# Finding Equity: Utilizing Artificial Intelligence to Identify Social and Infrastructural Predictors of Diabetes Mortality in Florida

Cistola, AS1

1Department of Health Services Research, Management, and Policy

University of Florida

# Finding Equity: Utilizing Artificial Intelligence to Identify Social and Infrastructural Predictors of Diabetes Mortality in Florida

## Abstract

DM outcomes represent one of the largest avoidable cost burdens with opportunity for improvement in the U.S. health care system. Improving health equity in the context of DM will require targeted community improvements, infrastructure investments, and policy interventions that are designed to maximize the impact of resource allocation through the use of available data and computational resources. By using an Artificial Intelligence approach to evaluate over 2000 socio-economic and infrastructural predictors of DM mortality, this study used a specific series of modeling techniques to identify significant predictors without human selection and compare their predictive ability with all possible factors when passed through artificial neural networks. The final regression model using zip code and county level predictors had an R2 of 0.863. Significant predictors included: Population % White, Population % Householders, Population % Spanish spoken at home, Population % Divorced males, Population % With public health insurance coverage, Population % Employed with private health insurance coverage, Manufacturing-Dependent Designation, Low Education Designation, Population % Medicare Part A & B Female Beneficiaries, Number of Short Term General Hospitals with 50-99 Beds. Using a multi-layered perceptron to predict zip codes at risk the C-statistic for all 2000+ predictors was 0.7938 while the 13 selected predictors was 0.8232. This indicates that these factors are highly relevant for DM mortality in Florida. This process was completed without the need of human variable selection and indicates how AI can be used for informative precision public health analyses for targeted population health management efforts.

# Finding Equity: Utilizing Artificial Intelligence to Identify Social and Infrastructural Predictors of Diabetes Mortality in Florida

In 2018, 34.2 million people (13.0% of the U.S. Population) were estimated to have DM in the U.S., including 26.9 million with diagnosed DM and 7.3 million with undiagnosed DM. Furthermore, another 88 million people (27.0% of the U.S. population) were estimated to have PD. This represents a significant increase since 2003, with most of the increases represented in racial and ethnic minority populations (Centers for Disease Control and Prevention, 2020). In 2016, there were 16 million Emergency Department (ED) visits, 7.8 million hospitalizations with or for DM, and accounted for approximately 1 in 5 hospital 30-day readmissions (Soh, 2020). In 2017, there were 270,702 deaths where DM was listed as primary or underlying cause making DM related mortality the 6th leading cause of death in the U.S. (Centers for Disease Control and Prevention, 2020). Adults with DM represent and accounted for approximately 1 in 4 health care dollars ($327 billion) spent in the U.S. representing an increase of a 26% from 2012 after adjusting for inflation (American Diabetes Association, 2018). From 2001–2014, the total cost for preventable hospitalizations related to DM increased from $4.5 to $5.9 billion (Shrestha, 2019). While utilization from DM outcomes are costly, evidence-based interventions are extremely cost-effective for almost all segments of the population. Based on these trends, DM outcomes represent one of the largest avoidable cost burdens with opportunity for improvement in the U.S. health care system.

**Background**

Diabetes Mellitus (DM) is a chronic metabolic disease that is characterized by a deterioration in the production of insulin for managing blood glucose. For those with DM, the pancreas is limited in ability or no longer able to produce enough insulin to keep blood glucose levels within an appropriate range. Without treatment, those with DM will experience hyperglycemia when blood glucose levels remain persistently elevated. Symptoms of hyperglycemia include dehydration, excessive urination, sudden weight loss, dizziness or nausea. If hyperglycemia persists without diagnosis and treatment, diabetic ketoacidosis can occur leading to the need for acute hospitalization and sometimes resulting in disability or death.

Diagnosis of DM is confirmed with a hemoglobin A1c (HbA1c) of 6.5% or higher or a fasting blood glucose (FBG) of 126 mg/dL or higher. Treatment for DM often includes insulin therapy, which involves the injection of artificial insulin at appropriate times for managing healthy blood glucose levels. Along with insulin therapy, living with DM includes consistent glucose monitoring, pharmaceutical regimens, and intensive lifestyle modifications. DM management is a complex process and requires the individual to work closely with providers to control many aspects of their daily living. This often includes food intake, meal timing, medication regiments, sleep cycles, physical exercise habits, and infectious disease prevention. Those with DM are at risk for developing retinopathy (eye damage) and neuropathy (nerve damage) and are at higher risk for cerebrovascular (stroke) and cardiovascular (heart) diseases. DM is also the leading cause of End Stage Renal Disease (ESRD) leading to long-term dialysis or death from kidney failure.

There are multiple types of DM that share common diagnostic criteria and similar methods of treatment, but with different pathologies and prevention practices. The primary forms of DM are Type 1 Diabetes, Type 2 Diabetes, and Prediabetes.

