A Framework for Data Quality: Case Studies

October 2023

Recommended citation: Mirel LB, Singpurwalla D, Hoppe T, Liliedahl E, Schmitt R, Weber J. 2023. *A Framework for Data Quality: Case Studies*, FCSM-23-02 Data Quality Framework Implementation Subcommittee, Federal Committee on Statistical Methodology. October 2023.

2023 Federal Committee on Statistical Methodology

Members

Ann E. Carson
Bureau of Justice Statistics

Chris Chapman
National Center for Education Statistics

Jennifer Hunter Childs
U.S. Census Bureau

Tala Fakhouri
U.S. Food and Drug Administration

John Finamore
National Center for Science and Engineering Statistics

Ellen Galantucci
Federal Maritime Commission

Michael Hawes
U.S. Census Bureau

Travis Hoppe
National Center for Health Statistics

Rochelle (Shelly) Wilkie Martinez
Office of Management and Budget

Wendy Martinez
U.S. Census Bureau

Lisa B. Mirel
National Center for Science and Engineering Statistics

Jennifer Nielsen
National Center for Education Statistics

Jennifer M. Ortman
U.S. Census Bureau

Anne Parker
Internal Revenue Service

Jennifer Parker
National Center for Health Statistics

Mark Prell
Economic Research Service

Michael Ratcliffe
U.S. Census Bureau

Rolf Schmitt
Bureau of Transportation Statistics

Darius Singpurwalla
National Center for Science and Engineering Statistics

Bob Sivinski
Office of Management and Budget

Matt Spence
U.S. Census Bureau

Ed Strocko
Bureau of Transportation Statistics

G. David Williamson
Centers for Disease Control and Prevention

Linda Young
National Agricultural Statistics Service

Data Quality Framework Implementation Subcommittee Members

(in alphabetical order)

Travis Hoppe

National Center for Health Statistics

Erika Liliedahl

Office of Management and Budget

Lisa B. Mirel

National Center for Science and Engineering Statistics

Rolf Schmitt

Bureau of Transportation Statistics

Darius Singpurwalla

National Center for Science and Engineering Statistics

Julie Weber

National Agricultural Statistics Service

Table of Contents

Introduction

The Federal Committee on Statistical Methodology (FCSM) released A Framework for Data Quality in September 2020 to provide a common foundation upon which federal agencies can make informed decisions about the quality of data products and their management throughout their life cycle. The framework introduces data program managers and analysts—including statisticians, chief data officers, and evaluation officers—to a broad range of quality considerations and provides a common, systematic language for communicating data quality issues and methods for resolving or accepting those issues. The framework includes quality issues beyond bias and accuracy, the traditional focus of statisticians, and encompasses data from sources including surveys, administrative records, monitors, and blended data.

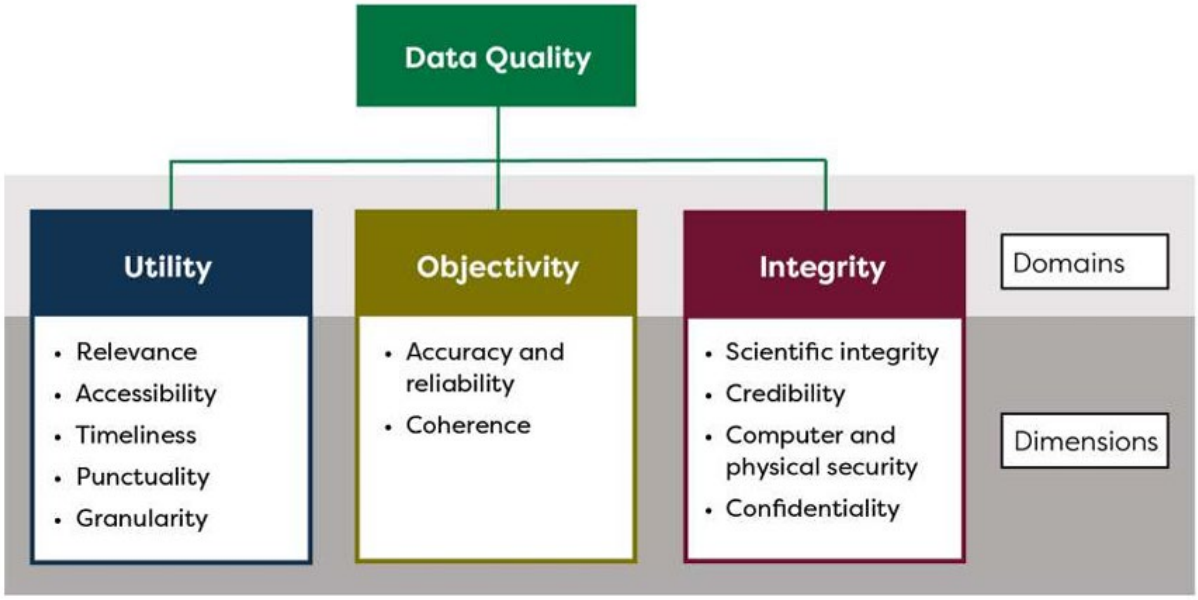
The framework has evolved into a tool for considering data quality challenges in the design of a data program or analysis, documenting how each challenge is resolved during implementation of the data program or analysis, converting documentation into user guides and summary caveats to encourage proper use of the results, and evaluating the quality of the data program or analysis itself.

The FCSM Framework for Data Quality implementation subcommittee has compiled seven case studies from a variety of agencies (e.g., Centers for Disease Control and Prevention [CDC]/National Center for Health Statistics [NCHS], Bureau of Labor Statistics [BLS], and the Department of Transportation [DOT]) to illustrate how the framework was applied in different scenarios. The case studies cover how the framework can be used to assess data quality with different data sources, in different stages of data collection and release, and in response to specific threats to data quality.

The case studies were selected for inclusion in this report to provide readers with a variety of scenarios demonstrating how the framework could be used to address various threats to data quality.

Each case study addresses the three domains and 11 dimensions of the framework. Each data source provides unique insight into different aspects of the framework.

Figure 1. FCSM Framework for Data Quality



The first case study assesses the quality of the NCHS's Linked Mortality Files which blend survey and mortality data. Since blended data can increase the disclosure risk of a dataset, this case study describes the NCHS's procedures and methods to reduce disclosure risk, an important aspect of maintaining confidentiality, in addition to other aspects of data quality.

The second case study describes the data quality assessments of a new method for collecting Consumer Price Index gasoline price data from retailers or data aggregators instead of a sample survey at BLS. While crowdsourcing data directly from retailers leads to efficiencies in both collection efforts and costs, the method does introduce the potential for increased errors during collection. This case study highlights how BLS mitigated threats to the accuracy and reliability of these crowdsourced data.

The third case study describes a new tool, SafeSpect, to collect roadside inspection data at DOT, highlighting the framework's utility domain and, more specifically, the relevance and timeliness dimensions. With respect to relevance, the tool continues to provide highly relevant data that support the Federal Motor Carrier Safety Administration's (FMCSA) mission. The case study describes in some detail how the new tool improves the timeliness of making the data available for review.

The fourth case study assesses the utility of physical activity monitor data that are collected as part of the National Health and Nutrition Examination Survey (NHANES) conducted by CDC/NCHS. The case study focuses specifically on questions of accessibility that can arise when working with large data files that require subject matter expertise to analyze, in addition to other aspects of data quality. Resources are offered to help mitigate some potential threats.

The fifth case study examines the creation of import and export price indexes at BLS and how the FCSM framework can be used to effectively assess the objectivity of the data. This case study discusses the

need for coherence when creating an index as well as accuracy and reliability to mitigate the threat of bias. The authors offer potential strategies to assess possible bias.

The sixth case study uses the FCSM Framework for Data Quality to demonstrate how the data quality framework can help data collectors plan a new study. In this case study, a data producer intends to launch a new study that investigates the public's perceptions of the science and engineering enterprise within the United States. The FCSM Framework for Data Quality was utilized during the planning stages of this study to help account for and mitigate potential threats to data quality.

The seventh case study examines how the FCSM Framework for Data Quality serves as a tool to determine the utility of administrative data in determining program effectiveness and provides considerations for the data's use that are paramount to upholding rigor and ethics as principles of program evaluation.

This report brings together the case studies into a cohesive document and provides a quick guide and template for implementing the framework. These diverse case studies highlight how the framework can be used by policymakers and researchers to gain a deeper understanding of the data and its fitness for purpose at different stages of data collection and release while providing strategies for documenting and reporting data quality in a systematic and consistent way.

A Roadmap on How to Use the FCSM Data Quality Framework

All data have quality problems, whether the data come from surveys, administrative records, sensors, or a blend of multiple sources. Our challenge as creators and users of data is to minimize quality problems within the time and resources available and to articulate the problems that remain to encourage appropriate uses of the data and analyses by both sophisticated and novice consumers of our work.

Data quality assessments are vital to understanding the fitness for purpose of a certain data resource. Decisions are increasingly driven by -empirical data, and the data quality framework is part of an evolving set of tools to help decision makers and the public learn from the past, understand the present, and plan for the future.

The FCSM Framework for Data Quality provides a comprehensive list of quality challenges for consideration to address the following objectives:

- Design new data collections, estimation methods, and analyses.
- Take notes for future managers of a data or analysis program to explain how one overcame each quality challenge.
- Turn notes into guides and documentation for users of data collections and analyses.
- Explain succinctly to executive decision makers and the public how far the data and analyses can be taken to answer the questions at hand.
- Learn the quality of the data collections, estimation methods, and analyses and determine how it could be done better in the future.

Because the framework is designed to cover all forms of data, some elements may not be relevant to a specific data collection, estimation method, or analysis. Therefore, some case studies do not address every dimension. Specific quality issues and trade-offs among quality challenges vary by data source and application. Data program managers and analysts must make judgments on the importance of data quality challenges and deal with those challenges as a normal part of their work. The framework provides a qualitative assessment to help assure that all relevant quality problems are considered, and the framework provides a common language for communicating how the relevant challenges were ameliorated and what challenges or caveats remain.

- Click here to learn about the structure and roots of the data quality framework as well as how the framework was constructed.
- Click here to learn about individual elements of the framework.
- Click here to learn about a case study using blended data.
- Click here to learn about a case study using monitoring data.
- Click here and here for examples of how the framework was used to assess design and implement data collections.
- Click here and here for examples of how the framework was used to evaluate the quality of a data source to create an index.
- Click here for a case study from program evaluators and evaluation officials.
- Click here to send comments and questions to the authors of the framework.

Case Study 1: The NCHS Linked Mortality Files

Authors: Lisa B. Mirel, National Center for Science and Engineering Statistics, Jonathan Aram, National Center for Health Statistics (NCHS), Christine Cox, NCHS

Overview: This case study uses the Framework for Data Quality to assess the quality of National Center for Health Statistics (NCHS) Linked Mortality Files (LMFs), which blends survey and mortality data. Since blended data can increase the disclosure risk of a dataset, this case study describes the procedures and methods used by NCHS to systematically address the dimension of confidentiality ,in addition to other dimensions of the framework.

Introduction: Federal data are used to inform federal policy, evaluate programs, drive business and economic decisions, and improve lives. New sources of data, particularly linked or integrated data sources, must include transparent assessments of data quality to allow users to understand the strengths and limitations of the data as well as the appropriateness of the data for its intended use. The FCSM recently published A Framework for Data Quality to support federal agency efforts to report on aspects of data quality for all federal data collections across three main domains: utility, objectivity, and integrity[1]. This case study demonstrates how the FCSM framework can be used to effectively assess the integrity of NCHS linked data products, particularly the risk of disclosure posed by linking datasets.

Description of data being assessed for quality: Data from various NCHS surveys have been linked with death certificate records from the National Death Index (NDI). Linkage of the NCHS survey participant data with the NDI mortality data provides the opportunity to conduct a vast array of outcome studies designed to investigate the association of a wide variety of health factors with mortality. These NCHS LMFs have been cited in more than 1,000 peer-reviewed publications since their initial release [2]. The NCHS LMFs have recently been updated with mortality data collected through December 31, 2019. A partially synthetic public-use version of the LMFs can be downloaded from the NCHS website[3], and a restricted-use version is available through the NCHS Research Data Center (RDC)[4,5]. The public-use LMFs include data perturbation to reduce disclosure risk and limit the amount of detail available for specific mortality variables (e.g., cause of death information is summarized as the leading causes of death at the national level). Despite these identified data quality limitations, the public-use files remain highly utilized, indicating that the limited data included on these files continues to meet data user needs. Users who require more detailed linked mortality data continue to have the option to access these data through the NCHS RDC.

