

LEXSEE 2010 U.S. DIST. LEXIS 37621

UNITED STATES OF AMERICA, ex rel. REBECCA GONZALEZ, Plaintiff/Relator, v. FRESENIUS MEDICAL CARE NORTH AMERICA, et al., Defendants.

EP-07-CV-247-PRM

UNITED STATES DISTRICT COURT FOR THE WESTERN DISTRICT OF TEXAS, EL PASO DIVISION

2010 U.S. Dist. LEXIS 37621

March 24, 2010, Decided

SUBSEQUENT HISTORY: Motion granted by, in part, Motion denied by, in part United States ex rel. Gonzalez v. Fresenius Medical Care North America, 2010 U.S. Dist. LEXIS 37632 (W.D. Tex., Mar. 31, 2010)

PRIOR HISTORY: United States ex rel. Gonzalez v. Fresenius Med. Care N. Am., 2010 U.S. Dist. LEXIS 37636 (W.D. Tex., Jan. 21, 2010)

COUNSEL: [*1] For Rebecca Gonzalez, Plaintiff, Relator: Mark Dawson Jarmie, LEAD ATTORNEY, Jarmie & Associates, Albuquerque, NM; Mark D. Standridge, LEAD ATTORNEY, PRO HAC VICE, Jarmie & Associates, Las Cruces, NM; Thomas E. Stanton, LEAD ATTORNEY, Law Offices of Thomas Stanton, El Paso, TX.

For Fresenius Medical Care North America, Defendant: A. Kevin Troutman, LEAD ATTORNEY, Fisher & Phillips LLP, Houston, TX; Edward L. Dowd, Jr., James F. Bennett, Megan S. Heinsz, LEAD ATTORNEYS, St. Louis, MO; James O. Darnell, LEAD ATTORNEY, Jim Darnell, P.C., El Paso, TX; Jennifer L. Aspinall, LEAD ATTORNEY, PRO HAC VICE, Dowd Bennett LLP, Clayton, MO; Juanita Rose Brooks, LEAD ATTORNEY, PRO HAC VICE, Fish & Richardson P.C., San Diego, CA; Roger Allen Denning, LEAD ATTORNEY, Fish & Richardson, PC, San Diego, CA; Stephen J. Roppolo, LEAD ATTORNEY, Fisher & Phillips L.L.P., Houston, TX.

For M.D. Alfonso Chavez, Defendant: Richard Andrew Bonner, LEAD ATTORNEY, Kemp Smith LLP, El Paso, TX.

For Larry Ramirez, Defendant: A. Kevin Troutman, LEAD ATTORNEY, Fisher & Phillips LLP, Houston, TX; Edward L. Dowd, Jr., James F. Bennett, LEAD ATTORNEYS, St. Louis, MO; James O. Darnell, LEAD ATTORNEY, Jim Darnell, P.C., El Paso, [*2] TX; Stephen J. Roppolo, LEAD ATTORNEY, Fisher & Phillips L.L.P., Houston, TX.

JUDGES: PHILIP R. MARTINEZ, UNITED STATES DISTRICT JUDGE.

OPINION BY: PHILIP R. MARTINEZ

OPINION

MEMORANDUM OPINION IN SUPPORT OF ORDER GRANTING DEFENDANTS' MOTION TO BAR TESTIMONY OF JASON BALDRIDGE

On January 29, 2010, the first day of trial, the Court granted Defendant Fresenius Medical Care North America et al.'s (Fresenius) ¹ "Motion to Bar Testimony of Dr. Jason Baldridge" ² [hereinafter Fresenius's Mot. to Bar Test.] (Docket No. 190), filed June 17, 2009, orally on the record, having also considered Relator Rebecca Gonzalez's (Relator) "Response to Fresenius' Motion to Bar Testimony of Dr. Jason Baldridge" [hereinafter Relator Resp.] (Docket No. 229), filed July 9, 2009, and Fresenius's "Reply in Support of Its Motion to Bar Baldridge" [hereinafter Fresenius's Reply] (Docket No. 238), filed July 16, 2009. On February 8, 2010, the Court issued a written order granting the motion and explaining that the testimony of Jason Baldridge is inadmissible

under Federal Rule of Evidence 702 and *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 113 S. Ct. 2786, 125 L. Ed. 2d 469 (1993), because it is the product of unreliable methods. Docket No. 413. The instant memorandum [*3] sets forth the Court's analysis in granting the motion.

