

University of New Mexico

## UNM Digital Repository

---

Family Community Medicine

HSC Data Community

---

2023

### UNM-OCH Diabetes Project Data Set

Janet Page-Reeves

*University of New Mexico, Family & Community Medicine, [jpage-reeves@salud.unm.edu](mailto:jpage-reeves@salud.unm.edu)*

Lidia Regino

*University of New Mexico Health Sciences Center, Albuquerque, USA*

Erik B. Erhardt

*University of New Mexico Health Sciences Center, Albuquerque, USA*

Cristina Murray-Krezan

*Department of Internal Medicine, Division of Epidemiology, Biostatistics, and Preventive Medicine, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA*

Blanca Pedigo

*See next page for additional authors*

Follow this and additional works at: [https://digitalrepository.unm.edu/hsc\\_data\\_othr](https://digitalrepository.unm.edu/hsc_data_othr)

---

#### Recommended Citation

Page-Reeves, Janet; Lidia Regino; Erik B. Erhardt; Cristina Murray-Krezan; Blanca Pedigo; Maria Tellez; Mark R. Burge; Shiraz I. Mishra; Elaine Bearer; Molly J. Bleecker; Denise Guerrero; and Daniel P. Rodriguez. "UNM-OCH Diabetes Project Data Set." (2023). doi:<https://doi.org/10.25827/MTPV-KG27>.

This Dataset is brought to you for free and open access by the HSC Data Community at UNM Digital Repository. It has been accepted for inclusion in Family Community Medicine by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).

---

**Authors**

Janet Page-Reeves, Lidia Regino, Erik B. Erhardt, Cristina Murray-Krezan, Blanca Pedigo, Maria Tellez, Mark R. Burge, Shiraz I. Mishra, Elaine Bearer, Molly J. Bleecker, Denise Guerrero, and Daniel P. Rodriguez

# UNM PCORI Diabetes Project (PDP) Data Introduction

Project information, data access, data formatting, codebooks

Erik Erhardt

2022-12-25 11:10:29

## Contents

<b>1 Project information</b>	<b>1</b>
<b>2 Data files and codebooks</b>	<b>2</b>
<b>3 Data</b>	<b>2</b>
3.1 Overview of formatting the data for analysis . . . . .	2
3.2 Data de-identification . . . . .	4
<b>Publications (incomplete)</b>	<b>4</b>

## 1 Project information

- Overview
  - Principal Investigator: Janet Page-Reeves, MA, PhD
  - Organization: University of New Mexico Health Sciences Center
  - Project Budget: \$2,276,443
  - Funding agency: Patient-Centered Outcomes Research Institute (PCORI)
- Study descriptions
  - PCORI
  - ClinicalTrials.gov
- Study Design
  - Study Type: Observational
  - Estimated Enrollment: 240 participants
  - Observational Model: Cohort
  - Time Perspective: Prospective
  - Official Title: A Patient-Centered Framework to Test the Comparative Effectiveness of Culturally and Contextually Appropriate Program Options for Latinos With Diabetes From Low-Income Households
  - Actual Study Start Date: February 1, 2017
  - Estimated Primary Completion Date: March 30, 2021
  - Estimated Study Completion Date: March 30, 2021
  - Actual dates:
    - \* Start September 13, 2016 (first pre-screening)
    - \* End August 15, 2020 (last follow-up collection)
  - UNM HSC IRB 16-303

