**IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA**

NATIONAL WOMEN'S LAW CENTER, et al.	))	
	)	
Plaintiffs,	)	
	)	Civil Action No. 17-2458 (TSC)
v.	)	
	)	
OFFICE OF MANAGEMENT AND BUDGET, et al.	))	
	)	
Defendants.	)	

DECLARATION OF SAMUEL C. HAFFER

I, Samuel C. Haffer, Ph.D., declare as follows:

1. I am the Chief Data Officer (CDO) and Director of the Office of Enterprise Data and

Analytics (OEDA) at the U.S. Equal Employment Opportunity Commission (EEOC or

Commission). I hold a Bachelor of Arts in Sociology (1988) from Loyola College, and a Master

of Public Policy (1991) and Ph.D. in Public Policy (1993) from University of Maryland –

Baltimore County.

2. I have over 27 years' federal experience in data, analytics, and research, including 14

years as Director of the National Medicare Health Outcomes Survey (MHOS) at the U.S.

Centers for Medicare and Medicaid Services (CMS), a component of the U.S. Department of

Health and Human Services. From its inception in 1997, the MHOS program collects

sensitive and confidential physical and mental health data from millions of Medicare

recipients enrolled in hundreds of Medicare managed health care plans. I led all aspects of

the MHOS program including question and measure design, testing, and implementation;

secure data collection systems design, programming, testing, and launch; data collection

training design and delivery; data collection; data cleaning; data quality assurance design, testing, and implementation; data validity and reliability studies; data analysis; and planning, construction, testing, creation, and dissemination of user specific analytical data files and documentation to agency enforcement and oversight staff as well as public data users. As a result of my work, I received a National Institutes of Health Award of Merit for my work in 2012.

3.   In the course of my work, I have authored or co-authored over three dozen peer-reviewed publications, and commissioned, oversaw, critically reviewed, and approved numerous data collection quality assurance guidelines, technical specifications, and methodological studies. These include the following related specifically to study design, data collection quality assurance, and quality control:

Study Design

Functional Health Outcomes as a Measure of Health Care Quality for Medicare Beneficiaries. 2001. Bierman AS, Lawrence WF, Haffer SC and Clancy CM. Health Services Research. December 2001. Volume 35(6) Part II: 90-109.

Measuring and Improving Health Outcomes in Medicare: The Medicare HOS Program. 2004. Haffer, SC and Bowen S. Health Care Financing Review. Summer 2004. Volume 25(4): 1-3.

Calculating HOS Performance Measurement Results. 2004. Rogers WH, Gandek B and Sinclair SJ.

Data Collection Quality Assurance and Quality Control

Medicare Health Outcomes Survey (HOS) Quality Assurance Guidelines and Technical Specifications. Washington DC: NCQA Publication. 1997-2010. www.ncqa.org

National Committee for Quality Assurance (NCQA). HEDIS® Volume 6: Specifications for the Medicare Health Outcomes Survey. Washington DC: NCQA Publication. 1997-2010. www.ncqa.org

Weight Adjustments in Estimates for the 1999 Medicare Health Outcomes Survey. 2002. Hwang Y, Bierman AS, Haffer SC and Wun LM. ASA Proceedings of the Joint Statistical Meetings. 2002. 1565-1570.

Using Multiple Survey Vendors to Collect Health Outcomes Information: How Accurate Are the Data? 2003. Haffer SC. Health and Quality of Life Outcomes. April 16, 2003. Volume 1(6).

Medicare Health Outcomes Survey: An Alternative Case-Mix Methodology. 2007. Selim AJ, Iqbal SU, Rogers W, Qian SX, Fincke BG, Rothendler J and Kazis LE.

Imputing Physical and Mental Summary Scores (PCS and MCS) for the Veterans SF-12 Health Survey in the Context of Missing Data. 2004. Spiro A, Rogers WH, Qian S and Kazis LE.

Imputing the Physical and Mental Summary Scores (PCS and MCS) for the MOS SF-36 and the Veterans SF-36 Health Survey in the presence of Missing Data. 2004. Rogers WH, Qian S and Kazis LE.

HOS/VA (Veterans Administration) Comparison Project Part 1: Measurement Equivalence of Medicare HOS SF-36 and VA Veterans SF-36 Spiro A, Lee AF, Kazis LE, Miller DR, Ren XS and Zhang M.

HOS/VA Comparison Project Part 2: Test of Reliability and Validity at the Scale Level for the Medicare HOS SF-36 and VA Veterans SF-36 (PDF, 63 KB) Kazis LE, Lee AF, Spiro A, Miller DR, Rogers W, Ren XS and Zhang M.