How the FCSM Framework for Data Quality (DQ) compares with previous efforts to assess DQ on this dataset: The NCHS Data Linkage Program publishes an accompanying Linkage Methodology and Analytic Considerations report with every linked data file release. The purpose of these reports is to offer users

[1] FCSM-20-04 A Framework for Data Quality (ed.gov)
[2] Linked Mortality Data Files Citation List (cdc.gov)
[3] NCHS Data Linkage - Mortality Data - Public-Use Files (cdc.gov)
[4] NCHS Data Linkage - Mortality Data - Restricted-Use Data (cdc.gov)
[5] RDC - Research Data Center Homepage (cdc.gov)

transparency in how the linked data were created and to describe analytic considerations for users prior to accessing the data. The NCHS was already incorporating many of the components of the FCSM Framework for Data Quality in its published linkage methods and analytic considerations documentation prior to the framework's publication. The FCSM Framework for Data Quality offers opportunities to evaluate aspects of data quality attributes, including disclosure risk, in a standardized reporting structure.

Description of implementation including human capital, technology needed, and cost: The production of the most recent NCHS LMFs took approximately 9 months, including research and development of an enhanced linkage algorithm, quality control, and disclosure review. The human capital required includes NCHS programmers and analysts, the branch chief, and other NCHS staff who review the LMFs for disclosure risks, perform scientific review of the documentation, and update the center's website. The technology needed includes statistical software, the NCHS's secure computer network, and the programming code used in the enhanced linkage algorithm, which includes machine learning and a sophisticated perturbation process.

Assessment of the NCHS LMFs using the FCSM Framework for Data Quality:

Utility

Relevance – There are over 1,000 peer-reviewed scientific publications based on the NCHS LMFs, which indicates that the LMF data are highly relevant to the NCHS data user community. NCHS has approved numerous researcher proposals to utilize the NCHS LMFs, which also demonstrates the relevance of this access option for some NCHS data users.

The release of LMF microdata files, as opposed to tabulations alone, increases the relevance to users by providing the opportunity to develop and answer their own research questions. The release of microdata also eliminates the need for NCHS staff to anticipate all of the research topics users will be interested in analyzing.

Accessibility – The release of both restricted-use and public-use LMFs balances privacy concerns with the desire to release granular and accessible microdata. Although there are additional accessibility challenges in analyzing the restricted-use file within a RDC environment, the additional data granularity (i.e., specific causes of death, exact dates of birth and death) and the fact that the data are not perturbed may offset these challenges—depending on the specific needs of each research project. Because NCHS provides users with both options, researchers are able to choose the level of accessibility and resulting data quality that best meets their research needs.

Timeliness – The linkage of NCHS survey data to mortality data is one of the longest-running linkage activities at the NCHS. The linkage is periodically updated to add new survey cohorts and linked mortality information. These periodic updates take place after the new survey data are collected and processed and after the NDI final files are released. Processing steps include standardization and cleaning of identification data collected during the survey data collection period[6]. Linking survey participants to death information and producing the linkage methods and analytic considerations report

[6] Appendix I of 2019 NDI Linkage Methods and Analytic Considerations (cdc.gov)

generally takes approximately 9 months to complete. Finally, the files are evaluated by the NCHS disclosure review board prior to their dissemination, which can take an additional 2 months. The most recent restricted-use 2019 LMFs were released in July 2021, and the most recent public-use 2019 LMFs were released in May 2022. The public-use LMFs require additional processing time to complete the data perturbation process and the assessment of validity and re- identification risk of the proposed files. The 2019 LMFs included survey cohorts collected through the end of 2018 and death data collected through the end of 2019, which allows for at least 1 year of mortality follow-up for all survey participants. The NCHS continues to incorporate automated processes whenever possible to shorten the processing time and increase the timeliness of LMF products. The periodicity of updated linkages is also subject to availability of timely administrative data updates.

Punctuality – The restricted-use and public-use LMFs are generally released within 3 months of the target date. Mortality linkages are periodically updated to include the most current NCHS surveys and administrative data sources.

Granularity – Once merged with the survey files, the LMFs provide access to granular linked demographic, socioeconomic, and health information, as well as mortality status, cause of death, and survival time. The restricted-use LMFs provide the most granular data, including date of survey interview, date of death, and detailed causes of death. Researchers utilizing the NCHS RDC network can incorporate other external data sources along with the restricted-use LMFs by utilizing geocodes. The public-use LMFs provide more granularity than many premade tabulations but limits access to variables that pose a greater risk to confidentiality. Specifically, the public-use LMFs include cause of death only for adults, only the nine most common causes of death, include a 10th category for "all other" causes of death if the cause of death was not among the nine leading causes, and provide less detailed information on time to death.

Objectivity

Accuracy and Reliability – The accuracy of the recently released 2019 LMFs was assessed through a comparison with the Medical Expenditure Panel Survey (MEPS), which uses active surveillance methods to determine vital status. MEPS households are a subsample of households that participate in the National Health Interview Survey (NHIS) approximately 6 months to a year prior to participating in the MEPS. There is an overlap of the participants that were in NHIS and those that participate in MEPS. For the participants that were in both surveys, the results showed almost perfect agreement between the two data sources, which supports the accuracy of the linkage methodology used to conduct mortality follow-up for NCHS surveys. The reliability of the most recent, 2019 LMFs was also assessed through a comparison with the previous LMFs, which was released in 2015. Hazard ratios were estimated for the risk of death among age, sex, race, ethnicity, and education groups using overlapping years of survey data and mortality follow-up. The results showed general agreement in the direction of the association between each of these characteristics and mortality risk. There were also few changes in the statistical significance of the results.[7]

Because the public-use LMFs are subject to statistical disclosure limitations methods to reduce re-identification risk, NCHS also conducts and publishes analyses to assess the comparability of results

[7] Appendix II of 2019 NDI Linkage Methods and Analytic Considerations (cdc.gov)

based on the restricted-use and public-use data files.[8,9,10] These comparative analyses demonstrate analytic scenarios where the two versions of the files produce very similar results and also highlight analytic scenarios in which analysts may wish to consider using the restricted-use files instead of the public-use files (e.g., estimates from analyses using the public use files may not align with restricted use data for numerically small population subgroups).

Coherence – The 2019 LMFs maintain coherence by using many of the same definitions as the previous versions of the 2015 LMFs, so much so, that minimal coding is needed to update the analyses, even with 4 additional years of mortality data. The 2019 LMFs also demonstrate considerable agreement with national mortality estimates generated using the national vital records file, as opposed to records for survey participants only.[11] This alignment with external statistical standards bolsters the coherence of analyses performed with the 2019 LMFs.

Finally, a potential threat to coherence of the NCHS LMFs is linkage bias. NCHS conducts person-level linkages only for survey participants who are considered eligible for linkage. Because linkage eligibility for the LMFs is based on whether survey participants provide the necessary personally identifiable information (PII) to enable linkage, bias can be introduced into the resulting linked data set if the characteristics of linkage-eligible survey participants differ from the nationally representative survey population. To mitigate this potential data quality concern, NCHS publishes analyses that evaluate potential bias and provide guidance to analysts on how to adjust sample weights.[12] Other sources of bias can also affect linked-data quality including linking error. NCHS also publishes analyses designed to inform data users of the anticipated level of false positive and false negative linking errors in linked data sets.[13] However, these known sources of error may not be the only source of bias present in linked data sets.

Integrity

Scientific Integrity – The 2019 LMFs were created using a state-of-the-art linkage algorithm that takes advantage of recent advancements in record linkage techniques, including utilizing a machine learning blocking technique. The production of the 2019 LMFs was free from political influence. In fact, until the files were released, the data linkage team did not know what research questions the data will be used to

---

[8] Lochner K, Hummer RA, Bartee S, Wheatcroft G, Cox C. The public-use National Health Interview Survey linked mortality files: methods of reidentification risk avoidance and comparative analysis. Am J Epidemiol. 2008 Aug 1;168(3):336-44. doi: 10.1093/aje/kwn123

[9] Mirel LB, El Bural Félix S, Zhang C, Golden C, Cox CS. Comparative Analysis of the National Health Interview Survey Public-use and Restricted-use Linked Mortality Files. National Health Statistics Reports; no 143. Hyattsville, MD: National Center for Health Statistics. 2020.

[10] Mirel LB, Zhang C, Cox CS, Ye Y, El Burai Félix S, Golden C. Comparative Analysis of the National Health and Nutrition Examination Survey Public-use and Restricted-use Linked Mortality Files. National Health Statistics Reports; no 155. Hyattsville, MD: National Center for Health Statistics. 2021.

[11] Vital and Health Statistics, Series 2, Number 147 (October 2008) (cdc.gov)

[12] Aram J, Zhang C, Golden C, Zelaya CE, Cox CS, Ye Y, Mirel LB. Assessing linkage eligibility bias in the National Health Interview Survey. National Center for Health Statistics. Vital Health Stat 2(186). 2021. DOI: https://dx.doi.org/10.15620/cdc:100468.

[13] Appendix I of 2019 NDI Linkage Methods and Analytic Considerations (cdc.gov)

answer and whether the research will have political implications. In addition, the public-use files were reviewed by the NCHS disclosure review board.

Credibility – The high demand for access to the previous LMFs indicates that users place a high level of confidence in the NCHS to produce accurate and reliable health data.

Computer and Physical Security – A restricted-access computer network is used to protect the information used in the creation of the LMFs and the LMFs themselves. Access is granted on a "need-to-know basis" and is limited to the data linkage staff actively involved in the production of the files. A formal output review process is in place to ensure that only information that does not pose a re-identification risk for survey participants is removed from the secure network. Data linkage staff complete annual security training and take an oath to protect the confidentiality of survey participants.

Confidentiality – Processes are in place to ensure the confidentiality of LMF data. The NCHS data linkage team applies statistically valid disclosure limitation techniques to the data, resulting in a partially synthetic public-use LMF.[14,15] Before the public-use LMF is released, the NCHS disclosure review board evaluates the risk of disclosure present in the data and the data perturbation techniques applied to the data. The restricted-use file is accessible only through the RDCs, which are staffed by statisticians trained to address confidentiality concerns before summary statistics are released for public use.

Lessons learned/sustainability: The NCHS conducts extensive evaluation of linked data files in order to assure the statistical validity of its linked data resources and to assist researchers in evaluating which LMFs (restricted use vs. public use) are most appropriate for their research goals, while protecting confidentiality. Because the public-use LMFs are by far the most utilized linked data product NCHS produces, the NCHS data linkage program continues to innovate to meet the demand for publicly available linked data files. However, these efforts must continue to appropriately balance the need for granular microdata with confidentiality protections. The NCHS will continue to explore new methods for developing publicly available linked data resources, including the creation of fully synthetic data files, while continuing to provide researchers with detailed information on the linkage methodologies and analytic considerations needed when analyzing the linked NCHS data.

Case Study 2: The Consumer Price Index Program's Use of Crowdsourced Gasoline Price Data

Author: John Bieler, Bureau of Labor Statistics

Overview: This case study describes the data quality assessment of a new method for collecting Consumer Price Index (CPI) gasoline price data from retailers or data aggregators instead of a sample survey at BLS. While crowdsourcing data directly from retailers leads to efficiencies in both collection efforts and costs, the method also introduces the potential for increased errors in collection. This case study highlights how the BLS mitigates threats to the accuracy and reliability of these crowdsourced data and is using the framework to guide the expansion of alternative data into CPI estimation.

Introduction: BLS traditionally collects prices by hand for goods and services sampled in the survey underlying the CPI.[16] However, collecting data directly from retailers or data aggregators in a more automated fashion ("crowdsourcing") leads to collection efficiencies and improved cost-effectiveness by reallocating resources. Additionally, collection automation can also lead to the efficient capture of significantly more price observations, thereby improving accuracy.

A case study is the CrowdSourced Motor Fuels Data project, which led to replacement of the traditional CPI gasoline sample. The BLS refers to the crowdsourced motor fuels data as CORP5 data. The CPI program uses pseudonyms for all alternative-data providers to protect their confidentiality.[16]

Description of data being assessed for quality: The CPI program has been collecting daily motor fuel price data for regular, midgrade, and premium gasoline from CORP5 since June 2017.

Crowdsourced CORP5 data are collected from all gas stations within CPI's 75 geographic sampling areas. While CORP5 data consist of millions more observations per month than the traditionally collected CPI data, they are not considered a census of all gasoline price observations.

Between 2017 and 2019, the CPI program conducted research on the data, including index simulations. After being vetted internally and in multiple outside venues, such as the American Economic Association (AEA) and EuroStat, the dataset was approved for implementation and the CORP5 data were included in the CPI in July 2021.

Table 1 – Data Comparison of Sources for CPI for Gasoline

	Survey data	CORP5
Frequency	Monthly	Daily
Number of price observations	4,000 price quotes/month	6.1 million observations/month
Number of retail outlets	1,400 outlets/month	91,272 stations/day
Data characteristics	Price, type of service, gasoline content, octane level, payment type, special pricing, brand name, address, collected throughout the month	Daily average price, number of valid reports, station ID, ZIP code, state, posted time

[16] Bieler, John, et al. *A Nontraditional Data Approach to the CPI Gasoline Index: CPI Crowd-Sourced Motor Fuels Data Analysis Project*, https://www.aeaweb.org/conference/2020/preliminary/paper/n8b4hBsT.

