The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Federal Award Number: 15PNIJ-22-GK-00246-BRND

Project Title: SMU Human Trafficking Data Project

Project Director / Principal Investigator:

Elizabeth M. Wheaton-Páramo, Ph.D.

Research Assistant Professor and Senior Lecturer II

Contact Information: htdr@smu.edu, P.O. Box 750496, Dallas, TX 75275, 214-768-6830

Award Recipient Organization: Southern Methodist University, 6425 Boaz Lane, Dallas, TX 75205

Project Period: 03/15/2022 – 09/14/2023

Award Amount: $1,187,000

Summary of the Project

Human trafficking (both labor and sex) continues to be on the agendas of federal, state, and local authorities. Reports and laws state the need for more and higher-quality human trafficking data. Human trafficking efforts by stakeholders, including government, law enforcement, nonprofit organizations, service providers, lawyers, and researchers, call for reliable data collection and analysis. Southern Methodist University's (SMU) Human Trafficking Data Project ("Project") directly addressed the deficiencies in human trafficking data collection and analysis and in the collaboration of expertise needed to ensure data efficacy and complete, effective human trafficking research.

The Project constructed the SMU Human Trafficking Data Warehouse ("Data Warehouse"), created a human computational-machine learning game to scrape and complete human trafficking datasets, and consulted with stakeholders who own datasets, are interested in collecting data, and have the need for human trafficking data for anti-human trafficking research and work. Multiple research projects were conducted to gain an understanding of how current data can be used to determine deeper understanding about human trafficking.

The Data Warehouse provides for the secure collection, storage, cleaning, and filling in the gaps of the wide variety of human trafficking data types and sources. The Data Warehouse can be made available on a secure basis to stakeholders with existing sources of human trafficking data as well as organizations and individuals involved in anti-trafficking work who are interested in collecting data. Secure, online dashboards provide data visualization to present data for dissemination in a format and method that practitioners, policy makers, and researchers can easily access, understand, and use.

Major goals and objectives

Goal 1: Creation of a functioning data warehouse dedicated to human trafficking data

During the 1-year project and within the existing structure of SMU's Office of Information Technology (OIT), a data architect and a data scientist were hired to build the Data Warehouse. These professionals used the existing infrastructure and facilities available to the SMU OIT Team. The SMU Human Trafficking Data Warehouse has been built and is fully functional for securely uploading, storing, cleaning, analyzing, and disseminating publicly available and privately held datasets related to human trafficking. Data dashboards are created for specific user groups and can be accessed to view visuals of specific data by contacting the SMU Human Trafficking Data Research (HTDR) Project team.

Objective 1: Construction of a data warehouse

The data architect designed the human trafficking data warehouse to contain data from multiple disparate source systems into a structured data model. The architect designed and implemented table structures for the different datasets used in the Project, and for new datasets that have been and will be added to the Data Warehouse. Human trafficking data includes a combination of quantitative and qualitative data from nonprofit and government organizations, universities and private researchers, and public-source data. The data architect created repeatable processes to extract from the source system, transform the data as needed, and load it into the destination database. Consultant Natalie Nanasi worked with the data architect to identify valid methods of human trafficking data collection from legal sources. Consultant Jake Schauer assisted in identifying ways that the data warehouse can created to be used with law enforcement (and other sensitive) data. Since each user of the data in the Data

Warehouse must be verified, secure access to the SMU Human Trafficking Data Warehouse is provided on a custom basis and can be requested at htdr@smu.edu.

Objective 2: Collection, analysis, modeling, and visualization of human trafficking data

The data scientist developed procedures to clean and format data entering the Data Warehouse, worked with data owners to determine the most appropriate way to bring de-identified data from the data owners into the Data Warehouse, and created online dashboards (using tools such as Tableau) for custom data visualization by data owners, subject matter experts, and other user groups. The data scientist assisted with data acquisition and formatting of data for the subject matter experts as they built models, performed statistical and predictive analyses, worked with geospatial data, and queried databases. Consultant Natalie Nanasi assisted to identify effective ways to integrate these techniques into data collection and visualization involving law, such as creating strict de-identification rules for data coming into the Data Warehouse to ensure overlapping datasets cannot lead to an individual's identity and adding caveats to each data visualization to make users aware of the limitations on data accuracy and to caution them against misrepresenting the data in reports. Consultant Jake Schauer worked with the data scientist to identify areas of need for law enforcement in human trafficking data and possibilities for using sensitive data.