### *Type 1 Diabetes*

T1D is a sudden loss in beta-cell function within the pancreas and the inability to create insulin necessary for blood glucose management. T1D is the result of an auto-immune reaction that attacks beta-cell function inside the pancreas, leading to the need for insulin therapy. It is not fully understood what leads to the auto-immune reaction and developing T1D is not related to family history or genetics. Environmental factors appear to have some increased risk for developing T1D but are currently not fully understood. Due to the sudden loss in insulin production, T1D is managed primarily through insulin therapy. Due to the chronic need for insulin therapy, individuals with T1D benefit from recent advancements in insulin delivery technology such as insulin pumps. T1D is commonly diagnosed earlier in life within children, adolescents, and young adults. There is currently no cure or vaccine for T1D.

### *Type 2 Diabetes*

Type 2 Diabetes (T2D) occurs after a prolonged period of insulin resistance and hyperinsulinemia culminating in the loss of beta cell function. Insulin resistance occurs when the insulin produced by the pancreases loses the ability to maintain the same amount of blood glucose, often due to persistent demand for insulin from carbohydrate consumption and a lack of physical activity. In response, hyperinsulinemia occurs when the pancreas produces extra insulin in order to manage blood glucose levels. Eventually, insulin resistance and hyperinsulinemia lead to depleted ability for the pancreas to create enough insulin and blood glucose levels rise above a healthy range for extended periods. Individuals that are obese, over 65, physically inactive, with high blood pressure, and with high cholesterol are at higher risk for developing T2D. While T2D is not reversible after a certain threshold, progression of the disease can be significantly decreased and many patients with T2D that experience adequate disease management will not become insulin dependent. T2D contributes to approximately 85% of cases of DM in the U.S.

### *Prediabetes*

The final phase of insulin resistance and hyperinsulinemia before T2D diagnosis is called Prediabetes (PD). PD is diagnosed with an HbA1c between 5.7% and 6.4% and often does not present clinical symptoms. This leads to widespread underdiagnosis for PD and an under deployment of preventative care. The process of insulin resistance and hyperinsulinemia can be significantly slowed or reversed with lifestyle modification and pharmaceutical intervention. Not all that develop PD will progress to T2D, but those with PD have a significantly higher risk of developing DM.

### Diabetes Treatment and Prevention

While insulin sensitivity rises during aging, subsequent PD and T2D diagnosis are not a natural result of the aging process. T2D diagnosis is preventable for those with high risk through non-pharmaceutical interventions including the Center for Disease Control’s (CDC) evidence-based Diabetes Prevention Program (DPP). DPP is a standardized, flexible, and relatively inexpensive peer support lifestyle intervention focusing on a 5% reduction in body weight and an increase of 90 minutes of physical activity (Diabetes Prevention Program Research Group, 2002). In the initial randomized control trial, DPP was shown to be twice as effective as Metformin in preventing progression to T2D (Knowler, 2002). In a 10 year follow up, DPP has been shown to be both consistently effective at prevention and while saving costs related to DM outcomes (Diabetes Prevention Program Research Group, 2012).

For those diagnosed with DM, it is possible to manage the disease and avoid hospitalization throughout the course of the disease (American Diabetes Association, 2016). Diabetes Self-Management Education (DSME) is standardized primary care approach for equipping patients to effectively live with the disease and prevent progression to negative outcomes (Beck, 2018). Accredited DSME programs have been consistently found to improve disease outcomes (UK Prospective Diabetes Study Group, 1998) and prevent health care utilization (Strawbridge, 2017). Most DM outcomes are potentially preventable (Sentell, 2016) and reflect a failure in adequate primary care delivery (Rubens, 2018). Based on the consistent evidence related to prevention and management, DM outcomes represent a reasonably preventable public health problem (Ricci-Cabello, 2013).

### *Disparities in Diabetes Outcomes*

Consistent with other chronic diseases in the U.S., minority populations and those with lower socioeconomic status (SES) experience much higher rates of DM diagnosis and worse DM outcomes. The observed disparities by race and ethnicity as well SES and deprivation have been well validated by recent evidence.

**Race and ethnicity**. Non-Hispanic blacks (NHB), Hispanics, and American Indian or Alaskan Natives (AI/AN) experience significantly higher rates of DM when compared to non-Hispanic whites (NHW) (Centers for Disease Control and Prevention, 2020). Genome-wide association studies have found that risk for the development and progression of DM is relatively consistent among ethnic groups (Saxena, 2012) and genetic risk related to ancestral origin did not account for the disparity in observed outcomes among historically disadvantaged populations (Waters, 2010). Over the past two decades, persistent historical inequality among racial groups has been clearly established as significant factors in many disease outcomes within the U.S., including DM (Nelson, 2002). differences in T2D outcomes among groups are better explained as the result of institutional forces rather than inherent biological risk (Ricci-Cabello, 2013).

While NHBs, Hispanics, and AI/AN have a signficiantly higher burden of T2D in the U.S., these groups received recommended T2D care less frequently regardless of insurance coverage (Canedo, 2018). NHBs have significantly higher rates of hospitalization and 30-day readmission for DM related outcomes when controlling for SES, insurance coverage, and hospital type (Jiang, 2005) (Rodriguez-Gutierrez, 2019) (Soh, 2020). Identification with a racial or ethnic group in the U.S has been shown to be a persistent predictor for earlier diagnosis and more aggressive DM disease progression when controlling for SES and insurance status (Peek, 2007). Technologies and therapeutics for managing T1D have been shown to be offered less consistently to NHBs when controlling for insurance coverage and education level (Addala, 2020) leading to marked differences in treatment outcomes (Willi, 2015). Hispanic youth have higher future risk for DM (Cistola, 2020) and are more likely to be diagnosed under the age of 30 (Magliano, 2020).