How the FCSM Framework for Data Quality (DQ) compares with previous efforts to assess DQ on this dataset:

The BLS and the CPI program have long been involved in the process of evaluating alternative datasets.[17] Prior to the FCSM Framework for Data Quality, the CPI and Producer Price Index (PPI) programs collaborated on a "scorecard" for alternative data that is very similar to the DQ framework produced by FCSM. The scorecard relied on the "qualitative" and "quantitative" analyses produced by researchers to evaluate the alternative data.

Researchers used the quantitative analysis to interpret summary statistics from the alternative data and used the qualitative analysis to summarize the data in narrative form, much like the DQ framework. The qualitative analysis in the scorecard included narrative summaries on how many observations will be provided, how the data are compiled in terms of database structure and software format, how the data meet the agency's coverage requirements for the product category and geography, sufficiency of the level of characteristic details, and timeliness, security, and reliability of data delivery within a monthly production cycle. The qualitative analysis rated the alternative data in the categories analyzed and provided a recommendation on whether to move forward with the data.

Description of implementation including human capital, technology needed, and cost:

The CPI program views the CORP5 data as a success and a significant step forward in expanding the use of alternative data into CPI estimation. The cost of research and development was not insignificant. The research spanned 2 years, and the development and testing required another year. Still, the CPI program views the cost associated with CORP5implementation as an investment that is spread out over all the CPI alternative data research and development projects, given that many other projects will use a similar approach.

Utility

Relevance - CORP5 provides daily gasoline prices for thousands of gas stations across the United States. The CPI program uses that data directly to produce the CPI and average price products for gasoline and individual fuel types, including regular, midgrade, and premium fuel types.

Accessibility - CORP5 is providing the data on a voluntary basis at no cost to the CPI. The data source does not limit the CPI program's ability to release the data to users. We have now implemented the CORP5 data into the production process, meaning we are currently releasing the CPI and average price products as scheduled using the CORP5 data and associated methodology. The CPI program is still releasing the same products at the same level of granularity as in the past.

Timeliness - The daily prices are collected throughout the month. The data also include weekend and holidays observations, providing pricing data for days that were previously not reflected in the traditionally collected data.

Punctuality - The CORP5 data were implemented into the CPI's monthly production schedule, and no schedule accommodations were needed. CORP5 typically provides the data in a timely manner with relatively few hiccups. If CORP5 data collection is missed for whatever reason, the CPI program either

[17] Crystal G. Konny; Brendan Williams and David M. Friedman, (2019), Big Data in the US Consumer Price Index: Experiences and Plans, NBER Chapters, National Bureau of Economic Research, Inc.

retrieves the missing data later in the reference period or publishes the CPI without the missing data. However, the CPI program has been able to collect data on over 95% of the days since we have begun using the CORP5 data.

Granularity - We produce price indexes and average price products at the same level of granularity as we have historically, including U.S.-, regional-, and city-level products. The large sample of gas stations protects confidentiality by including price changes across the thousands of locations. Thus, data users are unable to ascertain whether a particular gas station is in the sample or not. Station IDs provided by CORP5 also further mask the identity of the stations. With so many observations from stations across the country, the CPI program is not concerned about insufficient data.

The data provided by CORP5 are more granular in terms of timeliness. The CPI program receives daily prices from CORP5. A method was developed to convert these daily prices into monthly prices. To replicate the granularity in terms of unit of time within the CPI, we calculate an arithmetic average price across the days of the month for a particular station and fuel type.

Objectivity

Accuracy and Reliability - Research results from the CORP5 compared favorably to the CPI for gasoline at the U.S. level. The CORP5 research process did not find differences greater than 1.0% at the U.S. level over 3.5 years.

In terms of reliability, the CPI program studied the minimum amount of data needed from CORP5 to publish a gasoline index. The CPI program still collects gasoline data using its traditional method as a fallback to the CORP5 data, but is also working to reduce the amount of data collected in the traditional way to lower costs. Additionally, the fallback data will serve as a baseline quality check on the CORP5 data.

The CPI program considers the accuracy and reliability of the alternative data to be the most important factor to the CPI in the entire framework. Typically, the CPI program likes to see alternative data researched for an extended period before receiving final approval.

Coherence – As mentioned in the relevance section, the gasoline prices provided by CORP5 match the definition of gasoline used in the CPI. Furthermore, the index methodology for CORP5 aligns with best practices of price index theory as described in the International CPI Manual.[18] Finally, as mentioned in the accuracy and reliability section, the results of the alternative index (using the CORP5 data) closely match the results of the traditional CPI for gasoline.

Integrity-

Scientific Integrity- The probability and impact of the malicious or unintentional interference by data providers with the data in a way that impacts the estimates is low. The CPI program perceives no incentive for a provider to manipulate the information. Data providers have an incentive for their data to be as accurate as possible since they also publish this information on an even more granular level. In general, CORP5 and the CPI's incentives align for accurate and reliable data.

[18] Consumer Price Index Manual: Theory and Practice. International Labour Office (ILO), 2004.

Since we have begun researching the data, there has been particular interest in the gasoline index.

Credibility – As stated in the accuracy and reliability section of our assessment, the CPI program compared over 3 years' worth of price indexes and found little difference between the official CPI for gasoline and the experimental index using CORP5 data. We also compared additional months as part of our acceptance testing and parallel testing process during the implementation of the CORP5 data into our published index. CORP5 is also an often-cited source in news organizations and is widely accepted by its users as a credible source of gasoline price information.

Computer and Physical Security– The CPI program collects data from CORP5 via a Secure File Transfer Protocol (SFTP). The risk to computer and physical security is considered low based on our collection of CORP5 data over time, which has been relatively consistent.

Confidentiality – As previously mentioned, the CPI program collects data from thousands of gas stations across the country, which generally mask the confidentiality of individual data providers. Furthermore, the CPI program uses pseudonyms to mask the confidentiality of the data provider.

Lessons learned/sustainability: The FCSM framework is an important tool to guide the expansion of alternative data into CPI estimation. However, the CPI program quickly recognized it is neither a "one size fits all" nor a "be all and end all" approach. Rather, BLS views the DQ framework as a guideline that needs to be adaptable to an organization's unique circumstances and concerns.

For instance, BLS developed a set of questions to add context to the 11 dimensions. For example, to evaluate the dimension of relevance, BLS adds "Are the data a relevant input to our data products and measurement objective?" For evaluation of timeliness, the CPI program asks itself, "Are the data representative of the index reference period?" The CPI program also considers an additional question about cost-effectiveness to complement the DQ framework. We ask ourselves, "Are the new data and methods more cost-effective than the data and methods they are replacing?"

Additionally, since adopting the DQ framework, the CPI program has replaced the "quantitative" and "qualitative" reports mentioned above with a single "Alternative Data Methods Summary," a living document that is updated regularly to reflect changes in the methods over time. The summary document is very similar to the qualitative analysis mentioned above. Methods related to the collection of alternative data are subject to change throughout the approval process, as analysts and stakeholders familiarize themselves with the data. Stakeholders use the summary document to assess the adherence to the DQ framework over time.

Additionally, the CPI program has established an alternative data approval process that includes two approval groups, technical experts and the approval board. The technical experts group, which consists of senior BLS economists and statisticians, are working with BLS staff to develop a new methodology to ensure that the use of alternative data does not increase the total measurement error relative to traditional methods of data collection or previously implemented non-traditional methods. These groups will help the CPI program sustain an adherence to the DQ framework by ensuring that all future alternative data projects adhere to its guidelines.

Since there is no single metric to assess total measurement error, the technical experts make an overall qualitative assessment. In general, they consider both the statistical viewpoint (i.e., can the portion of the marketplace not in the sample be considered missing at random?) and the economic viewpoint (i.e.,

is the new data source and method consistent with the scope of the CPI, and is it measuring what we intend it to measure?). They consider each area of the methodology summaries including geography, price specifics, item definition, item eligibility, item classification, sampling, sample rotation, index methodology and index formula, item substitution/quality adjustment/comparability, and imputation.

Once the technical experts approve the methodology associated with an alternative data source, the proposal is then reviewed by the approval board for final authorization. The approval board is a cross-program group of managers charged with approving methodologies for implementation into the CPI. It is the responsibility of the approval board to ensure that the alternative data source and methodology being considered align with the data quality framework as outlined by FCSM.

The approval board can either approve the methodology for implementation or send it back to the research team with comments. A consensus agreement must be reached within 2 weeks of the recommendation for the approval board to approve or disapprove a proposal for implementation. The approval board sends their approval to the technical experts group, the alt data oversight group, and the CPI management group. In all other cases, the issues preventing approval must be documented and returned to the research team to mitigate the issues.

As the CPI program expands its use of alternative data sources, the application of the FCSM framework will continue to guide our data quality assessment process. However, alternative data sources are typically unique, and, thus, we recognize that the FCSM framework may require refining and adjustment as we encounter new data scenarios.

Table 2– Questions Added by BLS When Evaluating Data Quality with A Framework for Data Quality.

Domain	**Dimension**	**Definition**	**BLS Question(s)**
Utility	Relevance	Relevance refers to whether the data product is targeted to meet current and prospective user needs.	• What is the probability of unknown sources of bias? • What is the impact if there are unknown sources of bias?
	Accessibility	Accessibility relates to the ease with which data users can obtain an agency's products and documentation in forms and formats that are understandable to data users.	• Are the costs to access the data an effective use of resources? • Will the methodology limit our ability to release data to users? • How can we describe the methodology to data users?
	Timeliness	Timeliness is the length of time between the event or phenomenon described by the data and their availability.	• Did a lack of timeliness impact how the data for the index could be used for the reference period?
	Punctuality	Punctuality is measured as the time lag between the actual release of the data and the planned target date for data release.	• Can the methodology be implemented within the typical production processing schedule? • What is the probability that the production schedule will be affected by a delay in the delivery or processing of data? • What is the impact of such a delay?
	Granularity	Granularity refers to the amount of disaggregation available for key data elements. Granularity can be expressed in units of time, level of geographic detail, or the amount of detail on any of a number of characteristics (e.g. demographic, socio-economic).	• Are there adequate data to support the current level of granularity in data products? • Are there sufficient data to adequately protect confidentiality?
Objectivity	Accuracy and reliability	Accuracy measures the closeness of an estimate from a data product to its true value.	• Are there any concerns with the technical experts'

		Reliability, a related concept, characterizes the consistency of results when the same phenomenon is measured or estimated more than once under similar conditions.	qualitative assessment of total measurement error?
	Coherence	Coherence is defined as the ability of the data product to maintain common definitions, classification, and methodological processes, to align with external statistical standards, and to maintain consistency and comparability with other relevant data.	• Does the methodology impact the ability to compare CPI data with external sources? • Is the methodology internally coherent with other CPI methodologies (not just the one it replaces)?
Integrity	Scientific integrity	Scientific integrity refers to an environment that ensures adherence to scientific standards and use of established scientific methods to produce and disseminate objective data products and one that shields these products from inappropriate political influence.	• What is the probability and impact of the data provider (either maliciously or unintentionally) interfering with the data in a way that impacts estimates?
	Credibility	Credibility characterizes the confidence that users place in data products based simply on the qualifications and past performance of the data producer.	• Based on a review of the output of index simulations and an assessment of the differences, how much does the simulation deviate from production?
	Computer and physical security	Computer and physical security of data refer to the protection of information throughout the collection, production, analysis, and development process from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.	• What is the probability of a loss of data or data quality issues due to technical issues? • What is the impact of risks of a loss of data or data quality issues due to technical issues?
	Confidentiality	Confidentiality refers to a quality or condition of information that is protected by an obligation not to disclose the information to an unauthorized party.	• Are there any confidentiality concerns related to the use or announcement of this methodology?

Case Study 3: Motor Carrier Inspection Data

Author: Ankur Saini, Department of Transportation

Overview: This case study evaluates data collected from the U.S. Department of Transportation (DOT)'s new roadside inspection tool, SafeSpect, highlighting the utility domain in the DQ framework and, more specifically, the relevance and timeliness dimensions. With respect to relevance, the tool continues to provide highly relevant data that support the Federal Motor Carrier Safety Administration's (FMCSA) mission. The case study describes in some detail how the new tool improves the timeliness of making the data available for review.

Introduction: The FMCSA was established as a separate administration within the DOT on January 1, 2000, pursuant to the Motor Carrier Safety Improvement Act of 1999.[19] The primary mission of the FMCSA is to reduce crashes, injuries, and fatalities involving large trucks and buses. As of December 2021, there were approximately 758,000 motor carriers registered with the FMCSA, which together had traveled over 300 billion vehicle miles.[20] The FMCSA conducts its safety mission by registering motor carriers and other entities and enforcing regulations, through inspections and investigations, which maintains the balance between supply chain efficiency and roadway safety.