Final reports for the methodological studies are available at:
https://www.hosonline.org/en/publications/methodology/

4.   My entire career of nearly three decades has focused on research study design, measure development, data collection, data analysis, and data dissemination. I am knowledgeable of the federal government's requirements for data collection, protection, security, and dissemination. In addition, I am knowledgeable of the requirements of the Federal Information Security Modernization Act of 2014 (Pub. L. No. 113-283) and Foundations for Evidence-Based Policymaking Act of 2017 (Pub. L. No. 115-435), which direct federal agencies to adhere to stringent security and privacy standards.

5. Data collection is by definition a process built around a range of activities designed to maximize accuracy and consistency in the final data. The US Department of Health and Human Services Office of Research Integrity (ORI), created in May 1992 to implement activities and programs to teach the responsible conduct of research, promote research integrity, prevent research misconduct, and improve the handling of allegations of research misconduct,[1] defines the data collection process:

> The process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. The data collection component of research is common to all fields of study including physical and social sciences, humanities, business, etc. While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same.
>
> . . . .
>
> [A]ccurate data collection is essential to maintaining the integrity of research. Both the selection of appropriate data collection instruments (existing, modified, or newly developed) and clearly delineated instructions for their correct use reduce the likelihood of errors occurring.[2]

ORI warns that the consequences of improperly collected data may include:

> o inability to answer research questions accurately
> o inability to repeat and validate the study
> o distorted findings resulting in wasted resources
> o misleading other researchers to pursue fruitless avenues of investigation compromising decisions for public policy.[3]

6. I joined the EEOC in November 2017 in the newly created role of CDO (a career position in the Senior Executive Service). Among my duties, I am responsible for administration of the EEO-1, -3, -4, and -5 data collections (EEO Surveys). I am thoroughly familiar with how

---

[1] United States Department of Health and Human Services, Office of Research Integrity. https://ori.hhs.gov/about-ori. Accessed April 2, 2019.

[2] United States Department of Health and Human Services Office of Research Integrity. Responsible Conduct in Data Collection: Data Management. https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/dctopic.html. Accessed April 1, 2019.

[3] Ibid.

employers report, and the EEOC receives the information reported as part of the EEO-1

information collection, as I have led the agency through one entire reporting cycle. I also am

aware that, depending on their size, many employers prepare and submit multiple data reports.

Issues Identified in Current EEO-1 Data Collection

7.   When I arrived at the EEOC in November 2017, EEOC Acting Chair Victoria A. Lipnic

tasked me with reviewing all products and services of the EEOC's 20-year-old (now former)

Office of Research, Information, and Planning (ORIP), and to create a 21st century data and

analytics department. The primary role of this office was to aid the Commission in its efforts to

achieve its mission and to strategically plan its goals and objectives by researching, collecting,

analyzing, and disseminating relevant data and information, including administering all aspects of

the EEO Surveys.

8.   Upon my arrival, I also moved quickly to review and familiarize myself with the EEO

Surveys, including the record of the development of the revised EEO-1 information collection that

was approved by the Office of Management and Budget (OMB), under the Paperwork Reduction

Act (PRA), on September 29, 2016. I am aware that OMB authorized the revised EEO-1

information collection for a three-year term expiring on September 30, 2019.

9.   I proceeded on two different fronts at once. I assessed the EEOC's current EEO-1 data

collection process and concluded that it needed immediate improvement. Separately, I worked

to initiate a comprehensive data collection modernization project, in coordination with OMB, to

update the EEOC's data collection program (including the EEO-1), to use 21st century data

collection procedures and technologies.

10. As to the current process for collecting EEO-1 data, including the EEOC's online system

for collecting the data (called the Online Filing System (or the EEO-1 Survey Application)), my

initial assessment identified two key areas for immediate improvement: quality assurance and quality control. Quality assurance activities take place before data collection begins and may involve, for example, a well-designed and implemented data training plan to assure the collection of high-quality data.[4] I observed that the former ORIP did not have a 2017 training plan in place and had not revised its EEO-1 instruction manuals and training materials to reflect the 2017 data collection requirements and methods for data submission.

11. Quality control activities take place during and after data collection and include cross-checks within the data collection processes and systems to assess the potential for errors in individual data items, systematic errors, and violations of protocol.