## 2 Data files and codebooks

- Data introduction (this file)
  - `data_introduction.html` or `data_introduction.pdf`
- Codebooks
  - REDCap codebook: `codebook_REDCap_pdp.pdf`
    - \* REDCap codebook includes the data collection questions and conditional logic.
  - Data-summary codebook: `codebook_dat_pdp_PUBLIC.pdf`
    - \* Data codebook summarizes the data for each question.
      - Number of observations.
      - Number of variables.
      - Summary of each variable's label, variable name, data class, number of unique values, and percent missing.
      - For each variable, a more detailed summary:
        - Numeric variables: number of unique levels, five-number summary, and histogram.
        - Factor variables: number of unique levels, list of observed levels, and barplot.
- Data formats
  - CSV files factor variables values are labels (not levels).
  - RData files are R formatted data including factor labels and variable labels.
- PDP Data
  - Step 3
  - Includes all data for analysis
  - CSV file: `dat_pdp_clean_PUBLIC.csv`
    - \* Load with `dat_pdp <- read.csv(file = "dat_pdp_clean_PUBLIC.RData")`
  - RData file: `dat_pdp_clean_PUBLIC.RData`
    - \* Load with `load(file = "dat_pdp_clean_PUBLIC.RData")`
- Eligibility Data (probably not useful)
  - Step 2
  - CSV file: `dat_Elig_clean_PUBLIC.csv`
    - \* Load with `dat_Elig <- read.csv(file = "dat_Elig_clean_PUBLIC.RData")`
  - RData file: `dat_Elig_clean_PUBLIC.RData`
    - \* Load with `load(file = "dat_Elig_clean_PUBLIC.RData")`
- Pre-Screening Data (probably not useful)
  - Step 1
  - CSV file: `dat_PreScr_clean_PUBLIC.csv`
    - \* Load with `dat_PreScr <- read.csv(file = "dat_PreScr_clean_PUBLIC.RData")`
  - RData file: `dat_PreScr_clean_PUBLIC.RData`
    - \* Load with `load(file = "dat_PreScr_clean_PUBLIC.RData")`

## 3 Data

### 3.1 Overview of formatting the data for analysis

1. Read and format data
  1. Read current files
  2. Replace unicode to appropriate ascii characters
  3. Split Spanish / English options and keep English (code also allows to keep only Spanish)
  4. Save data
2. Data cleaning checks of numeric and categorical variables for data cleaning
  1. Numeric: Print non-numeric as well as min and max 5 values for each variable; plot histogram.
3. Longitudinal plots showed differences from the first observation to show unusual differences from baseline
  1. Some BMI unusual variance was corrected due to recording errors
4. Additional formatting
  1. Pre-screening survey

1. Converting 3 date/time variables to POSIXct format.
2. Eligibility survey
  1. Converting 4 date/time variables to POSIXct format.
  2. Removing special characters from income dollar formats (\$ and ,).
3. PDP surveys
  1. Exclude any “database development” no participant records
  2. Converting 9 date/time variables to POSIXct format.
  3. Separate participant ID from participant/social support label.
  4. Calculate new variables.
    1. Age from interview date and DOB
    2. Height median over time to exclude unusual values
    3. BMI from average of two height and average of two weight measurements at each time
    4. A1c numeric, any values with label **greater than 14.0** are assigned 14.1
  5. Collapse categories for variables with too many levels.
    1. Education from 14 to 6 categories.
    2. Marital status 6 to 2 categories.
    3. Language: how well speak English 5 to 3 categories.
    4. General health 5 to 4 categories.
    5. Household number, 12 to 6 categories.
    6. Patient to SS relationship, 6 to 5 categories.
    7. Binary Sex variable (Other to Male given bio of person).
  6. Zip code a factor variable.
  7. Determining language from combination of questions: Spanish, English, Other.
  8. Copy baseline responses to followup records for selected demographic variables.
5. Dates: Print non-numeric as well as min and max 5 values for each date; plot histogram.
6. Data Validation for every variable, range and admissible value checks, including conditional relationships between variables. Visualization of errors. All errors reviewed by data quality team.
7. Comorbidities are cleaned, categorized, and counted.
8. Annual Income is calculated from typical income dollar amount and income frequency, followed by a data cleaning check.
9. A numeric household number is determined, used later for poverty ratio determination.
10. Time since diagnosis in years is calculated as the difference from baseline date and the estimated diagnosis date.
11. All five possible “missing” categories for factor variables are set to NA.
12. A1c
  1. If both One Hope and UNM CDE measurements, then use UNM CDE.
  2. A1c categories assigned based on A1c: Neither, Pre-diabetes, or Diabetes
  3. Calculate A1c days before baseline, if reporting a value not drawn as part of study.
13. Smoking years as numeric.
14. Survey instruments
  1. Depression: PHQ-9 score categories and log2 score.
  2. Diabetes Knowledge Questionnaire DKQ-23: scoring did not include Question 2 about “effective insulin” or the added Question 25 about drinking sodas.
  3. Patient Activation Measure PAM-10: scoring via query to [insigniahealth.com](http://insigniahealth.com).
  4. CAHPS scoring: CAHPS instructions indicate that the researcher can design their own composite measures.
    1. For each question, we assigned values from -1 through +1 to the “negative” through “positive” ordinal responses.
    2. We decided to follow the survey’s five domains of cultural competence organized from 8 domains.
    3. Our questions allowed us to create the following scales:
      1. A&B. Doctor Communication-Behaviors (8 patient and 1 social support questions)
      2. F. Equitable treatment (3 patient and 2 social support questions)
      3. G. Trust (6 patient and 1 social support questions)