Goal 2: Construct a gaming system to assist with acquiring human trafficking data

During the 1-year project and within the existing structure of the SMU Guildhall and the SMU Computer Science Department, research was performed that includes human computational techniques to clean, analyze, and enhance human trafficking datasets (e.g. human trafficking case files, websites). This process is required for "dirty data" that, for

example, takes too long to use, is too expensive to process, or has structural problems that make it challenging to analyze. These newly enhanced datasets served as inputs for research and development of methodologies and techniques to integrate human-in-the-loop machine learning via human computation gaming for the identified human trafficking projects.

Objective 1: Creation of the part of the game that effectively captures and fills gaps in human trafficking data

During the 1-year project and within the existing structure of the SMU Computer Science Department, research was completed to develop and create an interactive experience within a commercial game engine (e.g. Unreal, Unity) that cleans, analyzes and/or enhances the human trafficking datasets for the project. One graduate game programmer focused on the programming and engine code required to develop and deploy a game and/or interactive experience. One graduate game artist/designer focused on designing the overall engagement and interaction loops for the interactive experience as well as develop the required art assets (e.g. textures, models, animations, etc.). This part of the game was created and is ready for implementation into a working computer game. The game is a fictional detective game that leverages human computation gaming, a crowd sourcing technique, to collect player feedback that can be used to train a language model and annotate real human trafficking cases. Consultant Natalie Nanasi assisted to identify effective ways to integrate these techniques into data collection involving law, such as creating gaming protocol that randomized data from "case files" within the game so that players of the game are not able to identify an individual though personally-identifiable information or to recreate an actual human trafficking event.

Objective 2: Creation of the part of the game that effectively translates a variety of data sources into useable human trafficking data

During the 1-year project and within the existing structure of the SMU Computer Science Department, the Project created a technique that can be integrated into interactive media (like computer games) and coupled with machine learning techniques. Human trafficking data is often stored in unstructured or inconsistent formats, which makes it difficult to analyze and extract meaningful insights. For example, reports may be written as Word documents without clearly defined fields for essential details such as "date" or "location." Without structured formatting, it becomes challenging to identify patterns, track trends, or compare data across reports. To address this issue, the Project developed a human computation technique for training language models that can process unstructured text, recognize key information, and convert it into a structured format that facilitates analysis and interpretation.

Goal 3: Conduct human trafficking research

Objective 1: Spatial and temporal patterns and outcome disparities in human trafficking data

During the 1-year project and within the existing structure of the SMU Statistics and Data Science Department, integer-valued time series regression models (more precisely, count time series regression models (INGARCH models)) were used to identify the significant economic for sex trafficking convictions in the United States. A wide range of economic variables were considered in the analysis, including consumer price index, labor force participation rate for women, black and Latino populations, and immigration policy. This analysis was conducted for the period 2011-2022 and our results indicate that regression

models indicated that equity market volatility tracker variables related to immigration policy, labor market regulations, along with changes in the consumer price index (CPI) are statistically significant in explaining sex trafficking-related convictions in the United States. A deeper look into the temporal dynamics of this problem showed that policy changes relating to immigration have a more immediate impact on sex trafficking convictions, whereas policy and regulatory changes in the labor market have a more time-lagged impact on sex trafficking convictions.