**Socioeconomic status and deprivation.** Those with low socioeconomic status (SES) are at the highest risk globally for acquiring DM (Glazier, 2006) as well experiencing more hospitalizations and higher death rates (Jaffiol, 2012). Lower SES has been identified as an independent effect modifier of mortality when experiencing specific DM complications (Anderson, 2018). Neighborhood-level deprivation similarly has been also been associated increased risk of DM outcomes (Mezuk, 2013) (Laraia, 2012) (White, 2016). Deprived neighborhoods are also at higher risk of having less access to specialists necessary for T1D care (Walker, 2020).

### Population Health Management

Population health management (PHM) refers to the activities for improving health outcomes of a defined population through “improved care coordination and patient engagement supported by appropriate financial and care models” (American Hospital Association, 2020) combining the disciplines of public health and health administration. In contrast to segmented approaches to clinical health, PHM involves comprehensive approaches that identify the organizational structures and institutional policies related to health in a defined location (Chin, 2007). Improving DM outcomes will require comprehensive population health management (PHM) (Golden, 2017) that considers marked disparities among minority groups (Jiang, 2011), seeks to address upstream factors related to SES (Finkelstein, 2020), and includes updates to institutional policies, particularly at the state level (Gibbs, 2006). PHM efforts will need to be driven by location-based analyses that incorporate social stratification and community infrastructure that impact DM outcomes (Cistola, 2018).

While studies have identified that neighborhood level disadvantage is a significant predictor of clinical outcomes (Diez Roux, 2010) often through the use of common indexes based on public data (Messer, 2006), the full extent of neighborhood level factors needs further investigation (Scaria, 2020). Recent studies have attempted to broaden the scope of multi-dimensional neighborhood level factors that are associated with health outcomes, however factors under consideration include limited sets (less than 20) possible predictors (Kolak, 2020). Current interoperability issues related to health care information technology present significant hurdles when trying to conduct clinically focused population or geographic studies informative to PHM efforts (Vest, 2016).

The process of applying geographically based data science techniques “drive public health assessment, policy, and implementation activities” is referred to as Precision Public Health (PPH) (Khoury, 2020). PPH research strives to incorporate with innovative tools in machine learning and artificial intelligence in public health research. One of implementations of PPH analyses has been the use of small-area analysis (below the County and State level) to identify more specific areas of interest for public health interventions that can often be unnoticed in county or state aggregate data (Kind, 2018). Studies using geographic information systems (GIS) with public data sources have produced innovative analyses on drivers of health outcomes (Moore, 2016) and assisted in population health management efforts (Beck, 2019), but more research using these methods are needed (Holder, 2016).

### Artificial Intelligence in Public Health

Artificial Intelligence (AI) is a subset of data science that uses computational resources to perform tasks that normally require human intelligence (Wang, 2019). While AI in health science is expected to bring an exciting new phase in care delivery (Bourne, 2015), the current deployment of AI in healthcare has been inconsistent (Nevin, 2018). Within the clinical practice, there has been a struggle to implement AI and identify meaningful use cases (Deo, 2015), and current advances above established baselines have only been modest (Ching, 2018). Literature reviews indicate limited dialogue on what public health problems can be addressed with existing AI tools (Miotto, 2018), what technological infrastructure is necessary for AI usage in healthcare (Chan, 2019), and how to practically address ethical concerns that arise from AI implementation for clinical practice (Reedy, 2020). Machine learning algorithms and artificial intelligence tools have struggled to provide more meaningful prediction capability than traditional statistical methods in the public health context (Christodoulou, 2019) (Bian, 2019) and current applications of machine learning to healthcare systems remain limited (Khoury, 2020).

### Purpose of this Study

In order for population health management efforts to be effective, public health researchers will need to investigate the full extent of social and infrastructural risk factors available in public data sources. These types of analyses will need to handle more variables, in multiple dimensions, and at higher complexity than commonly used statistical tests can currently handle without the use of AI processes. These analyses also need to be considerate of efforts relevant to health care decision makers and handle interoperability and data collection realities specific to the U.S. healthcare system. These analyses will need to integrate a wider scope of ecological factors with increased precision that respects patient privacy restrictions and are achievable by computational resources available to health care organizations. Improving health equity in the context of DM will require targeted community improvements, infrastructure investments, and policy interventions that are designed to maximize the impact of resource allocation through the use of available data and computational resources.