12. I found that, for several previous EEO-1 cycles, the former ORIP had not sufficiently tested its data collection system in a rush to open data collection on time. This necessitated untested programming code patches and workarounds, which introduced a potential source of error into the data collection and ultimately resulted in more delay. For instance, the EEOC announced that the 2016 EEO-1 report, which collected Component 1 data, would close on September 30, 2016. For the 2017 EEO-1 report, which also collected Component 1 data, the announced filing deadline was March 31, 2018. Due to such data quality control issues, the 2016 EEO-1 reporting cycle actually closed on August 4, 2017, eleven (11) months later than the announced deadline of September 30, 2016. The 2017 EEO-1 reporting cycle actually closed on October 1, 2018, six (6) months later than the announced deadline of March 31, 2018.

13. When the OMB stay of the PRA authorization for Component 2 was in effect, I understood that the EEOC could not collect Component 2 data from employers. Therefore, the

---

[4] Ibid. See also American Association for Public Opinion Research (AAPOR). *Best Practices for Survey Research*. https://www.aapor.org/Standards-Ethics/Best-Practices.aspx#best7. Accessed April 1, 2019.

EEO-1 data collection process for calendar year 2017 data, the first one I observed at the EEOC, collected only Component 1 data. In the short time I was employed as EEOC's CDO before this collection of 2017 Component 1 data commenced, I worked to address quality assurance and quality control issues with Component 1. I became aware during this time that the EEOC had not yet supplemented its current system to collect Component 2 data and, given the challenges identified with the current system, I recognized that a Component 2 data collection would likely have needed serious reassessment.

The Need to Modernize and Secure the EEOC's Data and Analytics Processes and Capabilities

14.   My responsibility as the new CDO was to bring the quality, usability, and actionability of products and services (deliverables) produced by the former ORIP up to industry standards. In addition to the challenges identified above, and among other issues, my assessment of the former ORIP found that the work products generated by the office were usually untimely and often methodologically and statistically unsound.

15. These findings led to a restructuring and reorganization of the former ORIP into OEDA, which was approved by Acting Chair Lipnic in May 2018, after vetting through the Commission's required internal review process. As part of the reorganization, OEDA enhanced its expert staff capacity by adding statisticians and data scientists with expertise in data collection and analytics. Hiring was completed in December 2018. In mid-December 2018, the EEOC held three days of stakeholder meetings to introduce OEDA and its staff to the public and regulated community.

16. Simultaneously, the EEOC had been working closely with OMB to begin to upgrade all of the EEO Surveys, giving rise to the EEOC Data and Analytics Modernization Program, a comprehensive evaluation of the collection, analysis, and dissemination of EEOC data.

17. As part of the Modernization Program:

o The EEOC is conducting a multi-year review of all aspects of current data collection processes for the EEO Surveys. With this work the EEOC is looking to ensure that the data collection instruments and implementation processes for each EEO Survey comply with federal policies and align with best practices.

o The EEOC is also looking to identify opportunities to improve efficiency, as well as to minimize burden and maximize data quality.

o The EEOC will be assessing data collection processes to see how they align with industry standards and making improvements as appropriate.

o The EEOC will also be exploring potential alternative data sources to see if there are any available that the EEOC could obtain to supplement information needed to fulfill the agency's data needs.

o The EEOC plans to conduct focus groups with relevant stakeholders to ascertain views on current processes – such as what is and is not working, as well as alternative data sources and alternative methods of collection. The EEOC will also be using focus groups to assist the EEOC in more accurately estimating the costs to employers of implementing the alternatives for each of the EEOC Surveys.

o The EEOC is in the process of gathering requirements for building out a modern centralized enterprise data analytics warehouse with secure remote access by EEOC employees.

o The EEOC is currently in the process of developing a secure network through which confidential data, such as identifiable information from the EEOC, can be stored and accessed by researchers remotely.

o To aid in releasing detailed data but still protecting our respondents, the EEOC will be creating public use files to share more detailed data with the public.

o The EEOC will be producing an agency-wide data inventory.

o The EEOC is examining how it can use other statistical agencies' data to improve on the quality of the data the EEOC produces.

o The EEOC created and is now utilizing an EEOC Data Governance Board, whose core mission is to provide executive leadership and oversight for the development and implementation of the policies and processes which govern the collection or creation, management, use, and disclosure of EEOC data.

18. To undertake the Modernization Program, in the fall of 2018 the EEOC entered into a multi-year contract with NORC at the University of Chicago (NORC), an industry-leading data and analytics organization. NORC's tasks under the contract include, among other things, the work described above in addition to the review and modernization of the EEO-1 data collection as a whole.