- 1 The Fresenius defendants include Fresenius Medical Care North America, Bio-Medical Applications of Texas, Inc., Fresenius Medical Care Holdings, Inc., and Larry Ramirez.
- 2 Defendant Alfonso Chavez (Chavez) indicated that he joined in Fresenius's motion through his "Notice of Defendant Alfonso Chavez M.D.'s Adoption of Fresenius' Motion to Bar Testimony of Dr. Jason Baldridge" (Docket No. 197), filed June 19, 2009.

I. FACTUAL AND PROCEDURAL BACKGROUND

A. Factual Background

The instant case is a qui tam action under the False Claims Act (FCA) brought by Rebecca Gonzalez, a former Fresenius employee. She alleges that Chavez, a nephrologist, 3 and Fresenius, a company that provides dialysis services to patients with end stage renal disease (ESRD), submitted false or fraudulent claims to Medicare. See generally Fourth Am. Compl. 2, 4-9. Chavez, under contract with Fresenius, served as the medical director and the attending physician at Fresenius's Cliffview and Gateway clinics, located in El Paso, Texas. Id. at 7; Chavez's Answer to Fourth Am. Compl. 5. Relator alleges that two non-physicians, Ramiro Devora and Arturo Orozco (the assistants), [*4] who worked for Chavez during different time periods relevant to this action (2000 to 2001, and 2005 to 2006, respectively), performed tasks and made patient-care decisions at Fresenius clinics in violation of applicable state and federal regulations. Fourth Am. Compl. 8, 97. Relator further alleges that, because Fresenius permitted the assistants to perform work for which Chavez was responsible, Chavez's referral of patients to Fresenius from his private practice for dialysis services violated federal anti-kickback laws. 4 Id. at 88-93.

- 3 A nephrologist is a medical specialist in nephrology, the branch of medical science concerned with medical diseases of the kidneys. STEDMAN'S MEDICAL DICTIONARY (27th ed. 2000).
- 4 Relator's claims pertaining to the alleged referral scheme and illegal kickbacks (under the Anti-Kickback Act and Stark Law) were the subject of successful Rule 50 motions at trial by Defendants Fresenius and Chavez. However, that

ruling is not relevant for purposes of the instant motion to exclude Relator's expert because the instant motion was made and ruled upon prior to the start of Relator's case at trial.

The alleged violations of state and federal regulations in the use [*5] of the assistants and the allegedly illegal referral scheme that resulted, as well as an alleged conspiracy to defraud, form the basis of Relator's FCA theories against Fresenius. ⁵ *Id.* at 97-106. In addition, Relator alleges that Chavez, through his medical practice Nephrology Associates, falsely billed Medicare for patient visits conducted by one of the assistants. *See id.* at 75, 94, 100.

5 Relator alleges that, by submitting claims for reimbursement to Medicare, Defendants falsely certified compliance with the applicable state and federal regulations. Fourth Am. Compl. 12.

Fresenius denies that the activities of the assistants violated applicable regulations, and argues that, in any case, its Medicare claims were not false because all of the services for which Fresenius billed were actually provided, and alleged non-compliance with state and federal regulations would not render their claims false for purposes of the FCA. See generally, Fresenius's Answer to Fourth Am. Compl. For his part, Chavez denies the allegation that he made false claims to Medicare through his practice, and asserts that he was indeed present during all patient visits for which he billed. See generally, Chavez's [*6] Answer to Fourth Am. Compl.; id. at 5. Chavez further argues that Devora and Orozco performed their work for him in accordance with law, and that his arrangement with Fresenius therefore did not violate anti-kickback laws. Id. Both Fresenius and Chavez deny that they conspired to defraud the Government in violation of the FCA. Fresenius's Answer to Fourth Am. Compl. 87; Chavez's Answer to Fourth Am. Compl. 37.