Utilizing a novel approach combining multiple open source machine learning algorithms, artificial neural networks, and statistical tests to meet specific considerations related to the forementioned issues, this study is designed to provide a PPH framework useful to PHM management efforts for DM mortality. By using an Artificial Intelligence approach to evaluate over 2000 socio-economic and infrastructural predictors of DM mortality, this study is designed to provide granular information on areas necessary to improve DM outcomes in the state of Florida. It is hypothesized that AI approaches using these methods will identify an actionable subset (under 20) of statistically significant socio-economic and infrastructural factors will be able to be identified without human selection. It is also hypothesized that these selected predictors will be able to provide improved predictive capability when fed through artificial neural networks when compared to all possible predictors. This would indicate that the selected predictors are highly valuable when identifying populations at risk for DM mortality in FL.

## Methods

In order to identify significant predictors of DM mortality, crude mortality counts for T1D, T2D, chronic kidney disease (CKD), and other related metabolic conditions were collected from the Florida Department of Health (FDOH) by zip code from the years 2014-2018. The counts were collected from death certificates tabulated by the FDOH with a primary cause of death for all ICD-10 codes K00-K99. While these codes contain deaths not caused by DM but other metabolic diseases, they account for a very small portion of deaths. The outcome variable was created by calculating population adjusted rates averaged from these raw counts for the five-year time period.

In order to identify socio-economic predictors, US Census American Community Survey (ACS) five-year percent estimates by zip code were collected from the 2018 data release. This dataset provides over 500 measurements of demographics, housing types, economic indicators, education levels and employment types for the population within each geographic boundary. In order to identify infrastructural predictors of DM mortality, the Health Services Resource Administration (HRSA) Area Health Resource File 2019 data release was collected and five-year averages were calculated by county. The AHRF provides over 2000 population adjusted rates for pertinent health care providers, organizational resources, and economic indicators pertinent to health care delivery. Zip codes were joined to county using crosswalk files from the Department of Housing and Urban development so each Zip code could be contained within a specific county based on where a majority of the population is located. Each of these datasets included observations for the years 2014-2018 from the latest publicly available data release.

In order to search among possible social and infrastructural predictors for DM mortality, a number of considerations need to account for specific needs related to multi-dimensional geographic data. While common statistical tests used to identify significant predictors for outcomes of interest can account for many variables, probability-based statistics struggles to identify relevant predictors above 1000 possible candidates. The following methods are designed to identify significant predictors with an AI approach that does not need require human selection.

### Modeling Techniques

In order to identify significant predictors among available data, a specific series of modeling techniques were used. For ACS variables at the zip code level, principal component analysis (PCA) was used to identify predictors with above average eigenvectors for all components with an eigenvalue above 1.0. This allowed for predictors with higher variation to be identified from among possible candidates. Random forests (RF) were then used to identify predictors with above average importance measured by gini impurity. This allowed for variables to be independently evaluated for predictive ability that adjusts for collinearity and confounding. Predictors that had both above average variation and importance were selected for recursive feature elimination (RFE) with cross validation. This process calculated AIC values for all possible combinations of selected predictors and identified the best possible selection with the lowest number of predictors. The resulting list of variables was then used to create a multiple linear regression model (MR) alongside population over 65 in order to create an informative model for evaluating overall prediction and significance of each predictor from the ACS.

The list of ACS variables was then used for multi-scale geographic weighted regression (GWR) involving a location based assessment of each variables predictive ability. Based on the variation of GWR coefficients, counties where zip code level predictors had higher predictive ability were able to be identified indicating locations where infrastructural differences had possible impact on socio-economic factors. Each county was coded with a multi-level categorical variable indicating which ACS variable had the highest GWR coefficient. Support vector machines (SVM) were used to identify county level predictors from the AHRF with the highest coefficients for each ACS variable. SVM were chosen for their ability to provide consistent accuracy with lower numbers of observations (67 counties in FL) and higher possible predictors (over 2000). In order to remove low variation county level predictors, PCA was conducted and component loadings of an absolute value over 0.5 on components accounting for 95% of cumulative variation were kept.

Once county and zip code level predictors were selected, a comprehensive MR model was used to evaluate significance and direction all predictors. In order to evaluate the predictive ability and practical usage of the selected factors, a multi-layered perceptron (MLP) using an artificial neural network (ANN) architecture was used to predict a zip code being in the 50th percentile for diabetes mortality. AUC scores were compared for all predictors and selected zip code and county predictors so that different selections could be compared.

Each of these modeling techniques were conducted with open source packages available in the Python Programming Language (Python) and available as open source for access and audit under the MIT license from GitHub at <https://github.com/andrewcistola/fracture-proof>.

## Results

The final regression model using zip code and county level predictors had an R2 of 0.863. Significant predictors at the 0.05 level for zip codes were: Population % White, Population % Householders, Population % Spanish spoken at home, Population % Divorced males, Population % With public health insurance coverage, Population % Employed with private health insurance coverage. Significant predictors at the 0.05 level for counties were: Manufacturing-Dependent Designation, Low Education Designation, Population % Medicare Part A & B Female Beneficiaries, Number of Short Term General Hospitals with 50-99 Beds. Using the multi-layered perceptron to predict zip codes above 50th percentile in diabetes related mortality and conducting a Receiver Operator Curve test, the C-statistic for all 2000+ zip code and county level predictors was 0.7938. When only using the 13 selected predictors, the C-statistic was 0.8232.