Count time series regression models (namely INGARCH models) were considered for modeling monthly counts of sex trafficking-related convictions, at the federal level, in the United States during March 2011 to September 2022. The advantages of the INGARCH models over other count time series models were outlined. The response variable in this model was a time series of counts of sex-trafficking related convictions aggregated monthly. The explanatory variables considered was a wide-ranging list of economic variables such as consumer price index, unemployment rate, GDP, labor force participation rate for women, black and latino populations, immigration policy variable. Sex trafficking convictions count time series data was obtained from the U.S. DOJ press releases and processed further at the SMU Human Trafficking Data Warehouse using natural language processing techniques (NLP). Data on the various economic variables were obtained from the Federal Reserve (https://fred.stlouisfed.org/). See the following publication for detailed results. https://link.springer.com/article/10.1007/s00181-023-02549-w

The economic factors considered in this analysis are not an exhaustive list and another direction of research is to gather time series data on other socio-economic variables such as crime rates, refugee numbers, immigration levels, and the number of border encounters, to

name a few, and assess their effects on sex trafficking. Country of origin information for the victims is needed to call these economic factors as pull factors for sex trafficking. Reliable time series data from other countries is also needed to perform a similar analysis on the push factors of sex trafficking and to understand the joint influence of push and pull factors on sex trafficking. The current work deals with federally prosecuted sex trafficking cases and with the availability of similar data at the state level, this modeling approach can be applied to data on state level prosecutions.

Objective 2: Research on extracting and completing human trafficking datasets using the integration of human-in-the-loop machine learning via human computation gaming

During the 1-year project and within the existing structure of the SMU Guildhall and Computer Science Department, research was completed using human computational techniques and methodologies and techniques to integrate human-in-the-loop machine learning via human computation gaming for the identified human trafficking data. This research established a process of creating effective and complete human trafficking data sets for analysis and policy making. At present, many human trafficking datasets are stored in unstructured formats that are difficult to analyze (e.g. a word doc, a press release). This approach can be used to structure datasets and discover buried information (e.g. dates, locations). Results from the research were presented at the (Institute of Electrical and Electronics Engineers (IEEE) Conference on Games in Milan, Italy in August 2024.

Objective 3: Research on human trafficking in the child welfare population

During the 1-year project and within the existing structure of the SMU Economics Department, research was performed to study youth who have experienced maltreatment,

neglect, or abuse during their childhood or have been involved in the foster care system are vulnerable to falling victim to human trafficking. This study included the entire population of children involved in the welfare system across the United States in the context of human trafficking research. By linking data from the National Child Abuse and Neglect Data System (NCANDS) Child Files and the Adoption and the Foster Care Analysis and Reporting System (AFCARS) Foster Care Files, a comprehensive dataset spanning Federal Fiscal Years (FFY) 2018 through 2021 was created that captures children's maltreatment experiences, foster care placements, and potential instances of trafficking. To better understand the vulnerabilities within this high-risk group, this study investigated the primary risk factors associated with the likelihood of becoming a victim of human trafficking among children within the U.S. child welfare system. The research employed logistic regression models to quantify these risk factors across five key categories: demographic characteristics, child-specific factors, caregiver-related factors, living conditions, and child welfare history. Odds ratios (ORs) with 95% confidence intervals were computed to assess the strength and direction of associations between variables. Findings revealed that children aged 16-17 and females are particularly susceptible to becoming victims of human trafficking. Additionally, systemic vulnerabilities, such as prior involvement in child welfare system and history of victimization, as well as non-traditional living conditions like group homes, substantially elevate the risk. The findings from this study can be used by policymakers to create evidence-based policies. Such policies can help detect vulnerability at an early stage, mitigating the risk of young individuals becoming victims of trafficking. Results from the research were presented at the International Atlantic Economic Association Conference in Philadelphia, PA in October 2023.

Objective 4: Research on human trafficking data within the National Incident-Based Reporting System (NIBRS)

During the 1-year project and within the existing structure of the SMU Economics Department, research was performed to compare the demographic characteristics of U.S. nationals and migrants who were trafficked in the United States. The research focused on gaining a picture of how human trafficking prevails in the U.S. based on NIBRS data which shows human trafficking data that was reported to the FBI from law enforcement agencies in the United States. Despite the growth of socio-economic literature on global human trafficking, little emphasis has been given to U.S. residents. Although there are reports about victims who migrated to the United States, there is a large pool of victims among U.S. nationals who remain absent from these studies. This research provided an idea on the different features of human trafficking in the United States. Data from the U.S. National Incident-Based Report System (NIBRS) was used which provided data related to both victims and offenders of human trafficking. There are limitations to the research that can be done with data from NIBRS. For example, the data provides only a picture of the portion of human trafficking that is seen by law enforcement and reported as human trafficking. Many law enforcement agencies have not and do not actively report to NIBRS, or report information incompletely, which leads to large amounts of inaccurate or missing data points. The NIBRS framework will be more accurate once all agencies are reporting. Due to data limits, no causal analysis was possible, however, this study used graphical representations of the human trafficking data found in NIBRS to get some idea of the state of human trafficking in the U.S. According to this data, most victims of human trafficking are white females in their early twenties or late teens and most of the offenders are

