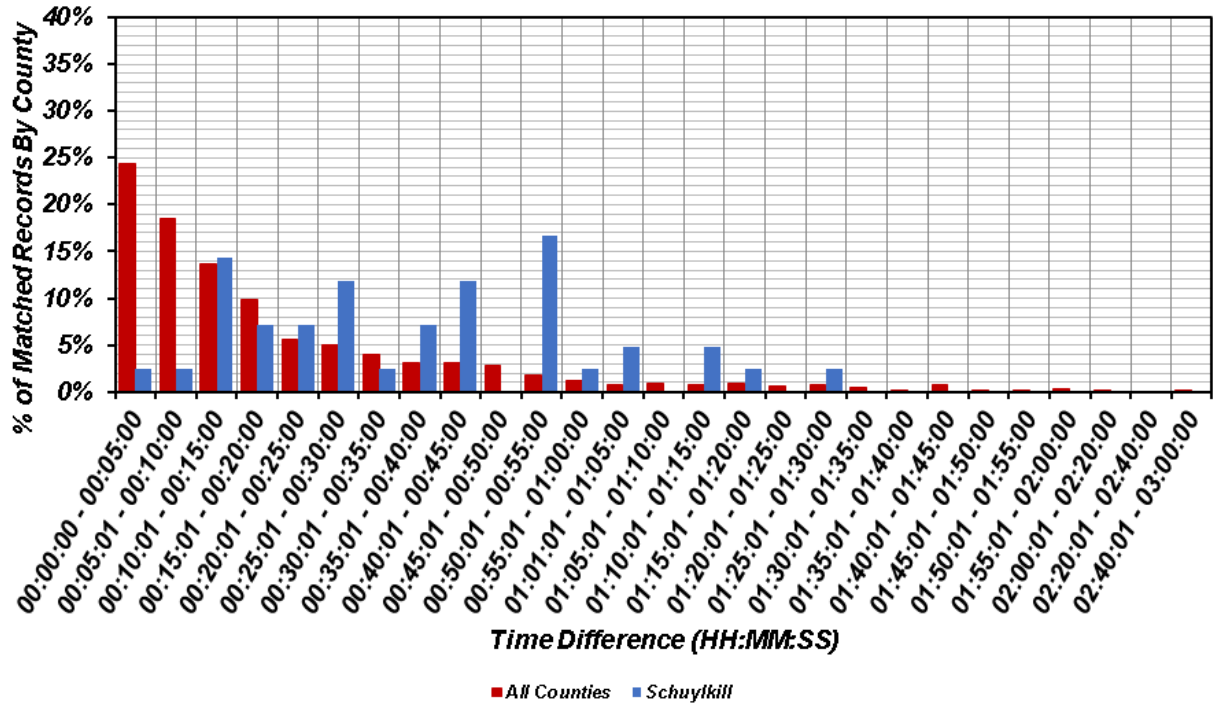



**(a)**

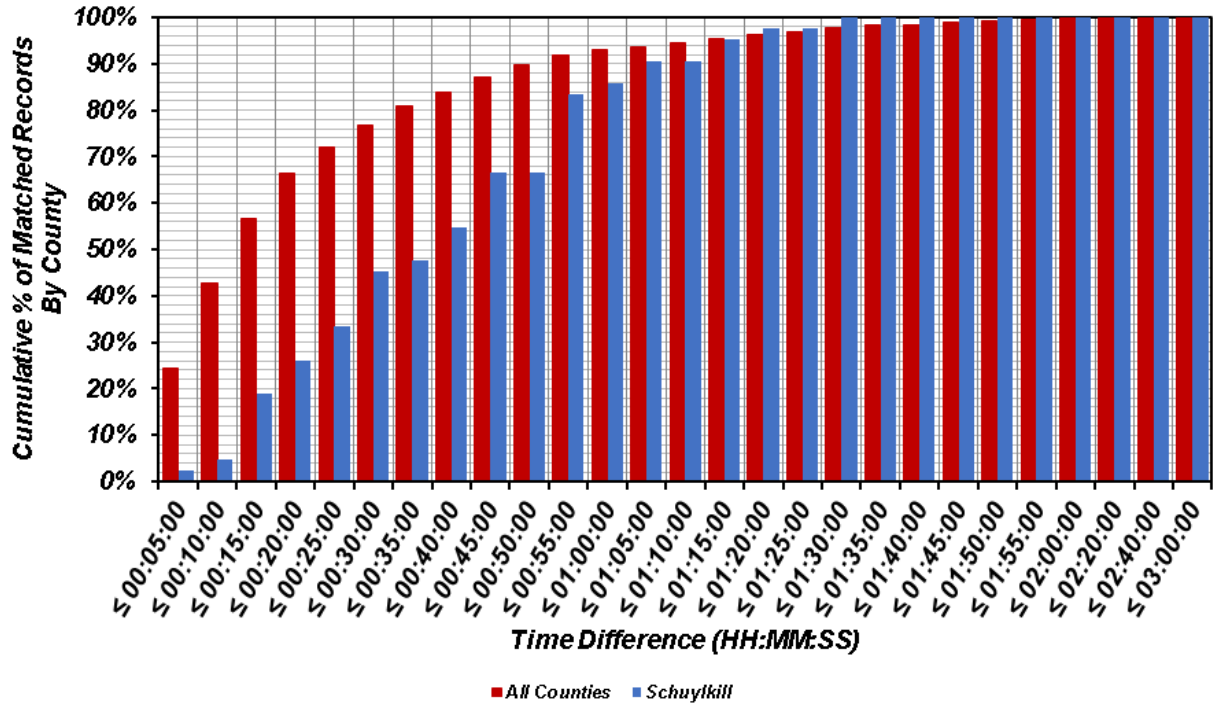



**(b)**

**Figure A.12. Comparison of time difference for matched records in Northampton County versus all counties: (a) histogram; and (b) cumulative distribution.**