B. Procedural Background

On September 26, 2006, Relator filed a sealed complaint alleging violations of the False Claims Act against Fresenius and Chavez on behalf of the United States of America (the United States, or the Government). On November 21, 2007, the Government informed the Court that it had elected not to intervene in the action. Docket No. 20, Cause No. 06-CV-336-PRM. On December 19, 2007, the Court unsealed the case pursuant to Relator's motion. Docket No. 25, Cause No. 06-CV-336-PRM. Meanwhile, on July 10, 2007, Relator filed a sealed complaint against Fresenius, Chavez, and Larry Ramirez (Ramirez) alleging retaliation under the FCA and various state-law tort claims in a new cause of action. Docket No. 2, Cause No. 07-CV-247-PRM. On December 21,

2007, the Court [*7] unsealed the retaliation case pursuant to Relator's motion. Docket No. 13, Cause No. 07-CV-247-PRM. On January 24, 2008, the Court consolidated the two cases in this action, Cause No. 07-CV-247-PRM. ⁷ Docket No. 18, Cause No. 07-CV-247-PRM; Docket No. 31, Cause No. 06-CV-336-PRM.

- 6 Larry Ramirez has been the manager of the Cliffview clinic since 2006. *See* Relator's Fourth Am. Compl. 83.
- 7 Hereinafter, all citations to the docket refer to Cause No. 07-CV-247-PRM.

Fresenius (together with Ramirez) and Chavez each filed motions to dismiss. Docket Nos. 26, 36. On September 2, 2008, the Court dismissed Relator's common law claims of payment by mistake and unjust enrichment for lack of standing, and Relator's retaliation claims against Ramirez. "Order Granting in Part and Denying in Part Defendants' Motions to Dismiss," Docket No. 64. The parties continued to conduct extensive discovery and filed a number of discovery-related motions, which the Court referred to a United States Magistrate Judge for disposition.

On June 5, 2009, Relator submitted her Fourth Amended Complaint. Docket No. 180. Therein, Relator alleged, in short, that the manner in which Fresenius and Chavez provided dialysis services [*8] to their patients--by using two "unlicensed, un[-]credentialed" assistants--rendered their subsequent claims for reimbursement to Medicare "false" or "fraudulent" for purposes of the FCA under seven distinct theories or causes of action. On June 17, 2009 and June 19, 2009, respectively, Fresenius and Chavez filed substantively identical motions to dismiss portions of the Fourth Amended Complaint (Docket Nos. 188, 201), which the Court granted in their entirety on January 21, 2010 (Docket No. 408). In granting the motions to dismiss, the Court dismissed two causes of action and imposed a six-year statute of limitations. Docket No. 408. Thus, the following causes of action remained viable at the start of trial for the FCA case against Fresenius and Chavez:

- . Count 1: Knowingly presenting fraudulent or false claims in violation of the FCA, 31 § 3729(a)(1) (West 2008); 8
- . Count 2: Knowingly making a false record or statement in presentation of false claims in violation of the FCA, § 3729(a)(2);
- . Count 3: Presenting false claims for Medicare reimbursement for services

rendered in violation of the Stark Law, 42 U.S.C. § 1395nn;

- . Count 6: Presenting false claims for Medicare reimbursement [*9] for services rendered in violation of the Anti-Kickback Act, 42 U.S.C. §§ 1320a-7b(b);
- . Count 7: Conspiring to submit false claims in violation of the FCA, § 3729(a)(3).