Description of data being assessed for quality: Inspections, as well as the data collected during the inspections, form a key component of the safety strategy employed by the FMCSA to reduce crashes and fatalities attributable to large trucks and buses. An inspection is an examination of an individual commercial motor vehicle (CMV) and/or driver by an authorized safety inspector.[20] The FMCSA, in collaboration with its state partners, conducts approximately 3.5 million inspections annually across the United States.[20] The inspection determines whether the driver and/or the CMV is in compliance with the Federal Motor Carrier Safety Regulations (FMCSRs) or the Hazardous Materials Regulations (HMRs), as appropriate. Serious violations result in the issuance of vehicle or driver out-of-service (OOS) orders.[20] These violations must be corrected before the affected driver or vehicle can return to service.[20] In addition to having immediate operational impact, the inspection data also serve as an input into the Safety Measurement System (SMS) that is used to identify carriers with potential safety problems and prioritize them for interventions. Approximately 3.3 million inspections[20] of vehicles and drivers are conducted each year by inspectors that are employed by state law enforcement agencies. The inspectors use a variety of inspection software, including inspection software provided by the FMCSA, and data gathering standards that are driven by both federal and state guidelines. The FMCSA is in the process of implementing SafeSpect, the replacement for the FMCSA's current roadside inspection software, which provides next-generation capabilities for information capture and processing during all facets of the inspection process.

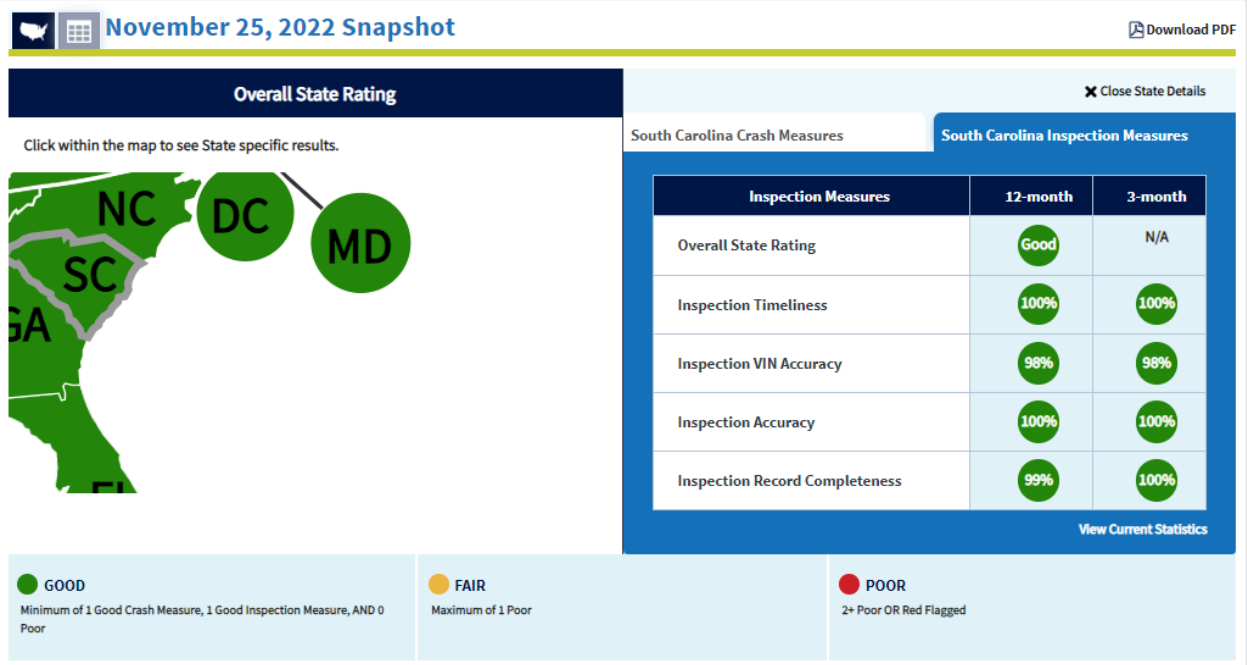
How the FCSM Framework for Data Quality (DQ) compares with previous efforts to assess DQ on this dataset: The FMCSA has implemented standards that guide the integration of state inspection systems with the FMCSA's repository of all inspection data, the Motor Carrier Management Information System (MCMIS). The methodology for State Safety Data Quality (SSDQ) was developed by the FMCSA to evaluate the completeness, timeliness, accuracy, and consistency of the state-reported commercial

[19] Federal Motor Carrier Safety Administration, https://fmcsa.dot.gov/mission
[20] Federal Motor Carrier Safety Administration, 2022 Pocket Guide to Large Truck and Bus Statistics

motor vehicle crash and inspection records in MCMIS.[21] The quality of these data are evaluated with each monthly snapshot, as shown in Figure 2, and the states receive ratings of Good, Fair, or Poor for nine SSDQ measures.[21]

In addition, the FMCSA has implemented DataQs, an FMCSA system that allows users to request and track a review of federal and state data issued by the FMCSA that the users believe to be incomplete or incorrect.[22] DataQs enables all users—motor carriers, drivers and their representatives, as well as the FMCSA and its state partners—to improve the accuracy of the FMCSA's data-driven safety systems.[21] Specifically, the DataQs system allows the user to submit a request for data review (RDR), provide supporting documentation, and track their request.

Figure 2. State Rating for South Carolina for Inspection Measures



Assessment of motor carrier inspection data using the FCSM Framework for Data Quality

Utility

Relevance – The data captured in the motor carrier inspection dataset are highly relevant to the safety mission of the FMCSA. The FMCSA, in collaboration with its state partners, conducts approximately 3.5 million inspections annually across the United States. The data collected during the inspections form a key component of the safety strategy employed by the FMCSA to reduce crashes and fatalities attributable to large trucks and buses.

Accessibility – The inspection records, once available in FMCSA data repositories, are easily accessible for reference to its key stakeholders: the inspection community and the motor carriers. However, there is unrealized potential in terms of leveraging the inspection data to run real-time and/or predictive

[21] Federal Motor Carrier Safety Administration, Data Quality, https://ai.fmcsa.dot.gov/DataQuality/
[22] Federal Motor Carrier Safety Administration, DataQs, https://dataqs.fmcsa.dot.gov/

analytics that guide outcomes for the FMCSA's safety mission. The FMCSA is in the process of modernizing MCMIS, the authoritative source of inspection records, to streamline the data exchange between the transactional data repository by establishing an authoritative data warehouse and other context-specific data marts that will allow for better reporting and outcome-oriented analytics.

Timeliness – An inspection record is considered timely if it is reported within 21 days of the inspection.[22] The FMCSA evaluates 12 months of data, excluding the most recent 3 months, to determine a state's rating on timeliness. The most recent 3 months of record are used to calculate a leading indicator that forecasts the timeliness rating going forward.[22] While the timeliness rating provides a snapshot of data quality based on the defined parameters, it presents programmatic challenges that present a risk to the safety mission. The lag between an inspection and the reporting of the record for review by the inspection community may allow unsafe motor carriers to operate and pose a hazard to other entities sharing the road. In addition, the lag may also contribute to inconsistencies in information presented in the record, which may (1) be detrimental to the probability of a successful enforcement action against the motor carrier, (2) have a negative impact on the ability of the motor carrier to operate in accordance with the regulations, and/or (3) result in an RDR. SafeSpect, the FMCSA's new inspection platform, provides significant efficiencies to improve timeliness of inspection records. SafeSpect presents a departure from the traditional workflow of an inspection, which requires inspection data to traverse a lengthy approval process before being deemed available for reference by the inspector community. Inspections, and resulting violations, conducted using SafeSpect are now immediately available for review by the entire inspection community—thereby empowering the safety organizations with timely information and implementing a deterrent for the bad actors that pose a safety hazard.

The inspection data undergo a data capture and review process that impacts the timeliness of the data at many stages. Most of the inspections entail motor carrier, vehicle, and driver data to be manually entered into the inspection software by the inspector. Depending on the software being used and the work habits of the inspector, the data may be entered during the inspection process, written down on a piece of paper and inputted in the system at the end of the inspection, or written on a paper inspection form that is later entered into the inspection software at the end of the day. Once the data are inputted, the inspection report is submitted for review. The inspection report becomes a record once approved. The approval is a manual process that includes the downloading and uploading of files by a designated authority. Given the labor-intensive nature of the process, it contributes to a lag between the date of an inspection and the date when the record becomes available. The FMCSA strives to improve the timeliness of data through the capabilities offered by SafeSpect. First, SafeSpect is the only product in the market that is device-agnostic. All it needs is a device with a Web browser to work. Therefore, one can work seamlessly using a phone, a desktop computer, or a laptop in multi-device environments, which facilitates inspections that are more efficient than those done using previous tools. In addition, SafeSpect allows use of multiple devices for conducting a single inspection. Similar to contemporary applications like Netflix, the SafeSpect application allows the inspector to start the inspection on one device and complete it on another. Such capability allows for electronic capture of information in areas, such as under the truck, that were inaccessible to desktops and laptops previously. Further, since the application was designed in close collaboration with the users, the application's workflow mimics the actual work. The users have the option of being guided by the software, if they are new to the inspection process, or guiding the software, if they are seasoned inspectors and have developed an implicit way of conducting an inspection over years. Such features are likely to encourage the users to

adopt electronic methods to capture inspection data as opposed to using pen and paper. SafeSpect also presents a unified interface to the process of conducting and approving an inspection, which moves the safety community away from the four-phase process of conducting the inspection, uploading it into a central repository, downloading it for approvals, and then re-uploading it for final archiving. An inspection performed using SafeSpect is immediately available for approval without any additional steps—thereby, improving timeliness and presenting an opportunity to institutionalize other metrics that are better indicators of timeliness.

Punctuality – Punctuality was not addressed in this case study.

Granularity – The inspection records provide an adequate level of detail in accordance with the North American Standard Inspection Program, which was developed by the Commercial Vehicle Safety Alliance to improve the safe operation of commercial motor vehicles by establishing a uniform and reciprocal roadside inspection and enforcement process for commercial motor vehicles. The program outlines minimum inspection procedures, standards, and requirements; ensures consistency in compliance, inspections, and enforcement; and minimizes duplication of efforts and unnecessary operating delays for the motor carrier industry.[23] The program describes eight levels of inspection (I-VIII), each of which defines a different data element for capture, e.g., driver credentials, record of duty status, and mechanical condition of vehicle.

Objectivity

Accuracy and Reliability– The accuracy of the inspection record is rated on two criteria: (1) inspection record matching a carrier registered with the FMCSA in the MCMIS; and (2) Vehicle Identification Number (VIN) meeting a validation criterion.[21] While the two criteria provide a rudimentary level of quality control, they do not ensure holistic quality across the inspection record. There are many scenarios where an inspection may pass the two accuracy criteria but is not accurate in reality, for example, because a driving license was captured incorrectly or violations were noted incorrectly. This is particularly true when the inspection details are captured manually by an inspector and processed through the inspection software later. SafeSpect provides a one-stop shop to the inspection community for all their inspection needs, which promotes accuracy by allowing the inspectors to log in once and conduct the entire inspection, from cradle to grave, without ever logging into another application. All the information required for inspections is at the inspector's fingertips and is furnished automatically without any intervention from the inspector. The human-centered design of SafeSpect will allow inspectors to tailor the experience of the application to their usage, carry an input device to the point of inspection to eliminate transcription errors, and easily input information with minimal keystrokes to minimize user input errors and inconsistencies.

Coherence – The North American Standard Inspection Program provides a good level of guidance to ensure coherence in inspection records. The primary risk to the coherence of the data set is the varying level of validations and checks built into the inspection software to enforce logical consistency. The FMCSA collaborated with its state partners and the CVSA to ensure that SafeSpect has validations and business rules in place that minimize the capture of information that is either not relevant to the

[23] Understanding the North American Standard Inspection Program, https://www.cvsa.org/wp-content/uploads/NASI-Program-Brochure.pdf

inspection or not permissible under rules. The application and automation of such data-driven decision-making paves the path for improved coherence during the data gathering process.

Integrity

Scientific Integrity– All inspections are conducted in accordance with minimum inspection procedures, standards, and requirements described in the North American Standard Inspection Program. The data captured during the inspection process are purely based on observations made by the inspector and reflect application of regulations. In addition, the inspection record is evaluated for completeness, i.e., certain elements of driver information and vehicle information are compared against validation criteria.

Credibility – All inspections are conducted by inspectors who are federal or state law enforcement officers. They undergo rigorous training and are certified to conduct inspections at one or more levels specified in the North American Standard Inspection Program. Therefore, there is a high level of confidence in the ability of the personnel engaged in data capture and processing. However, the current process to capture and validate all the information required for an inspection is convoluted and presents many areas and steps that are prone to errors, which leave an inspection record vulnerable to challenges targeted at the credibility of the record. Implementation of validations, application of business logic, and provision of a unified interface to input and process inspections through SafeSpect will contribute to enhancing credibility of the dataset.

