Harnessing Machine Learning to Assess Cervical Cancer Health Access

D. Morel[1], X. Zhao[2], W. Leo[1], A. Liu-Ferrara[1]
1. Health Economics Outcomes Research (HEOR) and Real-World Evidence, Data Science, BD
2. Single Cell Multiomics ,Waters Bioscience (Formerly BD Biosciences)

Background and Significance

Cervical cancer trends in the US

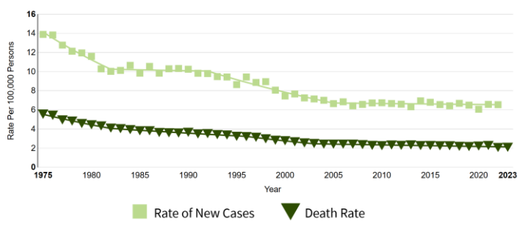


Figure 1. Cervical cancer trends in the US, showing progresses made in prevention, but mortality risks remain.

Cervical cancer by race and ethnicity

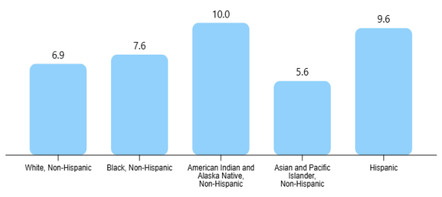


Figure 2. Rate of new cancers in United States 2022

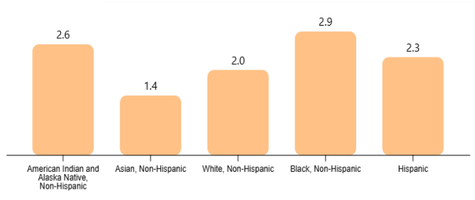


Figure 3. Rate of cancer deaths in United States 2023

Healthy People 2030 (HHS):

- Set national data-driven objectives to improve health and well-being
- Measure health access gaps and track status with 23 Leading Health Indicators
  - ➢ Increase the proportion of females who get screened for cervical cancer

Most Recent Data: **73.9** percent (2021) *
Target: **79.2** percent [1] *

Table 1. A comparison of various current health access frameworks

Framework	IHI Four-Step Approach	NQF Roadmap	NCQA Health Equity Accreditation	CMS Health Equity Framework	Our Framework
Primary Goal	Standardize measurement of disparities	Promote equity via quality measures	Accreditation and compliance	Integrate equity into hospital performance	Operationalize equity with predictive analytics
Steps / Structure	1. Identify; 2. Stratify; 3. Choose Reference; 4. Quantify	1. Identify; 2. Implement; 3. Invest; 4. Incentivize	Standards for data collection, stratification, and reporting	Strategic priority + stratified KPI dashboards	1. Identify; 2. Model Building & Metric Derivation; 3. Interactive Visualization
Data Emphasis	Strong focus on data quality (accuracy, completeness, timeliness)	Requires stratification by social risk factors	Requires demographic and social needs data	Mandates stratification and dashboard inclusion	Heavy: integrates EHR, claims, open-source, SDOH
Visualization	Not included	Optional	Optional	Basic dashboards	Advanced dashboards
Actionability	Conceptual guidance	Policy-driven	Compliance-focused	Linked to reimbursement	Direct intervention planning

Our proposed framework builds on these principles but introduces two key innovations:

- **Streamlined analysis:** Combining reference point selection and disparity quantification into an integrated phase.
- **Visualization capabilities:** Enabling rapid identification of equity gaps and actionable insights for improvement. By simplifying implementation and enhancing transparency, this approach aims to make equity measurement practical and impactful for healthcare organizations.

Figure 4. Basic flow chart of a user workflow using our dashboard.

Methods

After collecting data from both open source and private datasets we derive pap smear rates and cervical cancer mortality rates based off the 45+ female population of each county and the number of procedures performed per county, respectively. We built a regression model that predicts cervical cancer mortality rates based on racial demographics and conditions.

$\text{Mortality Risk Score}_{county} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$ where X_1 is Hispanic American Population X_2 is Rural/Urban Ratio X_3 is Asian American Population X_4 is Early-Stage Diagnostic Rate X_5 is African American Population X_6 is Smoking Rate

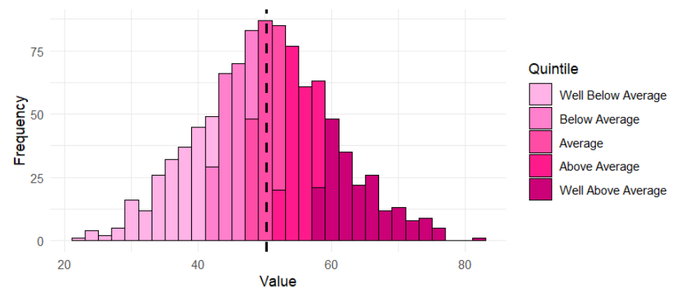
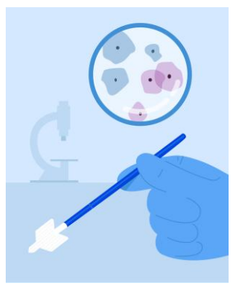


Figure 5. A visual aid for how the categorization is done

Increases Mortality Risk
- African American Population
- Hispanic American Population
- Smoking Rate
- Rural/Urban Ratio

Decrease Mortality Risk
- Early-Stage Diagnostic Rate
- Asian American Population



Visualizations

Using our dashboard is easy! There are 4 steps: locate → filter → explain → act. Let's say you wanted to plan a cervical cancer screening event but did not know where to start. We can go step by step to interact with the dashboard and filter on the desired output.