4. H. Access to Interpreter Services (5 patient and 1 social support questions)
4. Because not all questions were answered by all participants, each score is the ratio of the sum of their responses divided by the number of questions answered in each domain for each participant-social support pair.
  1. Scores are bounded between -1 and +1 with the interpretations of “negative” and “positive”, with 0 indicating “neutral”.
15. Federal Poverty level ratio: This ratio is the patient’s annual income divided by the poverty level based on each patient’s household size, cf. “Annual Update of the HHS Poverty Guidelines, A Notice by the Health and Human Services Department” on 01/25/2016, Document Citation: 81 FR 4036, Page: 4036-4037 (2 pages), Document Number: 2016-01450, <https://www.federalregister.gov/d/2016-01450>.
  1. A 1 means thier estimated annual income is equal to the poverty level for their family’s size, 2.5 means they make 2.5 times the poverty level.

### 3.2 Data de-identification

1. Data is de-identified for publication following the De-identification Standard via HIPAA Privacy Rule Section 164.514(a).
  1. Remove identifying columns.
  2. Zip codes truncated to use only the first 3 digits of zip codes, and no restricted ZIP code tabulation areas (ZCTAs) were in our data (otherwise, they would be set to “000”).
  3. Calculate durations in days between all dates relative to the date of the Patient Baseline assessment.