Computer and Physical Security – All inspection records are housed in cloud-based databases that serve applications that implement role-based access controls to furnish the inspection records to users on a need-to-know basis. The FMCSA has taken considerable care to ensure that the data are protected from unauthorized access. All access requests are reviewed and approved by the regional program staff and state personnel, who validate that the requesting party has a legitimate need for the request submitted. However, the current authentication mechanism does not provide two-factor authentication. SafeSpect addresses the security concerns presented by the current solution. It allows the inspectors to log in once, through a multi-factor authentication mechanism, and conduct the entire inspection, from cradle to grave, without ever needing to log in to another application. Further, as a Web-based application, SafeSpect allows the FMCSA to be a lot more responsive and agile in addressing cybersecurity risks and threats than it would be for a client-server application.

Confidentiality – Access controls are in place to ensure that (1) motor carriers have access to only their inspection records, and (2) enforcement personnel, including federal and state inspection personnel, have access to all inspection records to review.

Lessons learned/sustainability: Historically, the FMCSA's approach to implementing quality control for the inspection data has been accommodating. Since 95% of the inspection data are sourced from various states that historically have used a variety of inspection software, the accommodating approach has supported data collection in the absence of a common platform such as SafeSpect. With the use of SafeSpect, the FMCSA is transforming its approach to data collection and data quality. The modernization of SafeSpect and MCMIS are a step in the direction of ensuring that (1) the right data are being captured; (2) there is no redundant data gathering; (3) all data elements align with organization and regulatory policies; and (4) the technology platform is reliable, available, and maintainable.

Case Study 4: Physical Activity Monitor (PAM) Data from the NHANES

Authors: Lisa B. Mirel, National Center for Science and Engineering Statistics, Lara Akinbami, National Center for Health Statistics (NCHS), and Alan Simon, NCHS

Overview: This case study assesses the utility of physical activity monitor (PAM) data collected in the National Health and Nutrition Examination Survey (NHANES), which is conducted by the National Center for Health Statistics (NCHS) within the Centers for Disease Control and Prevention (CDC). This study focuses on evaluating the accessibility of large data files that require subject matter expertise to analyze, in addition to other elements of data quality. .

Introduction: Federal data are used to inform policy, evaluate programs, drive business and economic decisions, and improve health. Public release of new data sources must include transparent assessments of data quality to allow users to understand the strengths and limitations of the data as well as the appropriateness of intended use. The FCSM recently published A Framework for Data Quality to support federal agency efforts to report on aspects of data quality for all federal data collections across three main domains: utility, objectivity, and integrity.[24] This case study demonstrates how the FCSM framework can be used effectively to assess the accessibility of an NCHS data source that requires subject matter expertise for analysis.

Description of data being assessed for quality: The NHANES is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and health examinations. The NHANES is administered by the NCHS, the agency within the CDC responsible for producing national vital and health statistics. The NHANES program began in the early 1960s and has been conducted as a series of surveys focusing on different population groups or health topics. In 1999, the survey became a continuous program with core content focusing on chronic conditions and nutrition measurements while maintaining flexibility to periodically add content to meet emerging needs. The survey examines a nationally representative sample of about 5,000 persons each year from 15 counties across the country.

The PAM component was introduced to provide information about activity and exercise habits among the United States population. NHANES participants are asked to wear accelerometers for 7 consecutive days to collect objective information on 24-hour movement when awake and asleep. Self-reported interview data for physical activity are potentially biased because respondents' perceptions of activity intensity vary and recall of activity duration may be inaccurate. This may be especially true for proxy-reported data collected for children and adolescents. For example, children may spend large amounts of time away from home and engage in sporadic periods of activity that are difficult for a proxy respondent to recall and quantify. This case study focuses on the PAM component that was part of the NHANES 2011–2014 cycles.

Objective measurement of physical activity with accelerometers was first implemented from 2003 to 2006. The PAM used in NHANES 2011–2014 measured acceleration (i.e., on the x-, y-, and z-axes) every 1/80th of a second (80 Hz). The device also tracked movement by measuring ambient light levels every second (1 Hz). The 80 Hz accelerometer measurements and the 1 Hz ambient light measurements were

[24] fcsm.gov/assets/files/docs/FCSM.20.04_A_Framework_for_Data_Quality.pdf

summarized over each minute, hour, and day for each participant (https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/PAXDAY_G.htm). [25] Four datasets, including a header file and corresponding summary datasets at the minute, hour, and day level, were produced for public data file release.

How the FCSM Framework for Data Quality (DQ) compares with previous efforts to assess DQ on this dataset: The NHANES publishes an accompanying quality assurance and quality control section in its data file release documentation. For the acceleration data, the documentation specifies how data were assessed and how files were created for public release. In addition to the other dimensions, accessibility is addressed in the context of the size of these files. The NHANES already incorporated many of the components of the FCSM Framework for Data Quality in its published documentation. The FCSM Framework for Data Quality offers opportunities to use a standardized reporting structure.

Description of implementation including human capital, technology needed, and cost: The data production process for the most recent PAM data spanned 8 years, including research and development, quality control, obtaining approvals for file storage and public release plans, file preparation, and disclosure review.

The human capital required included expert consultants and NCHS programmers, analysts, supervisors, and staff who reviewed the files for disclosure risks, performed scientific review of the documentation, and updated the center's website. Required technology included statistical software, the NCHS's secure computer network, and the programming code to process the large files. The files are publicly available on the NHANES website (PAXLUX_G (cdc.gov) and PAXHD_H (cdc.gov)).

Assessment of NCHS PAM data using the FCSM Framework for Data Quality

Utility

Relevance – PAM provides an objective measurement of physical activity, a topic of great public health importance given the relationship between physical activity and chronic disease development and adverse outcomes. Self-reported interview data for physical activity, while easier to obtain, have limitations including: (1) varying respondent perceptions of activity intensity, and (2) recall difficulties for periods and duration of physical activity. NHANES interview data for children and adolescents younger than 16 years of age are provided by a proxy respondent. Children may spend large amounts of time away from home and engage in sporadic periods of activity that may be unknown to a proxy respondent or difficult for them to accurately quantify.

Accessibility – The PAM data are available as public-use files. However, there could be accessibility issues due to their unusually large size. The summary datasets include minute (one record per minute

[25] ActiGraph. GT3X+ and wGT3X+ Device Manual, 2013.

http://s3.amazonaws.com/actigraphcorp.com/wp-content/uploads/2018/02/22094126/GT3X-wGT3X-Device-Manual-110315.pdf

for each participant), hour, and day measurements. Acceleration measurements obtained on the x-, y-, or z-axes are summarized at the minute level. However, measures included in the summary datasets are standardized and based on publicly available methods. These summary measures in the minute summary file (PAXMIN_H) are specified in monitor-independent movement summary (MIMS) units, which is a nonproprietary, open-source, and device-independent universal summary metric developed by researchers at Northeastern University[26] .

Timeliness –Data releases can be impacted by several factors beyond data processing and quality assessment. For PAM, major issues included accommodating the large file size (negotiating storage agreements and providing manageable summary files for public release) and assessing meaningful intervals of data to release. Before release, NHANES files are evaluated by the NCHS disclosure review board, which adds additional time to prerelease data processing and evaluation. The 2013–2014 NHANES PAM data were released in November 2020.

Punctuality – The 2011–2014 NHANES PAM data were unprecedented in terms of file size. Since the PAM data were unprecedented, no official release date was ever announced and therefore we can't calculate punctuality as defined in the framework. The summary files were released in stages as quality control and file production continued. The initial summary files were released in November 2020. The most recent release of summary files, in October 2022, was the PAM ambient light raw data.

Granularity –The PAM data have minute, hour, and day measurements. PAM data can be merged with other NHANES data to provide demographic, socioeconomic, and health information. In addition, because NHANES data are geocoded, the PAM data can also be merged with other data sources such as weather or air quality.

Objectivity

Accuracy and Reliability – Not every eligible NHANES participant has a PAM data file. For example, data may be missing if a participant did not wear a PAM or did not return it, or the data could not be retrieved from a damaged PAM. Each participant may have up to 9 days of summary records. Per the protocol, the first and last day of data collection for each participant are partial days. For most participants, a complete data collection constitutes 193 hours. In some cases, e.g., due to battery depletion, the data collection period is shorter. A small proportion of participant data files include 194 hours. According to the 2013-2014 NHANES PAM documentation, overall, 96% of participants with data wear the PAM until the 9th day. About 2% of the participants with PAM data wear it for less than 7 days.[27] For those with complete data, the measurement accuracy/reliability is very high.

Coherence – The 2013–2014 NHANES PAM maintained coherence by using many of the same definitions as the previous version (2011–2012) of NHANES PAM data. As a result, minimal coding was needed to update the 2011–2012 analyses by adding the 2013–2014 PAM data. In addition, the PAM data uses similar standard and vocabulary as other types of physical activity monitoring data sources outside of

[26] John D, Tang Q, Albinali F, Intille S. An Open-Source Monitor-Independent Movement Summary for Accelerometer Data Processing, Journal for the Measurement of Physical Behavior. 2019:2(4),268–281.

[27] NHANES PAM Documentation: PAXHD_H (cdc.gov)

NHANES. Every effort is made to use universal definitions. However, because data collection and processing occur over a long period, maintaining up-to-date definitions that evolved over time is not possible.

Integrity

Scientific Integrity –The data files that are released ensure adherence to scientific standards and methods that are free from outside influence, including quality control (QC) flags that can be used for evaluation. For example, because data quality review and the MIMS-unit calculation are conducted independently, users may wish to assess MIMS units for all minutes for which QC flags are present to assess if these calculations should be excluded from their analyses.

In addition, aligning with scientific methods, an open-source, state of the art algorithm which integrates machine learning techniques is used with the 80 Hz accelerometer data to predict time periods of wake wear, sleep wear, or nonwear data and assign a confidence value ranging from 0.0 to 1.0 to indicate the algorithm's confidence for time period coding. In most cases, the algorithm identifies a category for each minute, including "unknown" below a specified level of uncertainty. This algorithm uses three steps which are described in detail in the documentation[27,28]

Credibility – The header file and the minute/hour/day summary files are reviewed for outliers and implausible values. The criteria used for implausible data are based on published literature and expert judgment.

Computer and Physical Security– The data are processed in a secure environment with limited staff having access to identifiers in accordance with NCHS standards.

Confidentiality – Processes are in place to ensure the confidentiality of NHANES participants. Before the public-use PAM data files are released, the NCHS disclosure review board evaluates the risk of disclosure presented by the data.

Lessons learned/sustainability: The PAM data will continue to be collected, and users should note that referring to the documentation released with the public-use data file is a best practice to understanding the data quality. Through the process of PAM data processing and dissemination, NHANES learned best practices for program planning to collect and release complex or unusual data as well as communicating with users on how to analyze the data.

Users are urged to refer to the NHANES Analytic Guidelines and the online NHANES Tutorial for further details on the use of the NHANES sample weights and other analytic issues. This case study summarized the dimensions of data quality for these data in a standardized way.

[28] John D, Tang Q, Albinali F, Intille S. An Open-Source Monitor-Independent Movement Summary for Accelerometer Data Processing, Journal for the Measurement of Physical Behavior. 2019:2(4),268–281.

Case Study 5: BLS Import/Export Price Indexes Measured Utilizing Census Administrative Trade Data

Authors: Susan Fleck, Bureau of Labor Statistics (BLS), and Steve Paben, BLS

Overview: This case study examines the creation of the import and export price indexes at BLS. This case study discusses the need for objectivity, specifically, coherence--when creating an index and how bias can be a threat. The authors offer potential strategies to assess this bias.

Introduction: Data provided by the federal government to the public have an important and unique place in society, and maintaining trust and integrity in federal data is pivotal to a democratic process. Historically, federal statistical programs used data collected through surveys to estimate or model the characteristics of the study population. With the evolution of improved computing methods, and an across-the-board decline in response rates in voluntary surveys, federal statistical programs are seeking to expand the sources of data used to estimate or model the datasets prepared for public use. In previous periods, the primary source of data has been direct collection of information from individuals, households, and businesses using surveys that are representative of the target population. In the current period, more data sources are available. Numerous public and private sources of data are collected, processed, and/or curated as a secondary product of regulatory or business activity, and some of these alternative data sources are sufficiently representative of the study population. When public federal statistical programs must provide transparent assessments of the quality of data outputs, the data quality of the inputs must also be addressed to allow users to understand the strengths and limitations of the data outputs as well as whether the intended use of the inputs is appropriate. The FCSM published A Framework for Data Quality to support federal agency efforts to report on aspects of data quality for all federal statistical programs across three main domains: utility, objectivity, and integrity.[29] This case study demonstrates how the FCSM framework can be used to effectively assess the quality of BLS Import and Export Price Indexes (MXPI) calculated from Census administrative trade data.