The following causes of action remained viable at the start of trial for the retaliation case:

- . Count 1: Retaliation in violation of the FCA, § 3730(h) (as against Fresenius);
- . Count 2: Retaliatory constructive discharge in violation of the FCA, § 3730(h) (as against Fresenius);
- . Count 3: Intentional infliction of emotional distress (as against Fresenius and Ramirez).
- 8 Congress amended several subsections of the FCA in May, 2009 by enacting the Fraud Enforcement Recovery Act of 2009, Pub. L. No. 111-21, § 386, 123 Stat. 1617 (2009) [hereinafter FERA]. The Court has assessed the retroactivity provision of FERA and determined that the FERA amendments to subsection (a)(2) do not apply to this action. The Court will therefore continue to cite to the former version of the FCA (and its subsection designations).

Relator, Fresenius, and Chavez also each filed motions for summary judgment and related evidentiary motions pertaining to certain affidavits and expert testimony. *See* Docket Nos. 125, 155, 156, 161, 190, 224, 228, 257, [*10] & 320. Among these evidentiary motions was Fresenius's "Motion to Bar Testimony of Dr. Jason Baldridge" (Docket No. 190), which forms the subject of the instant memorandum.

Jason Baldridge is an assistant professor in the Department of Linguistics at the University of Texas in Austin, Texas. Relator's Resp. Ex. D at 1 (Baldridge C.V.). He conducts research and teaches in the area of computational linguistics, which he defines as "the application of models and methods from computer science to bear on the linguistic problem." Baldridge Dep. 9:9-11, May 29, 2009 (Fresenius's Mot. to Bar Test. Ex.

B; Relator's Resp. Ex. C) [hereinafter Baldridge Dep.]. Relator engaged Baldridge in this case to attribute authorship of a large number of medical records (transcribed dictations, for the most part) from Fresenius patient files by identifying and comparing linguist patterns in documents of known and unknown authorship using a computer model. *See* Relator's Resp. 2; Fresenius's Mot. to Bar Test. Ex. A at 1-2 (Expert Witness Report) [hereinafter Baldridge Report]. If permitted, Baldridge would testify that, for medical records produced during two relevant time periods (the Orozco era and the Devora [*11] era), he was able to distinguish distinct authorship patterns and attribute a large number of documents to the assistants Devora and Orozco. Baldridge Report at 25.

9 By "author," the parties and Baldridge are referring to the individual who wrote or spoke the words, and not necessarily the person who takes responsibility for the content (in the case of medical examinations or instructions, for example) or who provided the information contained in the document. See Baldridge Dep. 19:15-24:6.

In its motion, Fresenius does not argue that Baldridge lacks the requisite qualifications to render an expert opinion, but rather that his methods and conclusions are not sufficiently reliable to meet the standards of *Daubert v. Merrell Dow*, 509 U.S. at 597, and Federal Rule of Evidence 702. The Court agrees that Baldridge's testimony is inadmissible because it is unreliable and therefore granted Fresenius's motion to exclude the testimony of Jason Baldridge.

II. LEGAL STANDARD

Pursuant to Federal Rule of Evidence 702, a court may allow expert testimony if it finds (1) that the expert is testifying to scientific knowledge based upon sufficient facts or data, (2) that the testimony is the product of [*12] reliable principles and methods, and (3) that the witness has applied the principles and methods reliably to the facts of the case. FED. R. EVID. 702.

Guided by Rule 104(a), ¹⁰ courts must "make a 'preliminary assessment of whether the reasoning or methodology underlying the [expert's] testimony is scientifically valid and . . . can be applied to the facts in issue." *Pipitone v. Biomatrix, Inc.*, 288 F.3d 239, 243-44 (5th Cir. 2002) (quoting *Daubert*, 509 U.S. at 592-93). In undertaking this preliminary assessment, a trial court, relying primarily on *Daubert*, must determine (1) whether the expert testimony assists the trier of fact and (2) whether the proffered testimony is "supported by appropriate validation" and grounded "in the methods and procedures of science." 509 U.S. at 590-92.

10 Federal Rule of Evidence 104(a) provides that "[p]reliminary questions concerning the qualification of a person to be a witness, the existence of a privilege, or the admissibility of evidence shall be determined by the court, subject to the provisions of subdivision (b)." FED. R. EV-ID. 104(a).