### Publications (incomplete)

- Bracht, Stephen, Krista Salazar, and Janet Page-Reeves. n.d. “Evaluating the Effectiveness of a Culturally Situated Type 2 Diabetes Education Program Serving Latino Adults (Unpublished Manuscript).”
- Janet Page-Reeves, Amy Anixter Scott, Maurice Moffett, Veronica Apodaca, and Vanessa Apodaca. 2014. “‘Is Always That Sense of Wanting . . . Never Really Being Satisfied’: Women’s Quotidian Struggles With Food Insecurity in a Hispanic Community in New Mexico.” *Journal of Hunger & Environmental Nutrition* 9 (2): 183–209.
- McGrew, Hannah Cole, Lidia Regino, Molly Bleecker, Maria Tellez, Blanca Pedigo, Denisse Guerrero, Virginia Sandoval, Loida Varela, and Janet Page-Reeves. 2020. “Training Patient Stakeholders Builds Community Capacity, Enhances Patient Engagement in Research.” *Journal of Community Engagement and Scholarship* 13 (1): 10.
- Mishra, Shiraz, Janet Page-Reeves, and Lidia Regino. 2012. “Results from a CTSC-Funded Planning Project to Develop a Diabetes Prevention Initiative with East Central Ministries.” Community {Report}.
- Page-Reeves, Janet. 2014a. “Conceptualizing Food Insecurity and Women’s Agency: A Synthetic Introduction.” In *Women Redefining the Experience of Food Insecurity: Life Off the Edge of the Table*, edited by Janet Page-Reeves, 3–44. Lanham, Maryland: Lexington Books of Rowman & Littlefield.
- . 2014b. “Conceptualizing Food Insecurity and Women’s Agency: A Synthetic Introduction.” *Women Redefining the Experience of Food Insecurity: Life Off the Edge of the Table*. Lanham, MD: Lexington Books.
- , ed. 2014c. “‘I Took the Lemons and I Made Lemonade’: Women’s Quotidian Strategies and the Re-Contouring of Food Insecurity in a Hispanic Community in New Mexico.” In *Women Redefining the Experience of Food Insecurity: Life Off the Edge of the Table*, p. 85–104. Lexington Books of Rowman & Littlefield.
- . 2014d. *Women Redefining the Experience of Food Insecurity: Life Off the Edge of the Table*. Lexington Books.
- Page-Reeves, Janet, Molly Bleecker, Maurice L. Moffett, Katharine Linder, Jeannie Romero, and Carol Krause. 2017. “Situating Food Insecurity in a Historic Albuquerque Community: The Whorled Relationship Between Health and Place.” *Journal of Health Disparities Research & Practice* In Press.
- Page-Reeves, Janet, and Enrique Cardiel. 2016. “GED Privatization as a Social Determinant of Health.” *Human Organization* 75 (1): 21–32. <https://doi.org/10.17730/0018-7259-75.1.21>.
- Page-Reeves, Janet, Ananda Marin, Molly Bleecker, Maurice Leon Moffett, Kathy DeerInWater, Sarah