black males. In addition, the data points to trafficking being more active during the night and increasing as the night progresses. Most reports show offenses take place at hotels and at the residence of the victim. The NIBRS shows limited information on victim-offender relationships and weapons involved (if any). The research limitations with NIBRS data are that it only paints the portion of human trafficking that is seen by law enforcement. Additionally, many agencies have not and do not actively report to NIBRS, or report incompletely, which leads to large amounts of inaccurate or missing data points. The NIBRS framework will be more accurate once all agencies are reporting. Consultant Jake Schauer worked with the researcher to understand the way law enforcement reports to and interacts with NIBRS. Results from the research were presented at the International Atlantic Economic Association Conference in October 2023.

Objective 5: Research on the economic cost of doing nothing about human trafficking

During the 1-year project and within the existing structure of the SMU Economics Department, research was performed to create a basic structure to estimate the cost of activities in the U.S. to prevent and fight human trafficking, recover victims, prosecute perpetrators, and restore survivors. Anti-human trafficking activities include, but are not limited to, work and funding activities by local, state, and federal governments, the creation of laws, law enforcement policies and procedures, first responder and medical services, prosecutions, data collection and research, education and training programs, victim and survivor care programs, shelter, social support services, businesses and policies related to human trafficking, and volunteer work by survivors, family of victims and survivors, and other members of society. With an enumeration of these and related activities, a counterfactual argument can be asked, what is the cost of doing nothing about human trafficking? This

counterfactual argument can be used to consider how current costs of anti-human trafficking efforts would change and the costs of human trafficking would escalate if human trafficking were not illegal and no efforts were made to prevent and fight human trafficking, recover victims, prosecute perpetrators, and restore survivors.

This research provides a foundation for the SMU Human Trafficking Data Research (HTDR) Project to use the SMU Human Trafficking Data Warehouse to create research partnerships with anti-human trafficking service providers and shelters. It allows research collaboration with nonprofit organizations and service providers to show the cost savings of their anti-human trafficking and support operations compared to the costs to victims, society, and taxpayers of doing nothing about human trafficking. Partnership agreements are in progress to conduct project-specific costs of doing nothing about human trafficking.

Results from the research were presented at the International Atlantic Economic Association Conference in Philadelphia, PA in October 2023.

Goal 4: Creating partnerships between anti-human trafficking stakeholders

The NIJ-funded SMU Human Trafficking Project was the foundation of creating the SMU Human Trafficking Data Research (HTDR) Project with the mission to make human trafficking data work (www.smu.edu/htdr). The SMU HTDR Project Team continues to create partnerships between SMU and individual researchers, universities, research centers, and other entities that have and need human trafficking data. Meetings were held with representatives of national and international universities, nonprofit organizations, the Texas Department of Public Safety, the Texas Office of the Attorney General, all three task forces in North Texas, and funding

agencies to discuss data and research collaboration. The HTDR Project is currently working on agreements to establish partnerships with several organizations for collaborative research.

The SMU HTDR Project Team talked with stakeholders of human trafficking data at multiple conferences and presented at the North Texas Coalition against Human Trafficking (NTCAHT) on November 16, 2023 and at the University of Houston Borders, Trade, and Immigration Institute Conference to End Human Trafficking in July 23, 2024. SMU HTDR Project was accepted into membership as part of the Engage Together North Texas Dream Team and as the center for data analysis as part of the University of Houston Borders, Trade, and Immigration Institute Center for Research Excellence to Counter Human Trafficking (CRECHT).

The first-of-its-kind SMU Human Trafficking Data Conference was held on Southern Methodist University campus on June 3-4, 2024. The conference brought together owners and key stakeholders of human trafficking data and those working with human trafficking data, presented research and innovations on enhancing human trafficking data efficacy, and created opportunity for conversation and collaboration on human trafficking data. The second annual SMU Human Trafficking Data Conference is scheduled for June 2025.