Despite the panoply of reliability factors enumerated in Daubert, 11 trial courts can determine "whether Daubert's [*13] specific factors are, or are not, reasonable measures of reliability in a particular case." Kumho Tire Co. v. Carmichael, 526 U.S. 137, 153, 119 S. Ct. 1167, 143 L. Ed. 2d 238 (1999). "Whether Daubert's suggested indicia of reliability apply to any given testimony depends on the nature of the issue at hand, the witness's particular expertise, and the subject of the testimony." Seatrax, Inc. v. Sonbeck Int'l, Inc., 200 F.3d 358, 372 (5th Cir. 2000); accord Kumho Tire, 526 U.S. at 153. In essence, a trial court's salient objective, as here, is to make certain that an expert, "whether basing testimony on professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field." Kumho Tire, 526 U.S. at 152.

11 Daubert suggested a nonexclusive, flexible test for district courts to consider when examining an expert's scientific reliability, including (1) whether the theory or technique the expert employs is generally accepted; (2) whether the theory has been subjected to peer review and publication; (3) whether the theory can and has been tested; (4) whether the known or potential rate of error is acceptable; and (5) whether [*14] there are standards controlling the technique's operation. 509 U.S. at 593.

In evaluating the reliability of proffered expert testimony, courts must look beyond mere credentials and must make sure that there is an adequate "fit" between the data and the opinion. See, e.g., Moore v. Ashland Chem., Inc., 151 F.3d 269, 276 (5th Cir. 1998). Ultimately, "[t]he proponent need not prove to the judge that the expert's testimony is correct, but [he] must prove by a preponderance of the evidence that the testimony is reliable." Curtis v. M&S Petroleum, Inc., 174 F.3d 661, 668 (5th Cir. 1998); accord Bocanegra v. Vicmar Servs., 320 F.3d 581, 585 (5th Cir. 2003) (citing *Moore*, 151 F.3d at 276). "The reliability analysis applies to all aspects of an expert's testimony: the methodology, the facts underlying the expert's opinion, [and] the link between the facts and the conclusion." Knight v. Kirby Inland Marine, Inc., 482 F.3d 347, 355 (5th Cir. 2007) (quoting Heller v. Shaw Indus., Inc., 167 F.3d 146, 155 (3d Cir. 1999)). With this legal standard in mind, the Court now turns to a specific examination of Relator's proffered expert, Jason Baldridge.

III. ANALYSIS

A. Baldridge's Method of Author Identification

In [*15] his expert report, Baldridge explains how he arrived at his conclusions regarding the authorship of certain medical records in this case. Basically, counsel for Relator provided Baldridge with various sets of documents in PDF format: a set of transcriptions known to have been dictated by Orozco, a set of documents of unknown authorship from the period in which Orozco worked in the clinics ("Period 1"), a set of transcriptions known to have been dictated by Devora, and a set of documents of unknown authorship from the period in which Devora worked in the clinics ("Period 2"). Baldridge Report at 1. Baldrige was told, and asked to assume, among other things, that Chavez was "active at the clinic during both periods and had likely authored some of the documents in the questioned sets." *Id.* at 2.

Baldridge decided to use the "Bernoulli mixture model," a "probabilistic mixture model," to identify each document with a "cluster" (later associated with a particular author) based on the combined probabilities of the presence or absence of a series of linguistic features. Id. at 4-5. To develop this model, Baldridge first identified thirty (30) linguistic features for the Orozco-authored documents [*16] from Period 1, and twenty-five (25) features for the Devora-authored documents in Period 2. Id. at 6. Each linguistic feature is either a single word, a multi-word pattern, or a "referring expression" (that is, the manner in which the dictator referred to himself or the patient). Id. at 7. Next, Baldridge's assistant "annotated" the unknown documents by examining each of the documents to determine whether or not each feature was or was not present. Id. at 6-7.