19. Until that modernization is completed, however, the EEOC must continue using its current Online Filing System (or the EEO-1 Survey Application) for Component 1, with as many improvements put into place by OEDA staff as possible.

The EEOC Cannot Begin to Collect Component 2 Data Immediately; the EEOC Could Undertake and Close a Collection of Component 2 Data by September 30, 2019

20. Currently, the EEOC's Online Filing System (or the EEO-1 Survey Application) used for the collection of EEO-1 data is programmed to collect 140 data fields (10 Job Categories x 7 Race/Ethnicity Categories x 2 Gender Categories) for each Component 1 report required to be filed by over 80,000 employers. The addition of the collection of Component 2 data to the EEOC's Online Filing System would increase the data reported to the EEOC by a large order of magnitude. For each Component 2 report collected from an employer, there are 3,360 data fields to be addressed (3,360 represents the sum of the 1,680 data fields for pay data (10 Job Categories x 12 Salary Bands x 7 Race/Ethnicity Categories x 2 Gender Categories), as well as the 1,680 data fields for hours worked data (10 Job Categories x 12 Salary Bands x 7 Race/Ethnicity Categories x 2 Gender Categories)).

21. I determined that, for the EEOC to make the necessary updates, enhancements, security testing, load and performance testing, data validations and verifications, and application testing to securely collect and store this significantly increased volume of highly sensitive Component 2 data using its current Online Filing System (or the EEO-1 Survey Application), approximately

nine (9) months preparation would be required using the resources in place. These preparations could not be concluded until December 2019, which I understand is after the EEO-1's current PRA authorization expires.

22. I am aware that employers believe that they are likely to experience significant issues regarding the immediate reporting of Component 2 data, based on a March 29, 2019 letter to Acting Chair Lipnic from the National Payroll Reporting Consortium.[5]

23. I am aware that requiring employers to collect and report calendar year 2017 Component 2 data at the same time as the collection of calendar year 2018 Component 2 data without more extensive and well-developed quality assurance and quality control processes in place (c.f. paragraphs 10-12) could decrease response rates and increase errors in the entire data collection process to a greater extent than might be experienced in collecting 2018 Component 2 data alone.

24. However, I have also determined that, through NORC, the EEOC could undertake and close a collection of calendar year 2018 Component 2 data by September 30, 2019. The contractor would provide in a condensed amount of time and in an expedited manner all processes, procedures, and systems to undertake and close a collection of calendar year 2018 Component 2 data by September 30, 2019. This system would be utilized one time for the collection of calendar year 2018 Component 2 data only. It would not be utilized after the EEOC makes its transition to the modernized data collection process.

25. The estimated cost to the EEOC to conduct this Component 2 data collection for calendar year 2018 by September 30, 2019, would be in excess of $3 million. Collecting 2017 Component 2 data at the same time as 2018 Component 2 data would increase the EEOC's fixed

---

[5] http://www.nprc-inc.org/blog/eeoc-may-reinstate-pay-data-reporting/. Accessed April 2, 2019.

and marginal costs by 25 percent to approximately $4 million and would likely further

undermine data quality, including the quality of the 2018 data.

26. The estimated timeline for the collection of Component 2 data is:

i.   Data Quality Assurance

   o   Prep Work Task – April 2019 - June 30, 2019: Create, develop, and deliver the
       online data collection training to employers; develop, implement, and maintain
       data collection processes, procedures, and systems; contact employers with
       information on secure processes for submitting data.

ii. Data Quality Control

   o   Data Collection Task – July 1, 2019 - September 30, 2019: Initial notification via
       email and mail to employers, starting on July 1; collect data for a 2.5-month
       period (July 15 - September 30); contact non-respondents during the data
       collection period through email and mail letter reminders; operate a helpdesk via
       email and phone for employers starting July 1.

   o   Data Cleaning Task – October 1, 2019 - November 30, 2019: Data cleaning (by
       contractor) to produce as accurate a data set as possible and ensure that data is
       fully comprehensive; conduct data accuracy studies to identify data submissions
       that cannot be certified as accurate or sufficiently complete.

   o   Data Delivery Task – December 1, 2019 - December 15, 2019: Cleaned data set
       with full documentation and formatting delivered to EEOC.

27. Despite these efforts, and as discussed below, I am concerned that Component 2 data

collected by September 30, 2019, may not have sufficient integrity to support data comparisons

or other analyses because of the limited quality control and quality assurance measures that

would be implemented due to this expedited timeline.