Research Questions

See Goal 3 Objectives 1 through 5.

Research design, methods, analytical and data analysis techniques

See Goal 3 Objectives 1 through 5.

Expected applicability of the research

See Goal 3 Objectives 1 through 5.

Participants and other collaborating organizations

Elizabeth (Beth) Wheaton-Páramo, Ph.D. is a Research Assistant Professor and Senior Lecturer II in the Department of Economics at SMU and is the author of the first textbook on the Economics of Human Rights (www.routledge.com/The-Economics-of-Human-Rights/Wheaton/p/book/9781138500167, including capital punishment, violence against women, seeking asylum, terrorism, child abuse, genocide, and hate) and a white paper modeling the economics of human trafficking (2010). She coordinated the Project work among faculty, staff, and students across the various aspects of the Project, recruited and assisted Project partners, built partnerships between the stakeholders, conducted research on the cost of doing nothing about human trafficking, and organized all NIJ and NACJD paperwork.

Corey Clark, Ph.D. serves as Deputy Director of Research at SMU Guildhall, Assistant Professor in the Department of Computer Science at SMU, and CTO of BALANCED Media | Technology. During the Project, Dr. Clark supervised three graduate students (two within SMU Guildhall and one within SMU Computer Science) on research and development of methodologies and techniques to integrate human-in-the-loop machine learning via human computation gaming for the identified human trafficking projects.

Raanju Sundararajan, Ph.D. is an Assistant Professor in the Department of Statistical Science at Southern Methodist University. Dr. Sundararajan worked with and supervised one SMU Statistics graduate research assistant doing research spatial and temporal patterns in human trafficking data and disparities in outcomes.

Natalie Nanasi, J.D., is an Associate Professor of Law and the Director of the Judge Elmo B. Hunter Legal Center for Victims of Crimes against Women at Southern Methodist University. She provided consultation on legal issues in human trafficking in relation to attorney,

prosecution, and court protocols, visualizing data needs, and creating plans for the collection of relevant human trafficking data related to law and legal procedures.

Lt. Jake Schauer is an Investigations Supervisor for the Alvin Police Department in Alvin, Texas. He provided consultation on law enforcement issues around human trafficking in relation to law enforcement protocols, visualizing data needs, and creating plans for the collection of relevant human trafficking data for law enforcement.

The Institute for Survivor Care ("Institute") is an umbrella nonprofit organization to support the data collection by human trafficking safe houses through a newly acquired client management system (www.instituteforsurvivorcare.org). The SMU HTDR Project worked with the Institute to determine how to collect historical data from human trafficking safe houses.

Changes in approach from original design and reason for change, if applicable

The original proposal for this Project focused on human trafficking research in the areas of public policy and economics and included plans to use a dataset of federally-prosecuted human trafficking cases (of which two exist and are owned by different organizations). However, the final proposal that received funding from NIJ specified that the Project focus on human trafficking data. This shifted the focus from one of research with existing datasets to one of building a human trafficking data warehouse and conducting research on methods of extracting and cleaning data to create useable datasets.

The primary research efforts in this Project were around spatial and temporal patterns and outcome disparities in human trafficking data and extracting and completing human trafficking datasets using the integration of human-in-the-loop machine learning via human computation gaming. Preliminary research was conducted on human trafficking in the child welfare system,

demographic characteristics of U.S. nationals and migrants who were trafficked in the United States, and the cost of doing nothing about human trafficking. A large part of this effort was spent on determining how to access data related to human trafficking and how to build research agendas that would in-part answer questions related to human trafficking. The benefit of this process has been to inform the SMU Human Trafficking Data Research (HTDR) Project on ways to assist data owners and stakeholders on creating datasets, assembling expert research teams, and conducting research utilizing the SMU Human Trafficking Data Warehouse.

Outcomes

- **Activities/accomplishments**

  See Goals 1, 2, and 4.

- **Results and findings**

  See Goal 3.