Unfiltered map and table show risk score for each county

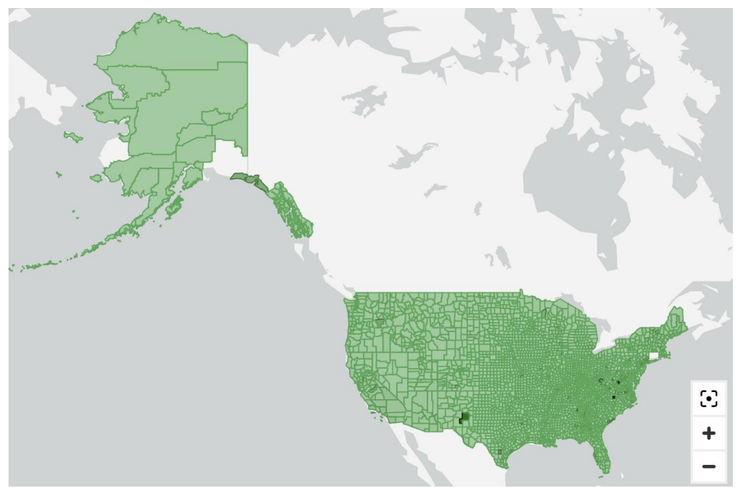


Figure 7. Spatial patterns of risk scores at the U.S. county level and social indicators.

Our model creates a mortality risk score based on several cervical cancer associated variables:
- Smoking Rates
- Area Deprivation Index (ADI)
- Rural/Urban
- Female & Racial Proportions

25 Segments can be created based on the criteria below:

state: All
county: All
risk category: All
intervention category: All

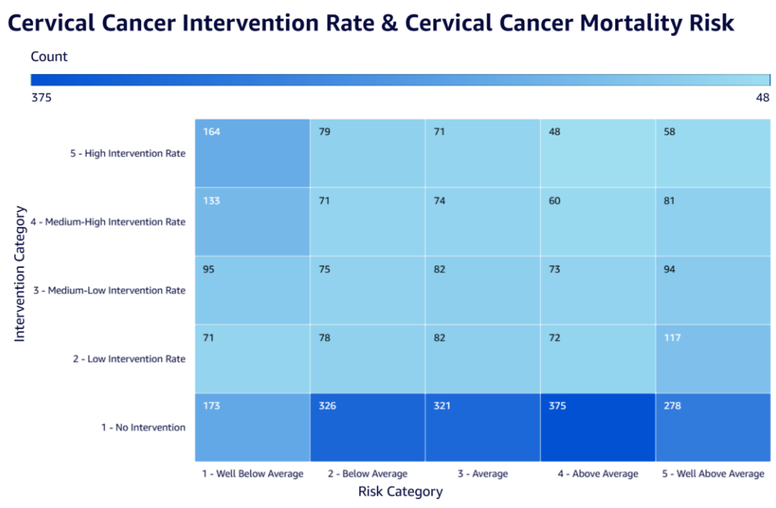
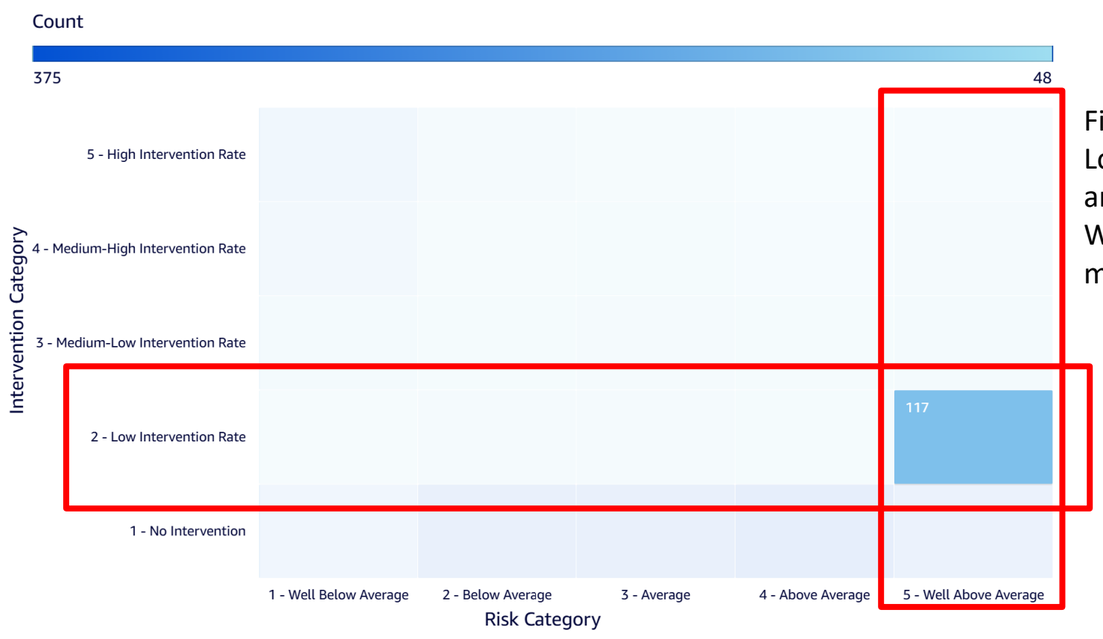