- EchoHawk, and Douglas Medin. 2017. "From Community Data to Research Archive: Partnering to Increase and Sustain Capacity Within a Native Organization." *Gateways: International Journal of Community Research and Engagement* 10 (0): 283–97. <https://doi.org/10.5130/ijcre.v10i1.4947>.
- Page-Reeves, Janet, Ananda Marin, Maurice L. Moffett, and Kathy DeerInWater. 2017. "Wayfinding as a Concept for Understanding Success Among Native Americans in STEM: 'Learning How to Map Through Life'." *Cultural Studies of Science Education* In Press.
- Page-Reeves, Janet, Shiraz I. Mishra, Joshua Niforatos, Lidia Regino, and Robert Bulten. 2013. "An Integrated Approach to Diabetes Prevention: Anthropology, Public Health, and Community Engagement." *Qualitative Report (Online)* 18: 1–22.
- Page-Reeves, Janet, Maurice L. Moffett, Leah Steimel, and Daryl T. Smith. 2017. "The Evolution of an Innovative Community-Engaged Health Navigator Program to Address Social Determinants of Health." *Progress in Community Health Partnerships: Research, Education, and Action* 10 (4): 603–10. <https://doi.org/10.1353/cpr.2016.0069>.
- Page-Reeves, Janet, Maurice Moffett, Molly Bleecker, Katharine Linder, Jeannie Romero, and Carol Krause. 2018. "Situating Food Insecurity in a Historic Albuquerque Community: The Whorled Relationship Between Food Insecurity and Place." *Journal of Health Disparities Research and Practice* 10 (4). <https://digitalscholarship.unlv.edu/jhdrp/vol10/iss4/7>.
- Page-Reeves, Janet, Cristina Murray-Krezan, Lidia Regino, Jackie Perez, Molly Bleecker, Daniel Perez, Bill Wagner, Susan Tigert, Elaine L Bearer, and Cathleen E Willging. 2021. "A Randomized Control Trial to Test a Peer Support Group Approach for Reducing Social Isolation and Depression Among Female Mexican Immigrants." *BMC Public Health* 21 (1): 1–18.
- Page-Reeves, Janet, Joshua Niforatos, Shiraz Mishra, Lidia Regino, Andrew Gingrich, and Robert Bulten. 2013. "Health Disparity and Structural Violence: How Fear Undermines Health Among Immigrants at Risk for Diabetes." *Journal of Health Disparities Research and Practice* 6 (2): 30–47.
- Page-Reeves, Janet, and Lidia Regino. 2018a. "Community–University Health Research Partnerships: Challenges and Concrete, Plain Language Strategies for Community Engagement in Research." *Anthropology in Action* 25 (2): 1–12.
- . 2018b. "Current Challenges of Community-University Health Research Partnerships and Concrete, Plain Language Strategies to Building Capacity." *Anthropology in Action* In Press.
- Page-Reeves, Janet, Lidia Regino, Hannah Cole McGrew, Maria Tellez, Blanca Pedigo, Amy Overby, Abigail Cunningham, Susan Tigert, and Mark Burge. 2017. "Collaboration and Outside-the-Box Thinking to Overcome Training-Related Challenges for Including Patient Stakeholders as Data Collectors in a Patient-Engaged Research Project." *Journal of Patient Experience* 5 (2): 88–91. <https://doi.org/10.1177/2374373517729506>.
- Page-Reeves, Janet, Lidia Regino, and Shiraz Mishra. 2012. "Culturally-Situating Diabetes Prevention in an Immigrant Community in Albuquerque, NM: The Importance of Culture." Albuquerque, NM.
- Page-Reeves, Janet, Lidia Regino, Cristina Murray-Krezan, Molly Bleecker, Erik Erhardt, Mark Burge, Elaine Bearer, and Shiraz Mishra. 2017. "A Comparative Effectiveness Study of Two Culturally Competent Models of Diabetes Self-Management Programming for Latinos from Low-Income Households." *BMC Endocrine Disorders* 17 (1): 46. <https://doi.org/10.1186/s12902-017-0192-4>.
- Page-Reeves, Janet, Lidia Regino, Maria Tellez, Blanca Pedigo, and Esperanza Perez. 2018. "Engaging Latino Patients in Diabetes Research: What We Are Learning." *Practicing Anthropology* 40 (3): 35–39. <https://doi.org/10.17730/0888-4552.40.3.35>.
- Page-Reeves, Janet, Sarah Shrum, Felicia Rohan-Minjares, Tamara Theideman, Jackie Perez, Carla Cordova, Ambrosia Murrietta, and Francisco Ronquillo. n.d. "Addressing Syndemic Health Disparities Among Latina Immigrants Using an Innovative Peer Support Group Approach." *Journal of Racial and Ethnic Health Disparities* In Press.
- Page-Reeves, Janet, Sarah Shrum, Felicia Rohan-Minjares, Tamara Theideman, Jackie Perez, Carla Cordova, Ambrosia Murrietta, Francisco Ronquillo, and Helene Silverblatt. n.d. "The Syndemics of Social Isolation, Depression, Diabetes, Food Insecurity, and Immigration: Innovating a Model of Peer Support with Structured Dialogue for Latina Immigrants in Albuquerque, New Mexico."
- Page-Reeves, Janet, Sarah Shrum, Felisha Rohan-Minjares, Tamara Thiedeman, Jackie Perez, Ambrosia Murrietta, Carla Cordova, and Francisco Ronquillo. 2019. "Addressing Syndemic Health Disparities Among Latin Immigrants Using Peer Support." *Journal of Racial and Ethnic Health Disparities* 6 (2):

380–92. <https://doi.org/10.1007/s40615-018-00535-y>.  
Tlapoyawa, Kurly. 2020. “Explaining the Concept of Randomization to Spanish-Speaking Potential Clinical Trial Participants: A Short Video and Video User Guide.” <https://www.youtube.com/watch?v=layi60vSQNQ>.