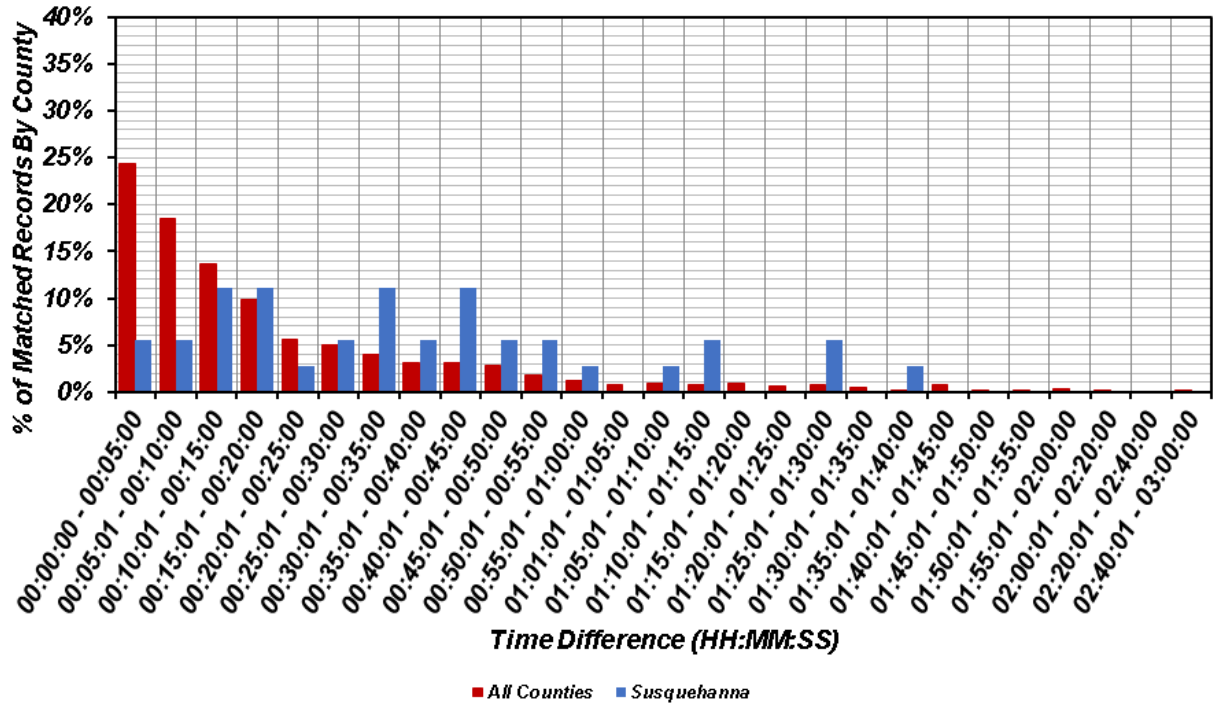



**(a)**

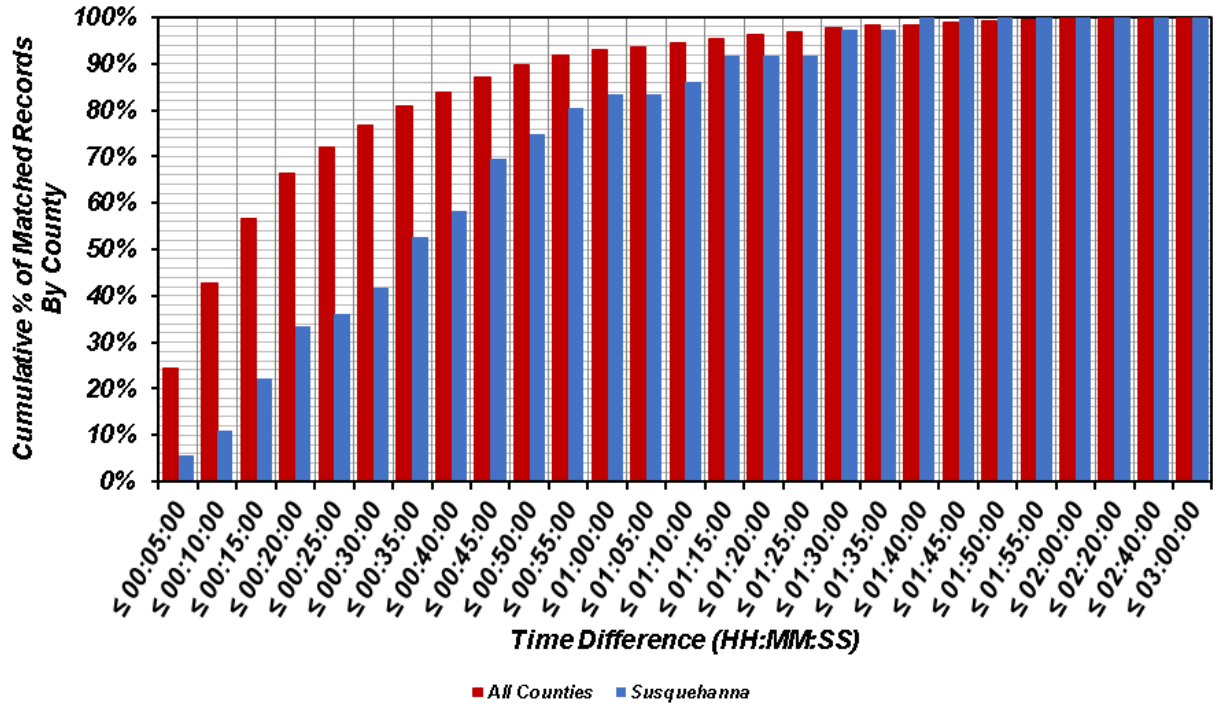



**(b)**

**Figure A.13. Comparison of time difference for matched records in Schuylkill County versus all counties: (a) histogram; and (b) cumulative distribution.**