## Discussion

The purpose of this study is to identify the socio-economic and infrastructural predictors of diabetes related mortality from relevant publicly available data. Using Artificial Intelligence to search among 2000+ possible options, this study identified 13 specific zip code and county level predictors that have significant association and strong predictive capability with diabetes mortality. These predictors remained significant even when controlling for population over 65. When evaluating those predictors, the strongest themes are related to employment and insurance coverage as well as isolation, education, and poverty. The selected 13 predictors also showed improved performance when fed into artificial neural networks when compared to all possible predictors, indicating that these factors are highly relevant for DM mortality in Florida. This process was completed without the need of human variable selection and indicates how AI can be used for informative analyses in PPH for targeted PHM efforts.

### Relevance

While these predictors are not radically different than other predictors existing in the literature, these predictors have a much higher level of precision and can inform targeted population health management efforts in Florida. These methods can also inform future research into multi-dimensional factors related to a number of population related outcomes reported in public data. Since Artificial Intelligence tools have not yet been actively deployed in the field of public health and welfare economics, this approach has the potential to empower a burst of informative studies related to pertinent public policy decisions.

### Limitations

This study focused only on the state of Florida. A multi-state study would need to incorporate informative data at the state level as well as identify consistent data reported from various states. If this can be achieved, state level policy approaches could be evaluated with significant precision. County level data was collected from only one source (HRSA). County and zip code level data is also publicly available from the USDA, EPA, CMS, and BEA. Future analyses should incorporate multiple county and zip code level datasets for a broader scope. Doing so will require significant time and infrastructure for acquiring and standardizing publicly available geographic data.

While zip codes can provide an informative measure of geographic distribution, census tracts are the preferred geographic layer for neighborhood level differences. Census tracts also cleanly fit into county equivalents and can more readily interact between different layers. While census tract mortality data from the FDOH has been public in the past, as of the time this study was conducted zip codes were the only geographic layer mortality counts were available below the county level. Due to the practical constraints for the methods used at the county level, only the top predictors were collected. It is possible that a lower threshold (top 3 or 5) could yield more informative results. A practical assessment of the practical steps in data processing could help these results.

The models used for the analysis were chosen based on their ability to perform specific tasks under defined parameters for multi-dimensionality data reduction and outcome prediction. While there is robust reasoning for each of these choices, there may be more optimal models that could be employed for increased significant associations and predictive capability. It is also entirely possible that the results generated by these methods are altogether meaningless. Since these methods take a hypothesis generating approach, practical use will need to be filtered through a more specific lens. Due to the observational nature of this study, the result should not be interpreted as causality inferring. However, these results would have the most meaningful use in the context of intervention design. Causality and practicality can be inferred only by interventions that utilize these results to inform a randomized, controlled, clinical trial seeking to improve the outcome of interest. This study is designed to be preliminary research that informs these types of studies and is not intended to supplant these types of studies.

### Conclusion

The purpose of this study was to identify the socio-economic and infrastructural predictors of diabetes related mortality from relevant publicly available data. Using Artificial Intelligence to search among 2000 social and infrastructural factors, this study identified 13 specific predictors relevant for targeted PHM efforts. This study also models how AI can be used in future PPH research.

**References**

Addala, A., Auzanneau, M., Miller, K., Maier, W., Foster, N., Kapellen, T., Walker, A., Rosenbauer, J., Maahs, D. M., & Holl, R. W. (2020). A decade of disparities in diabetes technology use and hba1c in pediatric type 1 diabetes: A transatlantic comparison. Diabetes Care. https://doi.org/10.2337/dc20-0257

American Diabetes Association. (2018). Economic costs of diabetes in the U.S. in 2017. Diabetes Care, 41(5), 917–928. https://doi.org/10.2337/dci18-0007

American Diabetes Association. (2016). Foundations of care and comprehensive medical evaluation. Diabetes Care, 39 Suppl 1, S23-35. <https://doi.org/10.2337/dc16-S006>

American Hospital Asssociation (2020). What is Population Health Management? AHA.org. https://www.aha.org/center/population-health/population-health-management

Anderson, S. G., Shoo, H., Saluja, S., Anderson, C. D., Khan, A., Livingston, M., Jude, E. B., Lunt, M., Dunn, G., & Heald, A. H. (2018). Social deprivation modifies the association between incident foot ulceration and mortality in type 1 and type 2 diabetes: a longitudinal study of a primary-care cohort. Diabetologia, 61(4), 959–967. <https://doi.org/10.1007/s00125-017-4522-x>

Beck, A. F., Anderson, K. L., Rich, K., Taylor, S. C., Iyer, S. B., Kotagal, U. R., & Kahn, R. S. (2019). Cooling the hot spots where child hospitalization rates are high: A neighborhood approach to population health. Health Affairs (Project Hope), 38(9), 1433–1441. https://doi.org/10.1377/hlthaff.2018.05496