Component 2 Data Utility Concerns

28. The goal of the EEO-1 data collection is to collect valid and reliable data that may be

used to enforce employment discrimination laws. Historically, the EEOC has also released

aggregated EEO-1 data to the public once the data is determined to be valid and reliable.

Monette, Sullivan, and DeJong[6] define validity as "the accuracy of a measure: Does it accurately

measure the variable that it is intended to measure" (p. 113). The measures used (i.e., the

questions asked) must actually measure the concept that one is interested in measuring (data

validity). Monette, Sullivan, and DeJong define reliability as "a measure's ability to yield

consistent results each time it is applied. In other words, reliable measures do not fluctuate

except because of variation in the variable being measured" (p. 116). The measures used must

gather the same information each time the data are collected independent of whomever is

reporting the data. Employers must understand exactly what data are being requested, and how

to appropriately gather, format, and report the data so that the data gathered are consistent across

all employers. Reliably reported data enables valid group comparisons within and between

employers (i.e., "apples to apples"). To ensure reliable data, employers should receive data

collection training, instructions, directions, and technical assistance. Employers use such data

collection training, instructions, directions, or technical assistance to program and test their

human resource management systems and to program and test their payroll systems, which may

not be interoperable, to produce the necessary data; other employers may use the information to

gather the data manually.

29. The proposed timeline for undertaking and closing a collection of Component 2 data by

September 30, 2019, raises significant issues with data validity and data reliability. Both the

National Academy of Sciences report referenced in the 2016 30-Day PRA Notice as well as the

HHS Office of Research Integrity recommend that prior to data collection, a true pilot study

collecting real data in as similar a manner as possible to the proposed data collection approach be

---

[6] Monette, Duane R., Thomas J. Sullivan, and Cornell R. DeJong. 1990. *Applied Social Research: Tool for the Human Services. Second Edition.* Fort Worth, Texas. Harcourt Brace Jovanovich College Publishers.

conducted. Conducting a true pilot study is a quality assurance step that can help reduce the

likelihood of introducing measurement error into the data collection by detecting:

- o   Uncertainty about the timing and methods for collecting data;
- o   Confusion about the data required for collection;
- o   Vague descriptions of data collection instruments to be used in lieu of rigorous step-by-step instructions;
- o   Specific content and strategies for training or retraining employer staff responsible for data collection.[7]

Failure to conduct a true pilot study increases the likelihood that measurement error will be

introduced into the data.

30. The EEOC has not conducted a true pilot study of the Component 2 data collection

measures, instrument, or processes. A pilot study is defined as "a small scale 'trial run' of all the

procedures planned for use in the main study."[8] During a pilot study, using a subset of actual

respondents, the data collection tool will be administered, and the processes and procedures for

selecting and reporting data will be tested. Utilizing a true pilot study "[i]mproves the validity of

the data collected and bolsters our confidence in the conclusions drawn."[9]

31. The "EEOC pilot study" referenced in the 2016 30-Day PRA Notice does not meet the

definition of a pilot study because it did not administer the data collection tool to a subset of

respondents. It did not test or validate any data collection processes or procedures for collecting

and reporting Component 2 data. It contained a recommendation as to the most appropriate

definition of pay and unit of pay to be collected. It reviewed existing compensation definitions

and how compensation data are collected in other contexts and for other purposes, and it

expressed opinions about how such definitions and data collection instruments may work in the

EEOC context**.**

---

[7] United States Department of Health and Human Services, Office of Research Integrity. Supra note 2.
[8] Monette, Sullivan, and DeJong. p. 11.
[9] Ibid.

32. Given the absence of a true pilot study leading up to the 2016 authorization of

Components 1 and 2 of the EEO-1, and given the abbreviated period available in which to

develop and implement quality assurance processes and procedures separate from the

Modernization Project (i.e., data training, instructions, directions, and technical assistance for

employers), it is likely that undertaking and closing the collection of Component 2 data by

September 30, 2019 would raise major data validity and reliability issues. Under the

circumstances, I perceive a significant risk that employers would not be reporting comparable

data that can be used by the government or others in meaningful comparisons or analyses.

33. Pursuant to 28 U.S.C. § 1746, I declare under penalty of perjury that the foregoing is true

and correct to the best of my knowledge.

Executed this 3rd day of April, 2019, in Coral Gables, Florida.

Samuel C. Haffer, Ph.D.
Chief Data Officer
U.S. Equal Employment Opportunity Commission

14