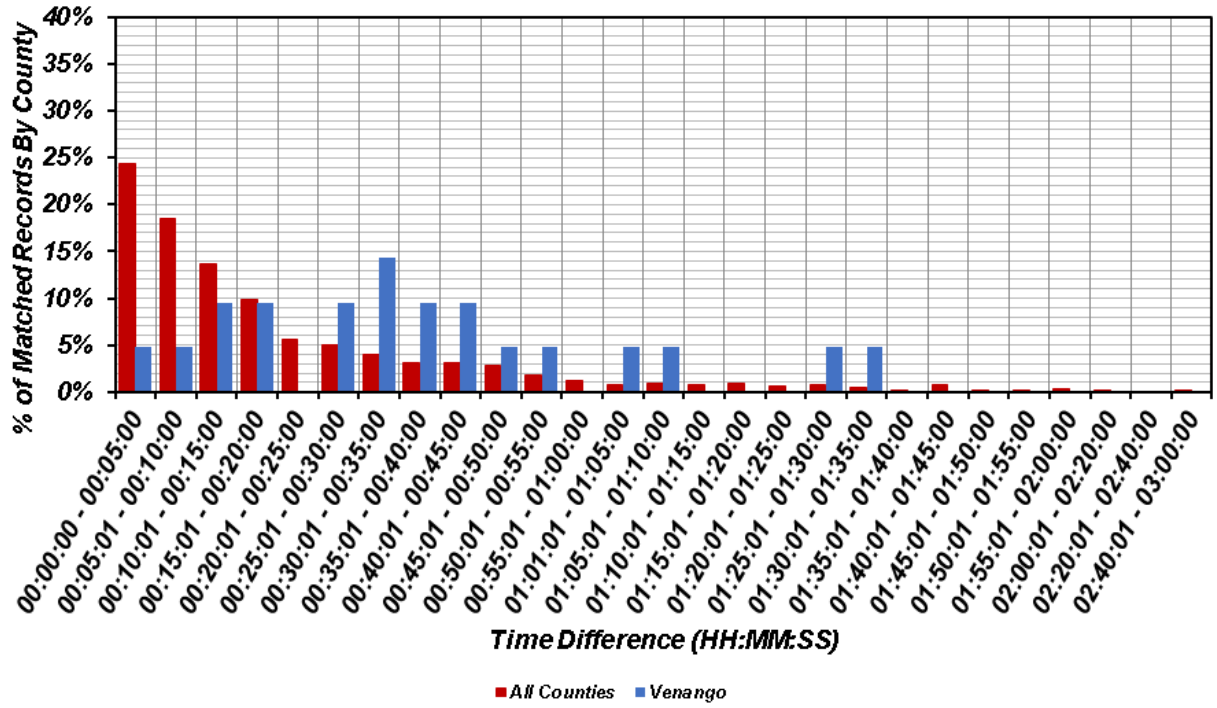



**(a)**

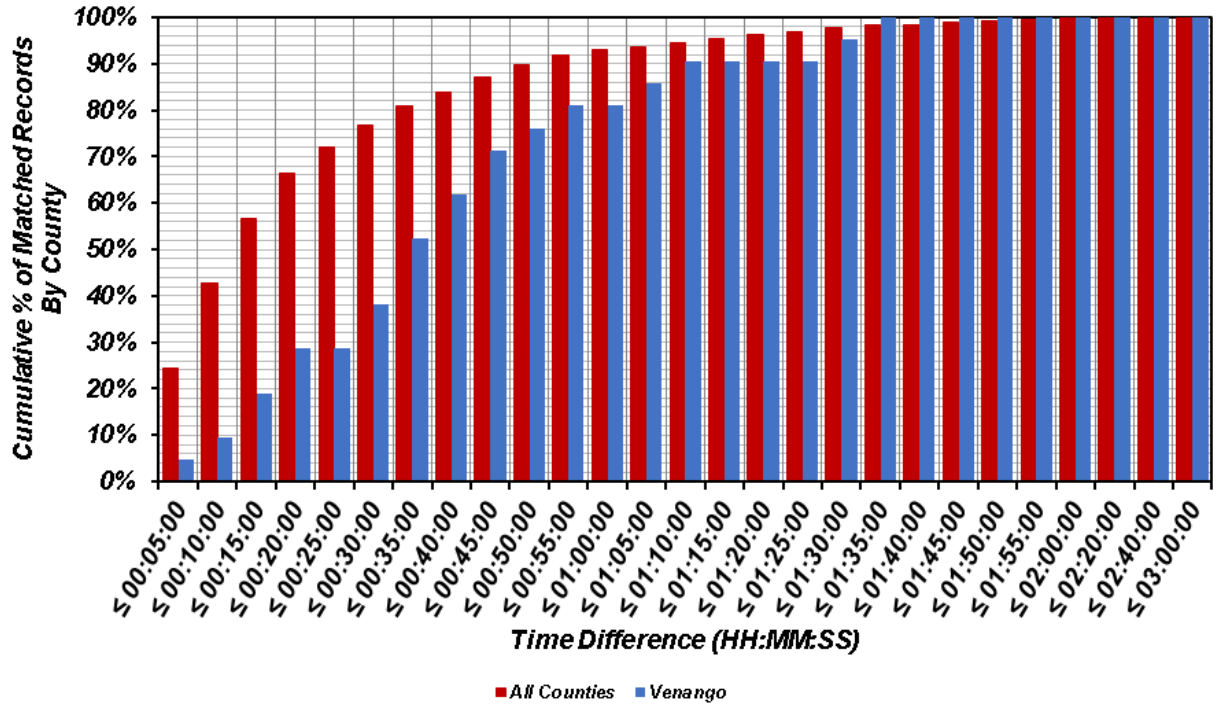



**(b)**

**Figure A.14. Comparison of time difference for matched records in Susquehanna County versus all counties: (a) histogram; and (b) cumulative distribution.**