Beck, J., Greenwood, D. A., Blanton, L., Bollinger, S. T., Butcher, M. K., Condon, J. E., Cypress, M., Faulkner, P., Fischl, A. H., Francis, T., Kolb, L. E., Lavin-Tompkins, J. M., MacLeod, J., Maryniuk, M., Mensing, C., Orzeck, E. A., Pope, D. D., Pulizzi, J. L., Reed, A. A., … 2017 Standards Revision Task Force. (2018). 2017 National Standards for Diabetes Self-Management Education and Support. The Diabetes Educator, 44(1), 35–50. https://doi.org/10.1177/0145721718754797

Bian, J., Buchan, I., Guo, Y., & Prosperi, M. (2019). Statistical thinking, machine learning. Journal of Clinical Epidemiology, 116, 136–137. <https://doi.org/10.1016/j.jclinepi.2019.08.003>

Bourne, P. E., Bonazzi, V., Dunn, M., Green, E. D., Guyer, M., Komatsoulis, G., Larkin, J., & Russell, B. (2015). The NIH Big Data to Knowledge (BD2K) initiative. Journal of the American Medical Informatics Association, 22(6), 1114.

Canedo, J. R., Miller, S. T., Schlundt, D., Fadden, M. K., & Sanderson, M. (2018). Racial/ethnic disparities in diabetes quality of care: the role of healthcare access and socioeconomic status. Journal of Racial and Ethnic Health Disparities, 5(1), 7–14. https://doi.org/10.1007/s40615-016-0335-8

Centers for Disease Control and Prevention. (2020). *National Diabetes Statistics Report, 2020*. U.S. Dept of Health and Human Services. https://www.cdc.gov/diabetes/data/statistics-report/index.html?CDC\_AA\_refVal=https%3A%2F%2Fwww.cdc.gov%2Fdiabetes%2Fdata%2Fstatistics%2Fstatistics-report.html

Chan, K. S., & Zary, N. (2019). Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMIR Medical Education, 5(1), e13930.

Chin, M. H., Walters, A. E., Cook, S. C., & Huang, E. S. (2007). Interventions to reduce racial and ethnic disparities in health care. Medical Care Research and Review, 64(5 Suppl), 7S-28S. https://doi.org/10.1177/1077558707305413

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P.-M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., … Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. Journal of the Royal Society, Interface, 15(141).

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of Clinical Epidemiology, 110, 12–22. <https://doi.org/10.1016/j.jclinepi.2019.02.004>

Cistola, A. S. (2019). Abstract P081: Applying Precision Public Health Tools to Address Health Disparities Among African-Americans in Alachua County, Florida. Circulation, 139(Suppl\_1).

Cistola, D. P., Ard, J. D., Brenner, M. H., & Dwivedi, A. K. (2020). Compensatory Hyperinsulinemia in Young Adults and the Risk of Future Diabetes: CARDIA 25-year Follow Up. Metabolism, 104, 154131. https://doi.org/10.1016/j.metabol.2019.12.077

Diabetes Prevention Program Research Group. (2012). The 10-year cost-effectiveness of lifestyle intervention or metformin for diabetes prevention: an intent-to-treat analysis of the DPP/DPPOS. Diabetes Care, 35(4), 723–730. <https://doi.org/10.2337/dc11-1468>

Diabetes Prevention Program (DPP) Research Group. (2002). The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care, 25(12), 2165–2171. <https://doi.org/10.2337/diacare.25.12.2165>

Diez Roux, A. V., & Mair, C. (2010). Neighborhoods and health. Annals of the New York Academy of Sciences, 1186, 125–145. https://doi.org/10.1111/j.1749-6632.2009.05333.x

Deo, R. C. (2015). Machine learning in medicine. Circulation, 132(20), 1920–1930.

Finkelstein, A., Zhou, A., Taubman, S., & Doyle, J. (2020). Health Care Hotspotting - A Randomized, Controlled Trial. The New England Journal of Medicine, 382(2), 152–162. https://doi.org/10.1056/NEJMsa1906848

Gibbs, B. K., Nsiah-Jefferson, L., McHugh, M. D., Trivedi, A. N., & Prothrow-Stith, D. (2006). Reducing racial and ethnic health disparities: exploring an outcome-oriented agenda for research and policy. Journal of Health Politics, Policy and Law, 31(1), 185–218. https://doi.org/10.1215/03616878-31-1-185

Glazier, R. H., Bajcar, J., Kennie, N. R., & Willson, K. (2006). A systematic review of interventions to improve diabetes care in socially disadvantaged populations. Diabetes Care, 29(7), 1675–1688. <https://doi.org/10.2337/dc05-1942>

Golden, S. H., Maruthur, N., Mathioudakis, N., Spanakis, E., Rubin, D., Zilbermint, M., & Hill-Briggs, F. (2017). The Case for Diabetes Population Health Improvement: Evidence-Based Programming for Population Outcomes in Diabetes. Current Diabetes Reports, 17(7), 51. <https://doi.org/10.1007/s11892-017-0875-2>

Holder, A. L., Wallace, D. J., & Martin, G. S. (2016). Hotspotting sepsis: applying analytic tools from other disciplines to eliminate disparities. Annals of Translational Medicine, 4(15), 295. https://doi.org/10.21037/atm.2016.07.19

Jaffiol, C., Fontbonne, A., Vannereau, D., Olive, J.-P., & Passeron, S. (2012). [Diabetes and social deprivation]. Bulletin de l’Academie Nationale de Medecine, 196(4–5), 953–975; discussion 976.