Baldridge then wrote code for a computer program to develop the model. Id. at 10. He relied on established mathematical equations that represent the probabilities for a given feature and the combination of those probabilities for an overall distribution. Id. at 5-6. To apply the program to a set of data, one must supply an assumption about the total number of authors represented. Id. at 14-15; Baldridge Dep. 166:16-167:22. Baldridge applied the program to the data using different numbers of authors to determine which produced the most "well separated clusters." See Baldridge Report at 14, 15. After establishing "the broad patterns of authorship" in this way, Baldridge conducted a "manual review" of the documents to make final [*17] determinations of authorship based on authorship patterns he identified. *Id.* at 10-11. Finally, Baldridge used the clusters resulting from the Bernoulli mixture model to "train" a "supervised model" which he applied to the documents of known authorship to confirm that the model would correctly assign them to the same cluster. *Id.* at 11-12.

B. Baldridge's Methods Are Unreliable

Baldridge's testimony is inadmissible at trial because his conclusions are the product of unreliable principles and methods, as evidenced by the following considerations.

1. Number of Authors

The Bernoulli mixture model that Baldridge employed in determining authorship of medical records required him to create models for, or test different numbers of, authors to determine the number of authors that best "fit" the data. Baldridge Dep. 166:16-167:22. For Period 2 (Devora), "[t]he assumption of two mixture components (= two authors) produced a model that found two very well separated clusters." Baldridge Report at 14. Thus, Baldridge felt confident that exactly two authors were represented in the documents for Period 2. *Id.*

For Period 1, however, the separation was not as clear given an assumption of two authors, and [*18] further models led Baldridge to find support for "at least three authors," and he could "not reject the possibility that there is a fourth or even a fifth author." *Id.* at 15, 17. Yet Baldridge decided there were probably three authors, ¹² *id.* at. 22, and the documents that some models had placed in additional clusters (beyond three) were assigned to Orozco, Baldridge Dep. 84:14-21.

12 When running the model for three authors, Baldridge noticed that the "separation [of clusters] improved significantly." Baldridge Report at 15. Baldridge identified and annotated five more linguistic features for that period, ran the model again with three authors, and then sorted and reviewed the documents in the third cluster. *Id.* He noticed that several of those documents bore the signature of Cesar Maldonado, MD, and he then used those documents to create standards for this third author. *Id.*

This uncertainty regarding an essential piece of information to the conclusions on authorship--the total number of authors represented in Period 1--alone demonstrates that the opinion is unreliable. Further, Baldridge acknowledged that he later learned from counsel for Relator that there was indeed a fourth author, [*19] Robert Metzer, and his conclusions were therefore in error. Baldridge Dep. 115:23-116:1. Baldridge didn't recognize the presence of a fourth author, as he did with the third author, because "it clustered similarly to Orosco [sic] and I didn't have -- I wasn't either given time or

didn't have time to see if there were further ways of teasing apart those patterns." Baldridge Dep. 115:3-6.

As stated above, Baldridge assigned the Metzer documents to the Orozco authorship cluster:

Q: . . . [Y]ou find strong evidence that "the identity of the author of the first cluster is the same as that identified to me as the author of the knowns, i.e. Orosco [sic]," right?

A: Yes.

Q: Now, we can also agree now, can't we, that that cluster also includes dictations by RM [Robert Metzer]?

A: Yes.

Baldridge Dep. 84:14-21.

Furthermore, Baldridge does not know how many documents were incorrectly attributed to Orozco in his final conclusions, and he would consider formulating a revised hypothesis:

Q: Do you know how many documents that you have attributed to Orosco [sic] would need to be moved to Metzer?

A: I can't determine that right now.

* * *

Q: Do you have an opinion to a reasonable degree of professional certainty [*20] of the number of documents that were attributed to Orosco [sic] that were done by Metzer?

A: No.

* * *

A: Given that the RM [Robert Metzer] documents were produced by a person with those initials, then absolutely I would consider that in formulating a revised hypothesis about these.

Baldridge Dep. 116:8-10; 117:20-24; 175:18-21.

The error regarding the number of authors for Period 1, which Baldridge acknowledged in his deposition and which affected an unknown number of authorship determinations, fatally undermines the reliability of the expert's method and his conclusions on authorship.