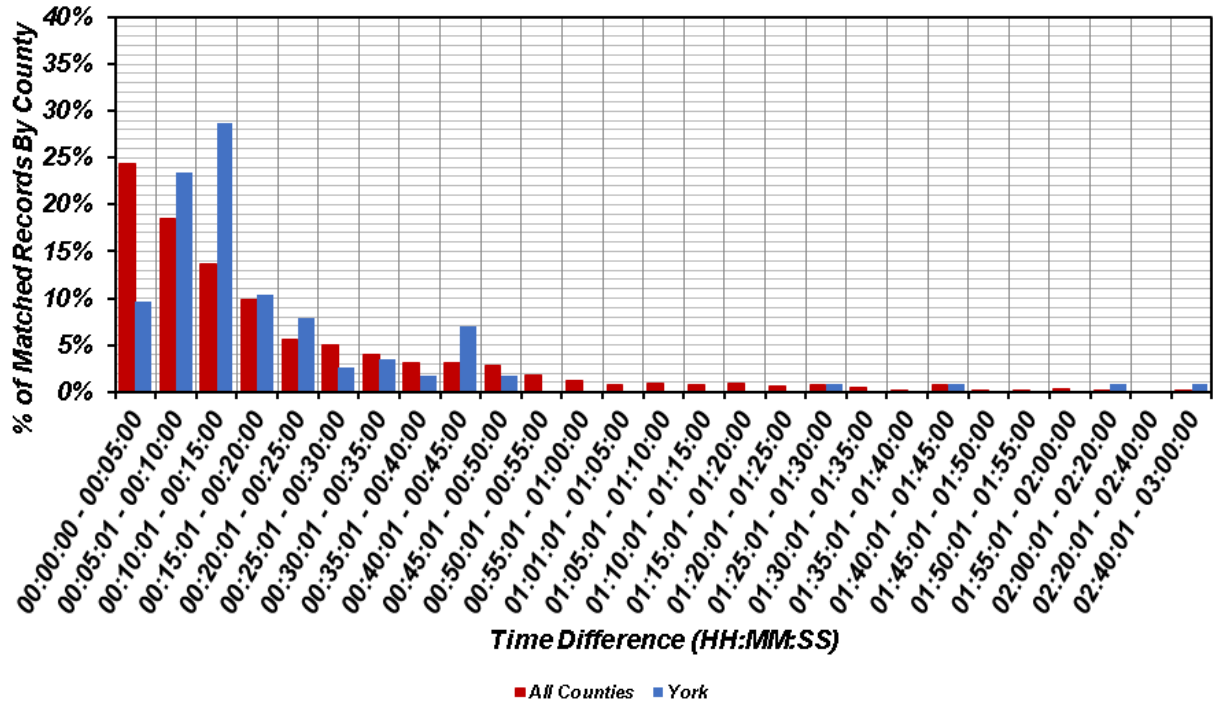



**(a)**

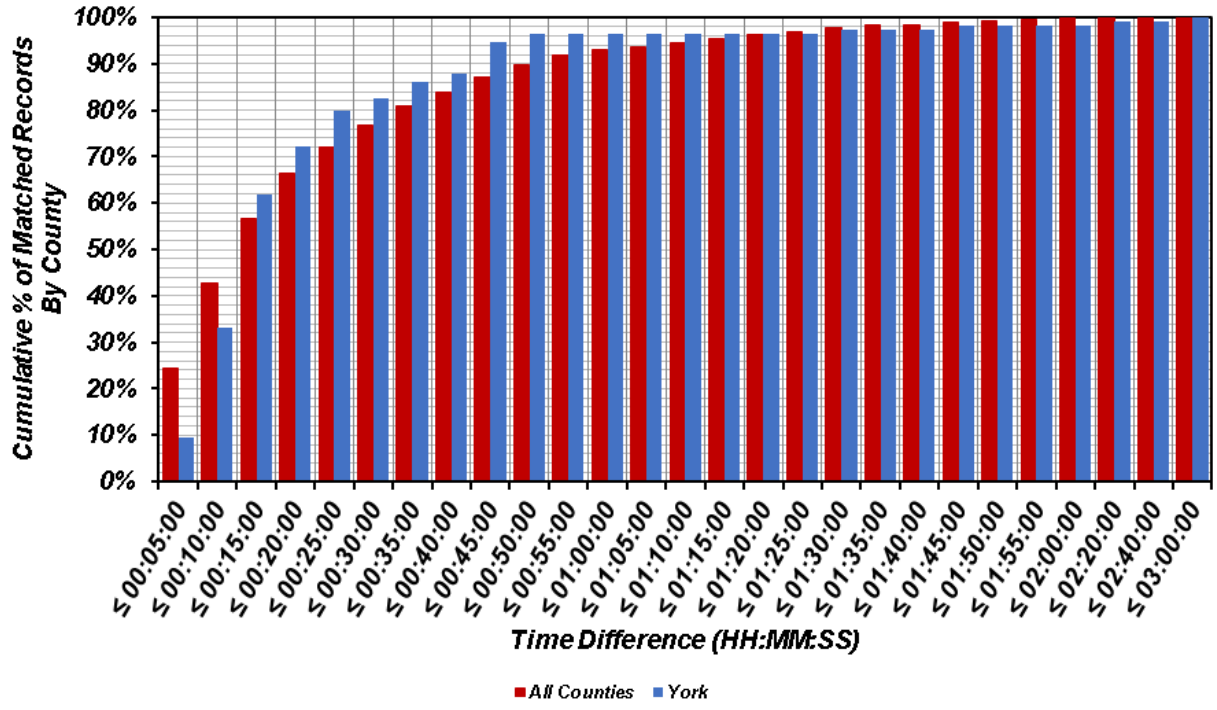



**(b)**

**Figure A.15. Comparison of time difference for matched records in Venango County versus all counties: (a) histogram; and (b) cumulative distribution.**




**(a)**




**(b)**

**Figure A.16. Comparison of time difference for matched records in York County versus all counties: (a) histogram; and (b) cumulative distribution.**