Jiang, H. J., Andrews, R., Stryer, D., & Friedman, B. (2005). Racial/ethnic disparities in potentially preventable readmissions: the case of diabetes. American Journal of Public Health, 95(9), 1561–1567. https://doi.org/10.2105/AJPH.2004.044222

Khoury, M. J., Armstrong, G. L., Bunnell, R. E., Cyril, J., & Iademarco, M. F. (2020). The intersection of genomics and big data with public health: Opportunities for precision public health. *PLoS Medicine*, *17*(10), e1003373. https://doi.org/10.1371/journal.pmed.1003373

Kind, A. J. H., & Buckingham, W. R. (2018). Making Neighborhood-Disadvantage Metrics Accessible - The Neighborhood Atlas. The New England Journal of Medicine, 378(26), 2456–2458. <https://doi.org/10.1056/NEJMp1802313>

Knowler, W. C., Barrett-Connor, E., Fowler, S. E., Hamman, R. F., Lachin, J. M., Walker, E. A., Nathan, D. M., & Diabetes Prevention Program Research Group. (2002). Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. The New England Journal of Medicine, 346(6), 393–403. https://doi.org/10.1056/NEJMoa012512

Kolachalama, V. B., & Garg, P. S. (2018). Machine learning and medical education. Npj Digital Medicine, 1, 54.

Kolak, M., Bhatt, J., Park, Y. H., Padrón, N. A., & Molefe, A. (2020). Quantification of Neighborhood-Level Social Determinants of Health in the Continental United States. JAMA Network Open, 3(1), e1919928. <https://doi.org/10.1001/jamanetworkopen.2019.19928>

Laraia, B. A., Karter, A. J., Warton, E. M., Schillinger, D., Moffet, H. H., & Adler, N. (2012). Place matters: neighborhood deprivation and cardiometabolic risk factors in the Diabetes Study of Northern California (DISTANCE). Social Science & Medicine, 74(7), 1082–1090. <https://doi.org/10.1016/j.socscimed.2011.11.036>

Magliano, D. J., Sacre, J. W., Harding, J. L., Gregg, E. W., Zimmet, P. Z., & Shaw, J. E. (2020). Young-onset type 2 diabetes mellitus - implications for morbidity and mortality. Nature Reviews. Endocrinology, 16(6), 321–331. <https://doi.org/10.1038/s41574-020-0334-z>

Messer, L. C., Laraia, B. A., Kaufman, J. S., Eyster, J., Holzman, C., Culhane, J., Elo, I., Burke, J. G., & O’Campo, P. (2006). The development of a standardized neighborhood deprivation index. Journal of Urban Health, 83(6), 1041–1062. https://doi.org/10.1007/s11524-006-9094-x

Mezuk, B., Chaikiat, Å., Li, X., Sundquist, J., Kendler, K. S., & Sundquist, K. (2013). Depression, neighborhood deprivation and risk of type 2 diabetes. Health & Place, 23, 63–69. https://doi.org/10.1016/j.healthplace.2013.05.004

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. Briefings in Bioinformatics, 19(6), 1236–1246.

Moore, J. X., Donnelly, J. P., Griffin, R., Howard, G., Safford, M. M., & Wang, H. E. (2016). Defining sepsis mortality clusters in the united states. Critical Care Medicine, 44(7), 1380–1387. https://doi.org/10.1097/CCM.0000000000001665

Nelson, A. (2002). Unequal treatment: confronting racial and ethnic disparities in health care. Journal of the National Medical Association, 94(8), 666–668.

Nevin, L., & PLOS Medicine Editors (2018). Advancing the beneficial use of machine learning in health care and medicine: Toward a community understanding. PLoS medicine, 15(11), e1002708.

Peek, M. E., Cargill, A., & Huang, E. S. (2007). Diabetes health disparities: a systematic review of health care interventions. Medical Care Research and Review, 64(5 Suppl), 101S-56S. https://doi.org/10.1177/1077558707305409

Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. Journal of the American Medical Informatics Association, 27(3), 491-497.

Ricci-Cabello, I., Ruiz-Pérez, I., Nevot-Cordero, A., Rodríguez-Barranco, M., Sordo, L., & Gonçalves, D. C. (2013). Health care interventions to improve the quality of diabetes care in African Americans: a systematic review and meta-analysis. Diabetes Care, 36(3), 760–768. <https://doi.org/10.2337/dc12-1057>

Rodriguez-Gutierrez, R., Herrin, J., Lipska, K. J., Montori, V. M., Shah, N. D., & McCoy, R. G. (2019). Racial and Ethnic Differences in 30-Day Hospital Readmissions Among US Adults With Diabetes. JAMA Network Open, 2(10), e1913249. https://doi.org/10.1001/jamanetworkopen.2019.13249