2. Transcription

There are additional reasons to find Baldridge's methodology unreliable under the standards set by *Daubert* and Rule 702. One pertains to the fact that many of the documents to which Baldridge assigned authorship were transcriptions of dictations.

Baldridge's method for identifying authorship, the Bernoulli mixture model, uses "binary" data representing the presence or absence of certain linguistic features. Baldridge Report at 5. The linguistic features chosen as the basis for the data were, for the most part, precise words or combinations of words. *Id.* at 7. However, the majority of documents [*21] Baldridge analyzed for authorship were transcriptions of audio recordings, produced by one of several transcriptionists who listened to the voice recording of the "author" and typed his words into a document.

Baldridge's method relies heavily, if not entirely, on word selection, and yet it does not account for the role of the transcriptionist. Baldridge acknowledges that "some of the linguistic patterns could be due to differences between transcriptionists rather than to different authors of the dictations." Baldridge Report at 23. In an attempt to address the problem, Baldridge checked that no strong correlation existed between any single transcriptionist and any single author. *Id.* at 23-24. He acknowledged that transcriptionists might have adjusted words in the documents, but chose to correct the issue by isolating it from his data:

Q: Do you know whether the transcriptionists did an accurate job of transcribing what was said into the Dictaphone?

A: Well, I was assuming they did. I did note that there were a few minor differences. In the Devora period there was one transcriptionist who would write down "the outpatient basis." Another would do it "an outpatient basis." I assumed that [*22] Devora probably said the same thing each time and that the transcriptionist -- one perhaps corrected a form that Devora used. So I attempted in my features that I extracted not to be tuned into linguistic indicators that were -- had that kind of indication with them.

Baldridge Dep. 26:12-24.

Yet Baldridge did not listen to any audio files to identify the effects of the transcription process on his data and conclusions:

Q: And in your work in this case did you attempt to evaluate whether the transcriptionists accurately took down what the person said into the Dictaphone?

A: No, I didn't.

Q: And that could have been done, though, at least for . . . some of the Devora documents, right? . . . [D]id you have recordings?

A: No.

* * *

Q: Did you listen to any recorded progress notes or other . . . documents?

A: No.

Baldridge Dep. 24:21-25:6; 25:11-13.

Given the method's reliance on word choice, the evidence that words were altered (by someone other than the author) during the transcription phase means that the method does not adequately "fit" the opinion offered on authorship. The failure to assess and account for the potentially serious effects of transcription of the documents reveals unreliability [*23] in the data and methodology utilized by Baldridge to reach his expert conclusions.

3. General Acceptance

The Relator has not identified, nor has the Court located, any case in which a court has accepted the expert testimony of a computational linguist for purposes of author identification. Furthermore, Baldridge is not aware of anyone applying computational linguistic techniques to text classification problems in a court case. Baldridge Dep. 87:5-8. One court that addressed the admissibility of expert testimony on author identification in depth found it unreliable due to the limitations of the scientific field.

In *United States v. Van Wyk*, the court analyzed proffered expert testimony in the area of "forensic stylistics" or text analysis, where, "[i]n cases of disputed authorship, the linguist analyzes and describes the style of writing of a document of questioned authorship and compares and contrasts its language to that of documents known to be written by a given author." 83 F. Supp. 2d 515, 517 (D.N.J. 2000) (internal citations omitted). The court considered whether to permit an FBI agent to testify regarding the authorship of certain documents. It excluded the testimony, finding it [*24] unreliable under *Daubert*:

The reliability of text analysis . . . is questionable because . . . there is no known rate of error, no recognized standard, no meaningful peer review, and no system of accrediting an individual as an expert in the field. Consequently, the existing data for forensic stylistics cannot definitively establish . . . that a particular person is "the" author of a particular writing.

Id. at 523.