Figure 6. An unfiltered heatmap of cervical cancer risk on the x-axis and intervention rate on the y-axis. (No Intervention means no cervical cancer testing and screening data in our data source.)

Select a risk category and intervention rate

Filtered heatmap show 117 US counties in selected risk level



Filter by
Low Intervention Rate
and
Well Above Average
mortality risk

Figure 8. A filtered view of Risk Category-"Well Above Average" and Intervention Category-"Low Intervention Rate".

Filter by Low Intervention Rate and Well Above Average mortality risk

Filtered map show risk scores for counties in selected risk level

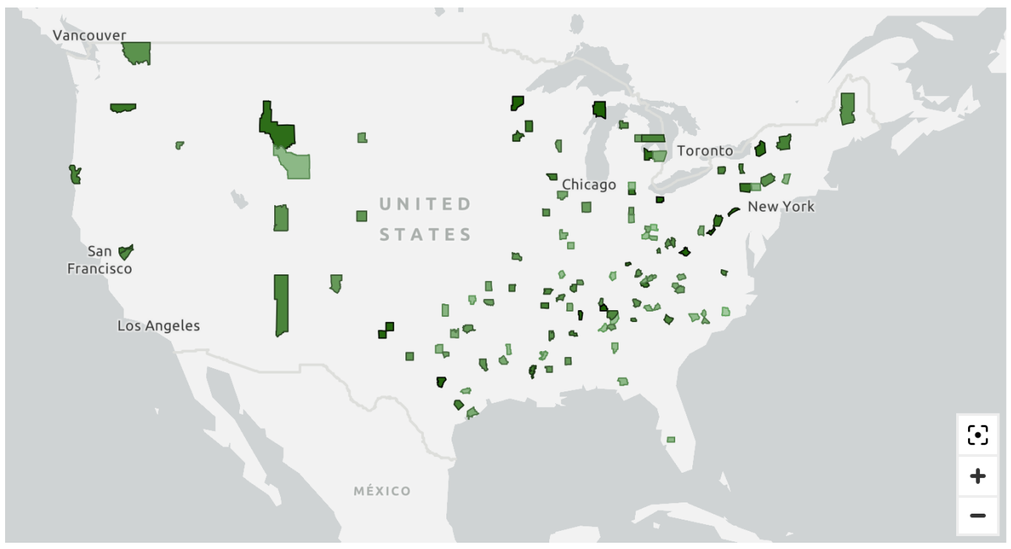


Figure 9. A filtered view of counties and their information filtered to Risk Category-"Well Above Average" and Intervention Category-"Low Intervention Rate" across the contiguous United States.

Conclusions

Our framework is designed to overcome the limitations of traditional equity models by **making health access measurements predictive, actionable , and easy to use.** Traditional frameworks often stop at identifying disparities, providing a snapshot of historical trends without offering guidance on subsequent actions. Our approach moves beyond description to prediction. By incorporating machine learning techniques, we predict risk in regions where it has not been reported using validated datasets, thereby extending risk characterization across the country. This predictive capability enables organizations to prioritize interventions proactively, ensuring resources are deployed where they will have the greatest impact. Health access challenges rarely stem from a single factor; they emerge from a complex interplay of clinical, social, and geographic determinants. To capture this complexity, our framework integrates **Electronic health records (EHR), claims data, national surveys, social determinants of health (SDOH) and geospatial information.**

By combining these data sources, we enable precise targeting of interventions and outreach programs, reducing the likelihood of overlooking vulnerable populations. With integrated use of this tool, we may be able to view how our efforts to promote cervical cancer screenings and pap smear procedures and awareness contribute to a decline in cervical cancer mortality.

References

1. HHS, CMS. (2022). CMS Framework for Health Equity 2022–2032.
2. NCQA. (2022). Advancing Health Equity: A Recommended Measurement Framework for Accountability in Medicaid.
3. Peterson, A. C. (2021). The Health Equity Framework: A Science- and Justice-Based Model for Public Health Researchers and Practitioners. *Health Promotion Practice*, 741-746.
4. The National Quality Forum. (2017). A Roadmap for Promoting Health Equity and Eliminating Disparities: The Four I's for Health Equity.

Contact Information

Ann Liu-Ferrara:
Ann.Liu-Ferrara@bd.com
Didier Morel:
Didier.Morel@bd.com
Walter Leo:
Walter.Leo@bd.com
Xueyin Zhao:
Xueying.Zhao@bd.com