Rubens, M., Saxena, A., Ramamoorthy, V., Khera, R., Hong, J., Veledar, E., & Nasir, K. (2018). Trends in Diabetes-Related Preventable Hospitalizations in the U.S., 2005-2014. Diabetes Care, 41(5), e72–e73. https://doi.org/10.2337/dc17-1942

Scaria, E., Powell, W. R., Birstler, J., Alagoz, O., Shirley, D., Kind, A. J. H., & Safdar, N. (2020). Neighborhood disadvantage and 30-day readmission risk following Clostridioides difficile infection hospitalization. BMC Infectious Diseases, 20(1), 762. <https://doi.org/10.1186/s12879-020-05481-x>

Saxena, R., Elbers, C. C., Guo, Y., Peter, I., Gaunt, T. R., Mega, J. L., Lanktree, M. B., Tare, A., Castillo, B. A., Li, Y. R., Johnson, T., Bruinenberg, M., Gilbert-Diamond, D., Rajagopalan, R., Voight, B. F., Balasubramanyam, A., Barnard, J., Bauer, F., Baumert, J., … Keating, B. J. (2012). Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. American Journal of Human Genetics, 90(3), 410–425. https://doi.org/10.1016/j.ajhg.2011.12.022

Sentell, T. L., Seto, T. B., Young, M. M., Vawer, M., Quensell, M. L., Braun, K. L., & Taira, D. A. (2016). Pathways to potentially preventable hospitalizations for diabetes and heart failure: a qualitative analysis of patient perspectives. BMC Health Services Research, 16, 300. <https://doi.org/10.1186/s12913-016-1511-6>

Shrestha, S. S., Zhang, P., Hora, I., Geiss, L. S., Luman, E. T., & Gregg, E. W. (2019). Factors Contributing to Increases in Diabetes-Related Preventable Hospitalization Costs Among U.S. Adults During 2001-2014. Diabetes Care, 42(1), 77–84. <https://doi.org/10.2337/dc18-1078>

Soh, J. G. S., Wong, W. P., Mukhopadhyay, A., Quek, S. C., & Tai, B. C. (2020). Predictors of 30-day unplanned hospital readmission among adult patients with diabetes mellitus: a systematic review with meta-analysis. BMJ Open Diabetes Research & Care, 8(1). https://doi.org/10.1136/bmjdrc-2020-001227

Strawbridge, L. M., Lloyd, J. T., Meadow, A., Riley, G. F., & Howell, B. L. (2017). One-Year Outcomes of Diabetes Self-Management Training Among Medicare Beneficiaries Newly Diagnosed With Diabetes. Medical Care, 55(4), 391–397. <https://doi.org/10.1097/MLR.0000000000000653>

UK Prospective Diabetes Study Group. (1998). Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). The Lancet, 352(9131), 837–853. <https://doi.org/10.1016/S0140-6736(98)07019-6>

Vest, J. R., Harle, C. A., Schleyer, T., Dixon, B. E., Grannis, S. J., Halverson, P. K., & Menachemi, N. (2016). Getting from here to there: health IT needs for population health. The American Journal of Managed Care, 22(12), 827–829.

Walker, A. F., Hu, H., Cuttriss, N., Anez-Zabala, C., Yabut, K., Haller, M. J., & Maahs, D. M. (2020). The neighborhood deprivation index and provider geocoding identify critical catchment areas for diabetes outreach. The Journal of Clinical Endocrinology and Metabolism, 105(9). https://doi.org/10.1210/clinem/dgaa462

Wang, P. (2019). On defining artificial intelligence. Journal of Artificial General Intelligence, 10(2), 1–37.

Waters, K. M., Stram, D. O., Hassanein, M. T., Le Marchand, L., Wilkens, L. R., Maskarinec, G., Monroe, K. R., Kolonel, L. N., Altshuler, D., Henderson, B. E., & Haiman, C. A. (2010). Consistent association of type 2 diabetes risk variants found in europeans in diverse racial and ethnic groups. PLoS Genetics, 6(8), e1001078. <https://doi.org/10.1371/journal.pgen.1001078>

White, J. S., Hamad, R., Li, X., Basu, S., Ohlsson, H., Sundquist, J., & Sundquist, K. (2016). Long-term effects of neighbourhood deprivation on diabetes risk: quasi-experimental evidence from a refugee dispersal policy in Sweden. The Lancet. Diabetes & Endocrinology, 4(6), 517–524. https://doi.org/10.1016/S2213-8587(16)30009-2

Wiljer, D., & Hakim, Z. (2019). Developing an Artificial Intelligence-Enabled Health Care Practice: Rewiring Health Care Professions for Better Care. Journal of Medical Imaging and Radiation Sciences, 50(4S2), S8–S14.

Willi, S. M., Miller, K. M., DiMeglio, L. A., Klingensmith, G. J., Simmons, J. H., Tamborlane, W. V., Nadeau, K. J., Kittelsrud, J. M., Huckfeldt, P., Beck, R. W., Lipman, T. H., & T1D Exchange Clinic Network. (2015). Racial-ethnic disparities in management and outcomes among children with type 1 diabetes. Pediatrics, 135(3), 424–434. https://doi.org/10.1542/peds.2014-1774