- **Limitations**

Due to the time and resources needed to build the human trafficking data warehouse, work with data owners and key stakeholders, and conduct research on methods for creating useable human trafficking datasets, time and resources were not available to conduct further research into the overlap in datasets. This was due in part because there was little discernable overlap in the few datasets used within this Project. The five research areas for the Project utilized specific datasets so that overlap of datasets was not necessary. New partnerships are in progress that will bring more datasets into the data warehouse. With this increase in data, research will progress on conducting data analysis on overlapping datasets.

Artifacts

- **List of products**
    - Website: SMU Human Trafficking Data Research (HTDR) Project. www.smu.edu/htdr
    - Customized, secure access to data within the SMU Human Trafficking Data Warehouse can be requested at htdr@smu.edu.
    - Publication: Jang, Y., Sundararajan, R.R., Barreto-Souza, W., Wheaton-Páramo, E. (2024). Determining economic factors for sex trafficking in the United States using count time series regression. Empirical Economics, 67: 337–354. https://doi.org/10.1007/s00181-023-02549-w
    - Conference paper: Buongiorno, S. and Clark, C. (August 2024). A Framework for Leveraging Human Computation Gaming to Enhance Knowledge Graphs for Accuracy Critical Generative AI Applications. IEEE Conference on Games in Milan, Italy. https://docs.google.com/document/d/1sfLTnHZ5szsaXmNEZgpTDvkpb8Ct1Wk3up564knjeK0/edit
    - Conference paper: Wheaton-Páramo, E., Hossain, M.T., and Mitra, A. (October 2023). Estimating the cost of doing nothing about human trafficking. 96th International Atlantic Economic Conference in Philadelphia, PA. https://iaes.confex.com/iaes/96am/meetingapp.cgi/Paper/18411
    - Conference paper: Hossain, M.T., Mitra, A., and Wheaton-Páramo, E. (October 2023). Human trafficking in the child welfare population. 96th International Atlantic Economic Conference in Philadelphia, PA. https://iaes.confex.com/iaes/96am/meetingapp.cgi/Paper/18343

- Conference paper: Mitra, A., Hossain, M.T., and Wheaton-Páramo, E. (October 2023). Human trafficking in the US: What do the data say?. 96th International Atlantic Economic Conference in Philadelphia, PA. https://iaes.confex.com/iaes/96am/meetingapp.cgi/Paper/18413

- **Data sets generated**
  - The United States Federal Bureau of Investigation collects data every year on incidents reporting by law enforcement agencies, known as the National Incident-Based Reporting System (NIBRS). The most granular level of data that is available is found on the FBI Crime Data Explorer at the yearly and state level. Human Trafficking was added as a crime category in 2013. We have downloaded all NIBRS data from 2013 to 2022.
  - The Department of Justice releases press releases for all federally prosecuted crimes. There have been organizations who use the provided API to download these press releases, sort them by category, and then process them using Natural Language Processing (NLP) techniques. The raw data pulled from the API for 2011-2022.
  - Client intake data was shared from urban shelters in the United States from 2010 to 2021 whose clients were survivors of human trafficking. All data was de-identified and standardized.

- **Dissemination activities**
  - The first-of-its-kind 2024 SMU Human Trafficking Data Conference was held in Dallas, Texas on June 3-4, 2024. The Conference convened over 100 professionals from different areas of anti-human trafficking to discuss the current state of data and data analytics in combatting human trafficking nationally and globally. Breakout sessions

included tracks for law enforcement, nonprofit organizations, and research and data science. Multiple collaborations were discussed on data initiatives to help survivors while developing new methods and techniques to track down and prosecute human traffickers.

- The **2025 Human Trafficking Data Conference**, to be held June 9 and 10, 2025 on the SMU campus, will allow for a larger audience of participants. There is a narrower focus to pinpoint specific areas of expertise (both national and international), strengths and weakness of current human trafficking data, effective strategies to increase data efficacy, and opportunities for collaboration. The conference scope is broader as breakout session tracks expand to include medical, legal, and survivor leadership.
- Multiple new partnerships are in progress that will bring more datasets into the data warehouse and allow for research into techniques to collect, store, clean, fill in the gaps, overlap, analyze, and disseminate human trafficking data.