Description of data being assessed for quality: The BLS currently calculates MXPI by surveying business establishments to collect price data for merchandise trade. To expand the quantity and quality of MXPI data, the BLS has researched the potential to use unit value indexes (UVIs) from administrative trade transaction data maintained by the Census Bureau in place of directly collected data. The data source is the official administrative trade dataset, which is collected by the U.S. Customs and Border Protection agency and later cleaned and edited by the U.S. Census Bureau for statistical purposes. This dataset comprises detailed shipment records for nearly all imports and exports of goods by the internationally consistent Harmonized System product classification for the United States. Currently, the Census dataset is shared with MXPI program after the dataset is processed to use as the sample frame for the MXPI business survey. The MXPI program plans to use the dataset on a month-to-month basis to calculate changes using statistical methods that identify unique items within the shipping transaction data, while overcoming unit value bias concerns that are known to occur with such data.

How the FCSM Framework for Data Quality (DQ) compares with previous efforts to assess DQ on this dataset: The BLS MXPI program published several research papers and formal publications which

[29] FCSM-20-04 A Framework for Data Quality (ed.gov)

included comparisons to currently calculated official price indexes.[30,31,32] The MXPI program also sought and incorporated feedback on the data concepts, source, and methods applied to incorporate the administrative trade dataset into the official price indexes. The venues for this feedback included the National Bureau of Economic Research (NBER) Conference on Research on Income and Wealth (CRIW), the Federal Economic Statistics Advisory Committee (FESAC), and the formal review process of the BLS technical advisory committee and the data users advisory committee. Additionally, the BLS published historical export price indexes for 2012–2018 in September 2021 and historical import price indexes for 2012–2018 in September 2021. Updates to the historical price indexes for 2019–2021 will be published in early 2023. The purpose of these reports is to offer users transparency and to describe analytic considerations for users, prior to switching to using UVIs that do not show any evidence of introducing unit value bias into production. BLS was already incorporating many of the components of the FCSM Framework for Data Quality in its published research papers and documentation. The framework offers opportunities to incorporate additional aspects of data quality attributes using a standardized reporting structure.

Description of implementation including human capital, technology needed, and cost: The production of the most recent historical import price indexes from 2012–2018 took approximately 9 months, including research and development, outlier treatment, and quality control. The human capital required included BLS data scientists, economists, mathematical statisticians, and the IT specialists who update the BLS website. Technology needed included statistical software, the BLS secure computer network, and the programming code used to calculate price indexes, which includes a sophisticated outlier detection and trimming process.

Assessment of BLS MXPIs calculated from Census administrative trade data using the FCSM Framework for Data Quality

Utility

Relevance – The relevance of the official MXPI measures is expected to surpass the current and prospective user needs. Currently, the major statistical agencies that use MXPI to deflate trade measures must use either PPIs or less detailed MXPIs because the representativeness and risks of respondent identifiable information, as well as other quality concerns, limits the ability to publish a number of detailed MXPIs. The use of administrative trade data will greatly expand the number of MXPIs that will be available to the public.

[30] Fast, D. and S.E. Fleck. 2019. "Measuring Export Price Movements with Administrative Trade Data." Working Paper 518. U.S. Bureau of Labor Statistics. June. Available at: https://www.bls.gov/osmr/research-papers/2019/pdf/ec190080.pdf.

[31] Fast, D., S.E. Fleck and D. Smith. 2022. "Unit Value Indexes for Exports – New Developments Using Administrative Trade Data." *Journal of Official Statistics*. Volume 38. Issue 1. March, pp. 83-106. DOI: https://doi.org/10.2478/jos-2022-0005.

[32] Fast, D. and S.E. Fleck. 2022. "Unit Values for Import and Export Price Indexes – A Proof of Concept," in *Big Data for 21st Century Statistics*. University of Chicago Press, edited by K.G. Abraham, R.S. Jarmin, B. Moyer & M.D. Shapiro. Available at: https://www.bls.gov/mxp/data/unit-values-import-export-price-indexes.pdf.

Accessibility –There is no change in accessibility of the official MXPI data products planned or expected. The main access points for the public will continue to be the BLS website and MXPI News Release.

Timeliness – The Census administrative trade data are made available to BLS to create price indexes immediately following the end of the month for more than 60% of all trade occurring during the month; in the subsequent revision month, 100% of all trade transactions are reported; a small portion of exports to Canada is finalized in the second revision month. Additionally, the monthly administrative trade data include transactions that actually occurred throughout the month. This receipt of the Census data is, on average, more timely than is receipt of directly collected MXPI data. For directly collected data, the reference period for each month is the period between the first of the month and the date that a price is first reported, but not all respondents provide prices in a timely fashion; the overall response rate measures the share of items sampled for which a price is received, and this response rate is currently 50%. The date of publication is expected to be comparable to the current schedule of publication, thus assuring the timeliness of publication.

Punctuality – Currently, the MXPI program publishes price indexes within about two to three weeks after the end of the month. In 2023, this could range anywhere between the 12th and 17th of the following month. MXPIs calculated from directly collected data have a reference pricing date of the first day of the prior month, and data are processed and indexes are run beginning in the second week of the reference month. Administrative trade data have a reference pricing date of the full month, and revised data are accepted for up to three months, with the fourth month of data being final. For the most current month of publication, the preliminary Census administrative trade data are not available until approximately one week after the end of the month. BLS will adhere as closely as possible to the current production schedule for the integration of UVIs. However, a few more business days are likely to be needed to calculate indexes from Census administrative trade data for the most current month, due to the timing of the receipt of data as well as the time required to process the large volume of data. Once the planned target date for release is established, it is not expected that there will be any problems with punctuality.

Granularity – The use of Census administrative trade data will provide BLS with the opportunity to release official MXPIs to the public at a more granular level than current sampled data, covering more detailed product and industry price trends. The homogeneous product areas account for approximately half of all MXPI product and industry categories, and the number of prices supporting the quality of the published indexes will increase significantly. Only heterogeneous product areas will be directly collected, and thus with no changes planned in resources, the number of prices and items representing heterogeneous product areas is expected to increase. The number of additional price observations greatly increases the quality of price index estimates and gives BLS the opportunity to provide more detailed index estimates. Currently BLS publishes approximately 700 MXPIs for product and industry groups, including BEA End Use, Harmonized System, and North American Industry Classification System; using the alternative data source, there will be approximately 500 more indexes that can be calculated, for a total of 1,200.

Objectivity

Accuracy and Reliability– For Census administrative trade data, the BLS will be constructing UVIs. The first step in creating accurate and reliable indexes is to construct items that can be tracked month-to-

month and are detailed enough to approximate the matched item model approach used by directly collected MXPI data and to mitigate unit value bias. Unit value bias is a mismeasurement of price trends reflecting changes in product mix within a unit instead of price changes for the products in that unit. The administrative data contain many characteristics, such as a Harmonized System code (10-digit classification number, quantity designation, country code, etc.). The MXPI program used match-adjusted R-squared (MARS) methodology to select which characteristics should be used to define an item for a UVI. This methodology described by Chessa (2021) features a trade-off between mitigating unit value bias by using more characteristics to define an item and selection bias.[33] Determining which data characteristics influence the price at which goods are traded is important so that these characteristics can be included in the item definition. Then, it can be determined whether the information available is sufficient to create items that are homogeneous enough to replicate under MXPIs current methodology.

Next, unit values for transactions are combined to form item unit values using a weighted arithmetic mean. Then, the price relatives of those items are aggregated to a classification group using a Tornquist index formula with a base strategy. Finally, the classification group indexes are aggregated to create price indexes using a modified Laspeyres formula, which allows this step of aggregation to remain consistent with current MXPI methodology.[34]

The official MXPIs have consistently been deemed accurate and reliable, and thus the primary standard to evaluate the accuracy and reliability of replacing directly collected data with the UVIs was a comparison of the index levels and month-to-month changes for the relevant product-specific MXPI. Several different statistics, such as ranking homogeneity of product categories by the coefficient of variation and comparing monthly and yearly trends (using root-mean-square error, t-tests, and regressions) were used to compare the fit between the two indexes over a historic 10-year period (2012–2021). A decision tree was used to classify the UVIs as either high quality, marginal quality, or poor quality. Those deemed as high quality showed consistent results for the same target population over the 10-year period; the impact of the change to the trends in the top-level MXPIs when replacing official price indexes with the high quality UVIs was minimal. The subset of data that underpin these high-quality UVIs for homogeneous product categories are set to replace the directly collected survey data currently used in the official MXPIs. For some marginal-quality UVIs, the current official MXPI may be of poor quality due to poor sample representativity, in which case the MXPI would be more reliable and accurate using the administrative trade data. Of all detailed BEA End Use Indexes calculated with UVIs, approximately 40% were of high quality, 13% were of marginal quality, and 47% were of poor quality.

Furthermore, research comparing the prices collected in the survey and the unit values calculated with the Census trade data showed consistent matches between items and prices at detailed product levels. This analysis validates the approach that granular descriptions of product varieties can result in a unit value that accurately reflects the price data collected with the survey. The study linked import and

[33] Chessa, A. 2019. “MARS: A Method for Defining Products and Linking Barcodes of Item Relaunches.” Presented at Ottawa Group International Working Group on Price Indices. Sixteenth Meeting - Rio de Janeiro, Brazil 08-10 May 2019. Available at: http://www.ottawagroup.org. (accessed April 2021).

[34] Smith, D. and Census Data Team. 2022. “Measuring Price Movements with Census Administrative Trade Data,” prepared for the BLS Technical Advisory Committee. June 17, 2022.

export transactions in the official MXPI survey data with the administrative trade data to assess how close unit values are to prices. There was a 20% to 25% match between product and Employer Identification Number (EIN) for each month of the study, and the percent difference in price between the matched records by product area was evaluated. The results showed that the average prices of the matched items were similar, although price levels tended to be higher for the survey data compared to the administrative trade data. This unpublished research provides solid evidence that, for homogeneous products, the survey data based on a representative sample approximate the administrative trade data, which is a census, and the specification to define a unique item for the administrative trade data approaches the matched item model for the survey data.

Coherence – The UVIs constructed from Census administrative trade data maintain coherence by using many of the same terms and definitions as the current MXPIs. It is constructed using the same source as the sampled data, the same classification schemes, and standard definitions for trade data. A potential threat to the coherence of the UVIs is unit value bias. However, the MXPI program has already evaluated the indexes from 2012–2021, which includes the COVID-19 global pandemic and recent periods of higher inflation. If the UVIs remained robust to unit value bias during this timeframe, it should remain so. However, the MXPI program will need to continually monitor UVIs with external sources where feasible.

Integrity

Scientific Integrity– The 5-year research project to evaluate the quality and suitability of the administrative trade data in lieu of survey data for homogeneous product areas has adhered to scientific standards and the use of established scientific methods to produce accurate and reliable MXPIs. In addition, the administrative data source is an improvement over survey data in two ways. First, the administrative trade data are a census of all U.S. trade and thus superior to a sample, for that subset of the target population that these data are deemed reliable to measure—homogeneous products. Second, the methodology to calculate UVIs is based on state-of-the-art price index methodology that takes advantage of recent advancements in creating price indexes from scanner data and applies those techniques here to administrative data. The methodology developed for incorporating UVIs into the official MXPIs has been disseminated to stakeholders and experts in the field of price index theory at conferences and official committees, subject to anonymous scrutiny and accepted in a peer-reviewed journal, and evaluated by eminent economists for inclusion as a chapter in an academic book.[35,36,37] Feedback from these experts and at these venues has been incorporated into the finalization of a methodology for calculating price change that improves upon the current methodology, by using current data and a superlative index formula to account for new and disappearing goods and substitutions. The production of the UVIs and decisions on which MXPIs to replace were free from political influence.

Credibility – The most visible users of MXPIs are the Census Bureau and the BEA, for which MXPIs are used to deflate gross domestic product (GDP). Each statistical partner has been highly engaged in this

[35] Fast, D., Fleck, S.E., and D. A. Smith. 2022. "Unit Value Indexes for Exports – New Developments Using Administrative Trade Data." Journal of Official Statistics. 38(1): 83-106. Doi: 10.2478/jos-2022-0005.
[36] Fast, D. and S.E. Fleck. 2022. "Unit Values for Import and Export Price Indexes – A Proof of Concept." in Big Data for 21st Century Statistics. University of Chicago Press, edited by K.G. Abraham, R.S. Jarmin, B.
[37] Moyer & M.D. Shapiro. Available at: https://www.nber.org/booksand-chapters/big-data-twenty-first-century-economic-statistics

administrative trade data project and is eager to have more reliable index estimates and more detailed MXPIs. The fact that BLS has been highly engaged with its statistical partners and has incorporated their feedback to evaluate the methodology and results reaffirms the credibility of the project and the final product.

Computer and Physical Security – A restricted-access computer network is used to protect the security of information used in the creation of the UVIs and the UVIs themselves. Access is granted on a "need-to-know" basis and is limited to BLS staff actively involved in the production of the indexes. A formal output review process is in place to ensure that only information that does not pose an identification risk is removed from the secure network. BLS staff complete annual security and privacy training and sign a nondisclosure agreement with Census to have access to and to protect the integrity of the administrative trade data.