Moreover, experts in the field of computational linguistics agree that their methods have not attained the level of rigor and general acceptance required under *Daubert*. In an article that Baldridge cites in his report and recognizes as authoritative, "Author Identification in American Courts," Baldridge Report at 2, 26, the authors examine different methods in the field and conclude that there are "serious questions about the admissibility of expert testimony on authorship, given the current state of the art" and that linguistic methods require further testing and improvement to be accepted in American courts. Lawrence Solan & Peter Tiersma, *Author Identification in American Courts*, 25 APPLIED LINGUISTICS 448, 463 (2004).

It is also important to note that Baldridge did not use [*25] the most rigorous methods available in the field of computational linguistics to generate his conclusions in this case. He would have preferred to use the more sophisticated "Gaussian mixture model," but the documents were in PDF rather than ASCII format, so he was unable to do so. Baldridge Report at 4-5; Baldridge Dep. 104:17-24. Due to the document formatting, the fact that documents were transcribed, and time constraints, Baldridge did not consider other linguistic elements often evaluated by experts in his field, such as punctuation, misspellings, actual word frequencies, syntax, and sentence length. Baldridge Dep. 95:19-98:1. Baldridge admits that, with additional time and the ability to convert the format of documents, he would have liked to incorporate several of these linguistic properties into his analysis. Baldridge Dep. 98:5-99:20.

Thus, the Court concludes that the methods of computational linguistics generally and the simplified method selected by Baldridge in particular have not gained general acceptance.

4. Error Rate

There is ample evidence that errors may have and did occur in Baldridge's analysis, *see supra* Section III.B.1, yet he does not supply an error rate for the

[*26] Court's consideration in his report or deposition aside from the incredible assertion that his model's classifications of authorship are 100 percent accurate:

Q: Are you testifying in this case that you're 100 percent accurate about all the classifications that you've done?

A: Yes.

Baldridge Dep. 123:9-12. While the lack of known error rate alone may be insufficient to find an expert's method unreliable, *see, e.g., United States v. Norris*, 217 F.3d 262, 269-71 (5th Cir.2000), it is one of several factors in this case that contribute to the Court's determination of unreliability under the *Daubert* standard.

5. Verification

Baldridge did not verify his methodology or conclusions in the manner he ordinarily would in the course of his academic work but instead relied on counsel for Relator to test his conclusions against other available data. Baldridge Dep. 33:17-38:14. Most importantly, Baldridge did not use a "test set," which is "typical []" in academic work in this area. Baldridge Dep. 35:9-13. He did not perform any of a number of other possible methods of checking the accuracy of his work. Baldridge Dep. 33:17-38:14.

C. Fraud Testimony

Baldridge's report contains a brief section in which [*27] he points out the infrequent use of first person pronouns in documents ascribed to Devora and Orozco, and then states that "social psychologists have found that people use first person pronouns less frequently when they are lying." Baldridge Report at 21-22. He further states that documents ascribed to Chavez include more "cognitively complex words" which are more often used when people are telling the truth. *Id.* Fresenius argues

that Baldridge's tesitmony on this issue is inadmissible. Fresenius's Mot. to Bar Test. 9-10. The Court agrees.

Baldridge has not elaborated any method upon which these conclusions were based, aside from a citation to a scholarly work. Further, Baldridge accepts that he is not an expert in social psychology:

- Q: Are you prepared yourself to render an opinion that Orosco [sic] and Devora chose particular words subconsciously because they were attempting to distance themselves from a lie?
- A: I would recommend calling an expert witness like James Pennebaker.
- Q: Because you are not an expert witness on that?
- A: I am not an expert witness in social psychology.

Baldridge Dep. 186:18-187:1. Therefore, Baldridge's testimony on whether certain individuals were lying [*28] or telling the truth based on word choices in documents they authored (according to authorship identification methods the Court finds unreliable) is inadmissible.

IV. CONCLUSION

Accordingly, and for the reasons elaborated above, the Court granted Fresenius' "Motion to Bar Testimony of Dr. Jason Baldridge" (Docket No. 190) by way of oral and written orders on January 29, 2010 and February 8, 2010, respectively.

SIGNED this day of March, 2010.

PHILIP R. MARTINEZ

UNITED STATES DISTRICT JUDGE