Confidentiality – Processes are in place to ensure the confidentiality of administrative trade data. Before the UVIs are publicly released, the BLS evaluates the risk of disclosure present in the data. The administrative trade data reside on restricted-use files only accessible to staff on a "need-to-know" basis.

Lessons learned/sustainability: BLS has greatly improved the methodology for calculating price indexes directly from administrative trade data since research began in earnest in 2018. It has calculated UVIs for all BEA five-digit indexes from the period of 2012–2022, selected the high and marginal quality to continue testing in a parallel test environment to simulate official price indexes in July 2023. The target goal for moving the UVIs deemed worthy of implementation into production is February 2025. The framework for data quality has offered a tool to assess elements of data quality for this project.

Case Study 6. Designing a New Study: Public Perceptions of Science and Engineering

Author: Darius Singpurwalla, National Center for Science and Engineering Statistics

Overview: In this case study, the FCSM Framework for Data Quality was used to anticipate potential data quality threats during the planning stages for a new study that is designed to measure the nation's perceptions of the science and engineering (S&E) enterprise.

Introduction: One factor that can influence the number of people choosing to pursue careers in science is their perceptions on science, scientists, and investment in science. Simply put, if the public's perceptions of the S&E enterprise are positive then the nation may be able to expect more people will pursue careers in science. To assist researchers and policymakers in understanding this landscape, a new study is being considered that will measure several facets of the public's perceptions of science and engineering.

Description of data being assessed for quality: The study data will consist of various measures of the respondents' perceptions of the S&E enterprise as well as their level of science literacy. In terms of perceptions, the study will measure respondents' perceptions of science and scientists, their personal involvement in science-based activities (e.g., visiting a zoo, participating in a research study), and their science information–seeking behaviors among other constructs. In terms of science literacy (defined as the ability to read and comprehend a scientific article in a popular publication such as the New York Times), the study will measure the respondents' understanding of the process of science and their epistemic science knowledge. These constructs will be measured through one questionnaire consisting primarily of Likert-style, fill-in-the-blank, and multiple-choice questions.

A sample of ~1,800 respondents will be selected from a nationally representative panel of individuals living in the United States. This sample will consist of individuals across a variety of ages, regions, and educational levels.

How the FCSM Framework for Data Quality (DQ) compares with previous efforts to assess DQ on this dataset: Given that this is a new collection, there are no other data quality frameworks or assessments to compare with the FCSM Framework for Data Quality. However, other data sources that have measured similar constructs have historically provided only high-level summaries of the methodology used to collect the data. These summaries do not address the entirety of the data quality threats itemized in the FCSM Framework for Data Quality nor do the assessments provide any information about how these threats were mitigated.

Utility

Since this is a new data collection, a great deal of importance was placed on addressing any threats to the utility domain, particularly the relevancy dimension. The prior sources of information for this study had last been reviewed for content several years ago. Therefore, there exists a risk that the information could be outdated or no longer salient to measuring perceptions. To ensure the relevance of this new

source, we conducted an in-depth literature review, engaged with subject matter experts, and tested the candidate questions developed to measure the selected constructs.

Relevance –

The new data source will include information related to the respondents' demographics, science literacy levels (i.e., process of science, content knowledge, epistemic knowledge), perceptions of science/scientists, science information–seeking behaviors, and engagement in science-related activities. In order to ensure that the most relevant information is collected for this study, we conducted a literature review of recent journal articles studying the public's perceptions of science as well as recruited a panel of experts to help decide which variables were most important to measure.

To identify the articles that could be salient to this study, we reviewed journal articles from the past decade that either measured or otherwise studied the constructs that are similar to what we intend to collect. We searched several databases of scholarly work for terms such as "science literacy," "public perceptions of science," "measuring science literacy," and "science literacy frameworks" to better understand the recent landscape of research on this topic. This review helped us identify the most contemporary frameworks on measuring public perceptions/science literacy. These frameworks helped us select the constructs that should be measured to properly study public perceptions and how to operationalize them for measurement.

While the literature review provided the initial guidance on public perceptions measurement, we also convened an expert panel to assist with finalizing the list of constructs to measure for this project. To organize this effort, we identified several individuals who had expertise in either science literacy or science perception data in general. We held at least two meetings with each of the experts. The first meeting was designed to set the stage for the group's work. During this meeting, we explained the goals of the study, how the data will be used, and how the experts would be engaged throughout the life cycle of the project. Subsequent meetings with the experts focused on the actual measurement items as well as how the data could be analyzed for inclusion in this study .

Accessibility

Individuals who are interested in public perceptions data will be able to access the data through a variety of means. First, the thematic report's author will select several constructs from the study to include as indicators in the report. We will also tabulate most of the variables collected for this study and make them freely available in a variety of formats (e.g., Excel, PDF). For users who wish to conduct tabulations of variable combinations that were not pretabulated, the study data will be disseminated through an online dissemination tool. Lastly, for sophisticated users who may want to model these data, public-use file containing record-level response data will be made available for users to analyze.

Timeliness

The planned periodicity of data collection for this study will be once every 2 years.

Punctuality

Because this is a new study, there is no planned release date for the data. Therefore, periodicity was not considered in this quality assessment.

Granularity

The data will be made available to users at the individual level, allowing researchers to tabulate the data on any characteristics that they desire to measure.

Objectivity

Accuracy and Reliability –Several steps ensure that the estimates from this new collection are both accurate and reliable. As stated earlier, this questionnaire will be developed in concert with several experts in science perceptions research. The experts will provide input both on the constructs to measure and on how to measure them. In addition to the expert review, the questionnaire will go through extensive pretesting before being fielded to participants. As part of the pretesting, each question will undergo cognitive testing to ensure that the respondents understand the question and, when necessary, to refine the questions. For instance, participants may be asked to clarify their understanding of terms (e.g., medical scientist) that are used in the survey. Respondents will also be probed to identify their thought processes and thinking while responding to the questionnaire.

After cognitive testing is complete, the revised questionnaire will be field-tested on a larger group of participants. The goal of this step is to ensure that the instructions, question wording, and survey design are understood by the respondents. Additionally, since testing will be done using a statistically valid sample, its results can be compared to other established estimates of similar constructs to evaluate their accuracy. The reliability of the survey will be evaluated using test-retest methods.

Coherence –The definitions and classifications that will be used in the new study will be both internally consistent and consistent with the definitions commonly used in the science perception field. These definitions and any classifications used in the study will be documented and made available to researchers.

Integrity

Scientific Integrity–The work will utilize modern scientific and statistical standards to uphold the work's scientific integrity. The network of experts that have and will be engaged with throughout the life cycle of this project has helped to ensure that the information collected for this project is contemporary. Additionally, the methods used to develop and disseminate the information from this data collection will also be up-to-date and scientifically sound.

Credibility – This dimension will not be covered in this case study.

Computer and Physical Security –Any data collected for this project will be stored on servers and security protocols that are well-protected from viruses and other malware through monitoring and firewalls.

Confidentiality–The data collected for this study will be protected using appropriate statistical disclosure limitation techniques.

Lessons Learned/sustainability

The framework is a robust tool that can be used in establishing a data collection effort. Because the framework details the various threats to data quality, it proved to be a useful tool during the planning stages of a new study.

Case Study 7: Program Evaluation Case Study: The National Directory of New Hires

Authors: Erika Liliedahl, Office of Management and Budget Evidence Team, and Janet Javar, Department of Labor, Chief Evaluation Office

Overview:

Federal agencies interested in assessing the effectiveness of government programs and services often face limited resources. Thus, agencies often rely on administrative data that are already collected by the government for a different purpose and use the data to aggregate program outcomes and estimate policy impacts. Evaluators have become entrepreneurial in identifying data sets that will support answering research questions of interest through rigorous program evaluation, which may necessarily involve matching across multiple sources or require some supplemental data collection. It is imperative that researchers are transparent about the fitness for use of existing data for program evaluation, to strengthen credibility of study design and findings, and to account for other factors including data privacy.

Introduction: This case study examines how the FCSM Framework for Data Quality serves as a tool to determine the utility of administrative data in determining program effectiveness and provides considerations for the data's use that are paramount to upholding rigor and ethics as principles of program evaluation.

Description of data being assessed for quality:

For a program evaluation regarding analysis of employment outcomes, the National Directory of New Hires (NDNH) from the U.S. Department of Health and Human Services (HHS) was used. It is important to note that this case study examines data quality of the NDNH for a specific study and is not meant to serve as a review of the data quality of the NDNH itself.

The NDNH is a repository of employment, unemployment insurance (UI), and quarterly wage data maintained by the Office of Child Support Enforcement (OCSE) at the HHS. The database, authorized in Sec. 453 of the Social Security Miscellaneous Amendments Act of 1996, was originally designed to help state and federal agencies locate noncustodial parents to establish and enforce child support orders, particularly across state jurisdictions. Since the creation of the NDNH, Congress has authorized specific and limited additional uses of the data. For privacy and security reasons, these authorizations clearly specify the entities that may access the data and/or the purpose for which the data may be used. For more information on the authority to use data contained within the NDNH for this study, please refer to the discussion later in the case study about data accessibility. Additionally, information on the data from HHS is readily available online at https://www.acf.hhs.gov/css/training-technical-assistance/overview-national-directory-new-hires.

The NDNH combines three types of data: employment information on all newly hired employees as reported by employers, quarterly wage information on individual employees, and UI information on individuals who received or applied for unemployment benefits. The information is derived from W-4 records submitted by employers to the State Directory of New Hires, quarterly wage and unemployment insurance data from the state workforce agencies, and new hire and quarterly wage data from federal agencies.

For this study, the quarterly wage data contained within the NDNH database was used to measure group employment outcomes and aggregate impacts of the Transition Assistance Program (TAP). The study sample included transitioning service members (TSMs) from the Army branch who completed a U.S. Department of Labor (DOL) employment workshop between October 1, 2014, and June 30, 2019, prior to transitioning from their military service to civilian life. Select data within the NDNH system were linked with Army data to develop the "treatment" and "control" (through a matched comparison) groups.

For more detailed information on the methods, data, and results of this program evaluation, please refer to the DOL's Chief Evaluation Office website.[38]

The data were used to answer the following questions:

- What is an appropriate matched comparison group for the selected sample of transitioning military members?
- How long did it take for TSMs to find work after they separated from the military?
- Were TSMs employed at 6 and 12 month after they separated from the military?
- What were their wages 6, 12, 24, and 36 months after separation from the military?
- How did employment outcomes differ by subgroup, such as by gender or occupation?

How the FCSM Framework for Data Quality (DQ) compares with previous efforts to assess DQ on this dataset:

In order to facilitate program evaluations such as the TAP impact evaluation, the DOL conducted a review of eight sources of earnings data.[39] The data source review included:

- Earnings data collected by the IRS for tax purposes
- Earnings data maintained by the Social Security Administration in the master earnings file
- Earnings data collected by the state unemployment insurance agencies, housed by each state
- Earnings data contained within the NDNH and obtained from states and federal agencies by HHS
- Longitudinal Employer-Household Dynamics (LEHD) program at the U.S. Census Bureau

[38] See https://www.dol.gov/agencies/oasp/evaluation/completed-reports/evaluation-of-transition-assistance-program . See also *Evaluation of the Transition Assistance Program (TAP) Impact Study Report* (July 2023). Accessed August 2023.

[39] Czajka, John L., Ankits Patnaik, and Marian Negoita. *Data on Earnings: A Review of Resources for Research.*. U.S. Department of Labor, Chief Evaluation Office. Washington, DC. September 2018. Available at: https://www.dol.gov/sites/dolgov/files/OASP/evaluation/pdf/Data-on-Earnings-Report.pdf. Accessed August 2023.

Description of implementation including human capital, technology needed, and cost:

In order to use the NDNH for the TAP evaluation, human capital and other costs were involved. The use of the NDNH requires the allocation of time to develop and execute memorandums of understanding between three federal agencies. Use of the NDNH also requires full reimbursement of the costs of data matching services. In terms of technology costs, users must maintain a secure network with government authority to operate (ATO) and analytical software to access and analyze the data. Human capital and technology costs may come into play in administering and supporting access to the data to authorized users.

Assessment of using the NDNH in the TAP Evaluation using the FCSM Framework for Data Quality

Utility

Relevance –When considering whether the NDNH would provide data of high quality for assessing program outcomes for the TAP, it is important to assess if the data contained in the NDNH are relevant to employment and earnings outcomes. Specifically, if information could be understood about aggregate wages 6, 12, 24, and 36 months after separation and what are the employment outcomes following program completion.

The NDNH includes the new hire file, employee date of hire, employee state of hire quarterly wage file, quarterly employee wage amount contained in the UI file, and the UI benefit amount. The new hire file contains information on the employee's date and state of hire, as well as some information about the employer, including the employer's federal employer identification number (FEIN) and city and state. The quarterly wage file includes information on the gross quarterly wage amount and the employer's FEIN; when an individual works more than one job during the reporting period, separate records are established for each job. The UI file contains quarterly information on all UI claim applications to state workforce agencies (even if the claim is rejected, denied, or suspended), and includes information on the claimant's gross benefit amount and the reporting period. The user can also provide an external pass-through file with sufficient PII for HHS/OCSE to match the external file to individuals' records in the NDNH, thus enabling external data to be linked to the new hire, quarterly wage, or UI files.

A key advantage of using the NDNH for the TAP evaluation instead of the State Directory of New Hires or state UI quarterly wage data is that information is available about people who have obtained work or claimed UI in another state. This is particularly helpful for the TAP evaluation given that it is a multistate evaluation with populations that are highly mobile, likely to commute across state lines, or work for employers operating in multiples states.

Accessibility —Although the NDNH includes existing data collections that are housed within a federal agency, it is important to understand if these data would be accessible by DOL and its evaluators to measure aggregate program outcomes of TAP participants by cohort and to apply methods to calculate aggregate program impacts across the treatment and control groups. Access to the NDNH is explicitly limited by statute. Title IV-D of the Social Security Act specifies that researchers outside covered federal agencies may be given access only to de-identified NDNH information "to conduct research found by the Secretary of Health and Human Services (HHS) to be likely to contribute to achieving the purposes of part A or part D of the Social Security Act" (42 U.S.C. §653(j)(5)), that is, to contribute to achieving the

mission of Temporary Assistance for Needy Families and Child Support Enforcement programs. Before accessing NDNH information, users must sign an agreement or memorandum of understanding with HHS/OCSE that describes the purpose, legal authority, justification, expected results of the match, description of the records, retention and disposition of information, reimbursement, and user's performance reporting requirements, as well as a security addendum that details the security requirements and safeguards which users must have in place before receiving NDNH information.[40] HHS/OCSE also requires NDNH users to provide a written description of the performance outputs and outcomes attributable to using NDNH information. Users must reimburse HHS/OCSE for the costs of obtaining, verifying, maintaining, and comparing the information. It has been noted "that the usefulness of NDNH data for research is limited by its data deletion and de-identification requirements"[41] which also relates to the timeliness domain.

Timeliness –Data timeliness requires complex planning for the TAP evaluation, as considerations of employment and earnings status prior to program participation, program completion timing, and reasonable timing for short- and long-term outcomes are equally important factors. Then, considerations of when the data are available for access and analysis on TAP outcomes is a complicating consideration. There can be a lag of up to 4.5 months from the end of a calendar quarter before data become available in the NDNH. All information entered into the NDNH is purged within 24 months; quarterly wage and UI data that do not result in a child support match are often purged within 12 months. However, when a memorandum of understanding is signed, HHS/OCSE does have the ability to retain data if they are part of an ongoing research sample for specific users. Obtaining authorization for the data and working out the details of the implementation process (such as the fee structure and testing and validation of the match) can take several months. Through agreements for the TAP evaluation, DOL can make data requests to HHS every quarter, and HHS is timely in holding the data. However, the lag time of the employment data itself is up to 4.5 months. Overall, the timeliness considerations are paramount in order to plan a successful evaluation of the TAP using the NDNH; the timing details are critical not only for the evaluation's design phase but also for determining and planning for accessibility and carrying out the analysis.

Punctuality- This dimension will not be covered in this case study.

Granularity –Once merged with the Army participant files, the NDNH matched dataset for the TAP evaluation provides granular linked demographic, military career, and program participation characteristics. For example, demographic characteristics include age at separation, gender, race and ethnicity, marital status, number of dependents, level of education before military, and disability status at separation. Military characteristics include pay grade at separation, military occupational specialist, Armed Forces Qualification Test percentile score, base assignment at separation, deployments, years of service, and type of discharge. Program participation data include participation in specific components of the program, such as financial planning, military occupation crosswalk, DOL elective courses, etc. Together the data on these demographics and characteristics allowed for important subgroup analysis that was useful in addressing research questions. Due to NDNH restrictions, restricted- or public-use

[40] See https://www.acf.hhs.gov/sites/default/files/documents/ocse/ndnh_research_request_form_fillable.pdf.
[41] Congressional Research Service. *The National Directory of New Hires: In Brief.* Washington, DC. Updated October 1, 2019. Available at: https://crsreports.congress.gov/product/pdf/RS/RS22889. Accessed August 2023.

datasets cannot be created, so although the blended dataset could be used for analysis, the results did not include identifiable or re-identifiable information, nor would the resulting dataset be available for any uses beyond the TAP evaluation study.

Objectivity

Accuracy and Reliability – For the TAP evaluation, the participants being observed were in the military prior to TAP participation and had become former military when the employment outcomes were being assessed in aggregate form. The program outcomes of the treatment and of the control groups are then used to calculate program impacts. It is critical to understand if the populations being observed will be captured in the data and consistently over the course of program participation and completion.

The universe of records in the NDNH is formal employment of individuals, but excludes self-employed individuals, most independent contractors, railroad employees, some part-time employees of nonprofit institutions, employees of religious orders, and some students employed by their schools. The NDNH does not include employment data for people without a valid Social Security number. HHS/OCSE has reported prior that 97% of all records submitted to the NDNH are available for matching.[42] Quarterly wage and UI data derived from the state UI agencies are susceptible to the study universe and coverage issues due to underreporting.[43]

An advantage over the LEHD and state UI records is that the NDNH covers those employed by federal or military agencies. However, NDNH, state UI, and LEHD do not have data on self-employment or informal or off-the books employment.[44] Additionally, none of these data sources have information on job characteristics (such as industry, occupation, and hours worked). The NDNH is a not a source of historical data as it contains only up to 2 years of data at the time of application for access.

*Coherence***-** This dimension will not be covered in this case study.

Integrity

*Scientific Integrity –*The released data adhere to scientific standards and methods and are free from outside influence.

Credibility **–** This dimension will not be covered in this case study.

Computer and Physical Security –

Federal agencies must meet the HHS/OCSE requirements for receiving, maintaining, and providing access to data for authorized users, including as part of the agreement for the TAP evaluation. HHS/OCSE provides transparent, publicly available information on how to request the data as well as upfront expectations and limitations. A restricted-access government computer network is used to secure the data. Access is granted only to those who are listed as authorized users in data sharing

[42] Czajka et al, 2018.
[43] Czajka et al, 2018.
[44] Durham, Christin and Laura Wheaton. *Investigating Alternative Sources of Quarterly Wage Data: An Overview of the NDNH, LEHD, WRIS, and ADARE*. Urban Institute. New York. 2012. Available at: https://www.urban.org/research/publication/investigating-alternative-sources-quarterly-wage-data-overview-ndnh-lehd-wris-and-adare. Accessed August 2023.

agreements. All analysis must take place on the network and data cannot be exported. All researchers sign nondisclosure agreements and must comply with contractual requirements for securing and protecting data.

Confidentiality– The process outlined in the agreement for the NDNH to support the TAP evaluation followed HHS/OCSE requirements. HHS has clear parameters about external pass-through files, with the aim of limiting disclosure from data linkage. After the matching occurs, HHS/OCSE returns the pass-through file, stripped of all identifying information and including a masked identifier generated by HHS/OCSE, that can be used to link to a de-identified NDNH file that is given to the user and includes the same masked ID.

However, because files are de-identified, it is impossible for researchers to later incorporate additional years of data, link to new sources of data, or even correct problems with prior linkages after the de-identified file with NDNH data has been returned. This is a great example of how data users, including evaluators, must make important decisions across the domains and dimensions of the data quality framework before moving forward with using data and how the data quality framework offers a systematic structure to document these trade-offs.

Lessons learned/sustainability:

This case study offers insights into the use of the NDNH for the TAP evaluation, and the process of documenting the case study also provided some broader lessons on how the FCSM framework could be used to support program evaluation planning, including in determining whether a dataset is a good fit for a particular program evaluation. For example, Will the data allow for analysis of a sufficient sample to detect program impacts? Will the data facilitate subgroup analysis on important groups? This section offers insights that other evaluators might consider in applying the data quality framework for future evaluations.

The factors documented in this case study are typical of what evaluators consider as they plan and design evaluation methodologies and the data to support this analysis. These considerations for assessing data quality for use in a program evaluation project are mapped in Table 3 to the item in the FCSM data quality framework that is most closely associated.

Table 3. Data Considerations in a Typical Program Evaluation and Their Corresponding Dimension and Domain in the FCSM Data Quality Framework

Typical program evaluation data consideration	FCSM data quality framework	
	Dimension	Domain
Description of the data and their origin	• Relevance	Utility
Study universe from which the data are collected and any known coverage issues (both population and amounts)	• Relevance • Granularity	Utility
Information included (for example, the availability and quality of information)	• Accuracy and reliability • Credibility	Objectivity and Integrity
Options for data linkage	• Granularity	Utility

Procedures and practices surrounding access	• Accessibility • Computer and Physical Security • Confidentiality	Utility and Integrity
Timeliness of the data—from collection to availability and from application to receipt	• Timeliness	Utility
Key strengths and limitations for analysis	• Any dimension identified as a key strength and limitation	Utility, Objectivity, and Integrity

Overall, the case study documents that the data contained within the NDNH are useful for program evaluations that examine aggregate participant outcomes and/or program impacts. However, due to the restrictive access to the data, the methods that can be applied to the data are limited. For example, researchers receive de-identified results by specific groups, which is not useful for merging with other datasets for more complex analysis, for increasing match rates, or for the creation of a blended data sets, nor does it facilitate longitudinal analysis. Using the NDNH for program evaluation under current access authorities and procedures plays a large role in important methodological choices for an evaluation study. The methodological choices are also bound by considerations about data quality that would naturally occur in the course of data curation and analysis. The data quality framework is an incredibly useful decision tool when determining which source of administrative data on wages or employment may be most relevant and for conducting a particular program evaluation, and it also offers an opportunity to better understand the value of access to restricted-use data files.

The case study reveals that the FCSM Framework for Data Quality provides a consistent framework to discuss the difficult yet realistic data quality balance for program evaluation, such as:

- Balance between timeliness vs. data accuracy
- How to measure outcomes in meaningful ways
- Abilities, authorities, and capabilities needed to obtain data
- Important contextual information on the data critical to understanding the data and implications of use
- Data cleaning or validation (e.g., missing data imputations, elimination of obvious data reporting errors or inconsistencies) necessary in order to ready data for analysis on the outcomes of interest

As data generators and data users look to future work, it is crucial that owners of data are as transparent as possible about data quality so that data users, including program evaluators, can make well-informed choices about data quality. Conversely, it is imperative that researchers are transparent about the fitness for use of existing data for program evaluation, to strengthen credibility of study design and findings, and to account for other factors including data privacy. The FCSM Framework for Data Quality can serve as a valuable tool to determine the utility of administrative data in determining program effectiveness and provides considerations for the data's use that are paramount to upholding rigor and ethics as principles of program evaluation. The FCSM Framework for Data Quality also gives data collectors, data creators, and data users a common language to discuss data quality.

Appendix I. FCSM Framework for Data Quality Case Study Template

Overview:

Introduction:

Description of data being assessed for quality:

How the FCSM Framework for Data Quality (DQ) compares with previous efforts to assess DQ on this dataset:

Description of implementation including human capital, technology needed and cost:

Assessment [data source] using the FCSM Framework for Data Quality

Utility

Relevance –

Accessibility –

Timeliness –

Punctuality –

Granularity –

Objectivity

Accuracy and Reliability –

Coherence –

Integrity

Scientific Integrity –

Credibility –

Computer and Physical Security –

Confidentiality –

Lessons learned/sustainability:

Appendix II. Dimension Definitions

Utility

Relevance: Relevance refers to whether the data product is targeted to meet current and prospective user needs.

Accessibility: Accessibility relates to the ease with which data users can obtain an agency's products and documentation in forms and formats that are understandable to data users.

Timeliness: Timeliness is the length of time between the event or phenomenon described by the data and their availability.

Punctuality: Punctuality is measured as the time lag between the actual release of the data and the planned target date for data release.

Granularity: Granularity refers to the amount of disaggregation available for key data elements. Granularity can be expressed in units of time, level of geographic detail available, or the amount of detail available on any of the number of characteristics (e.g., demographic, socio-economic).

Objectivity

Accuracy and reliability: Accuracy measures the extent to which an estimate from a data product reflects the data's true value. Reliability, a related concept, characterizes the consistency of results when the same phenomenon is measured or estimated more than once under similar conditions.

Coherence: Coherence is defined as the ability of the data product to maintain common definitions, classification, and methodological processes, to align with external statistical standards, and to maintain consistency and comparability with other relevant data.

Integrity

Scientific integrity: Scientific integrity refers to an environment that ensures adherence to scientific standards and use of established scientific methods to produce and disseminate objective data products and shields these products from inappropriate political influence.

Credibility: Credibility characterizes the confidence that users place in data products based simply on the qualifications and past performance of the data producer.

Computer and physical security: Computer and physical security of data refers to the protection of information throughout the collection, production, analysis, and development process from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

Confidentiality: Confidentiality refers to a quality or condition of information that is protected by an obligation not to disclose that information to an unauthorized party.