# Analysis of Diagnostic Error Risk Factors and Resource Needs in Primary Care Using Bayesian Networks


Yukti Kathuria [1], Muge Capan PhD [1], Sonita Bennett [2], Kristen Miller [3], Elizabeth Selden [4], William Gallagher [3]

[1] Department of Mechanical and Industrial Engineering; [2] MedStar Center for Biostatistics, Informatics, and Data Science;
[3] MedStar Health Center for Diagnostic Systems Safety; [4] MedStar Georgetown University Hospital


## Background & Significance

- Healthcare systems are complex networks sharing limited resources (e.g., providers, equipment etc.)
- Primary care (PCP) serves as the initial point of contact for patients
- Resource-demand mismatch in primary care increases risk of diagnostic errors and lead to worse health outcomes for vulnerable populations [1].
- Patients with chronic conditions e.g., Type 2 Diabetes (T2D) are particularly impacted as their effective management relies heavily on primary care [2].

## Study Aims & Population

**Study Aims**

1. To mathematically demonstrate that two primary care locations within the same healthcare system serve significantly different patient populations.
2. To examine diagnostic error variation in PCP patients diagnosed with T2D depending on individual and geographical risk factors
3. To understand resource implications of populations at different diagnostic error risk

**Study Population:** Adult patients diagnosed with Type 2 Diabetes from 2 PCPs in the mid-Atlantic (N1 = 1758, N2 = 509)

**Outcome**: Diagnostic error criteria such as delay in time from first elevated hemoglobin A1C lab measurement to diagnosis, next lab measurement, next PCP visit etc. (categorized as 1 = at least one delay, 0 otherwise).

## Methods

- Statistical testing (t-test, chi-squared tests)
- Bayesian networks (BN) [3] are directed acyclical graphs G=(V,E) where V are the nodes and E are the edges of the graph that encode conditional dependencies as shown in Figure 1
  - 11 Nodes $(X_1, X_2, \ldots X_{11})$:
    - *Demographics* like sex, race, ethnicity, etc.
    - *Socioeconomic* such as employment, etc.
    - *Continuous* such as age, body mass index (BMI), etc.
    - *Outcome* such as diagnostic error (yes, no)

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Parents}(X_i))$$

*Figure 1: Joint Probability Distribution*

  - All nodes are discrete (continuous variables were discretized)
- Greedy search algorithm (hill climbing) used to learn the graph structure using scoring method like Bayesian Information Criteria

## Results

*Table 1: Study Population Characteristics*

| Study Characteristics | PCP 1 (N, %) | PCP 2 (N, %) | p-value |
|---|---|---|---|
| Biological Sex | | | |
| Female | 1076 (61.2%) | 265 (57.9%) | 0.7427 (PCP1), **0.0142 (PCP2)** |
| Male | 682 (38.8%) | 244 (52.1%) | |
| Employment | | | |
| Employed | 369 (21.0%) | 239 (47%) | 0.3076 (PCP1), **0.0068 (PCP2)** |
| Unemployed | 513 (29.2%) | 173 (34%) | |
| Unknown | 876 (49.8%) | 97 (19%) | |
| Age | | | |
| Q1 | 447 (25.4%) | 128 (25.2%) | **0.0007 (PCP1)**, 0.1045 (PCP2) |
| Q2 | 453 (25.8%) | 129 (25.3%) | |
| Q3 | 447 (25.4%) | 142 (27.9%) | |
| Q4 | 411 (23.4%) | 110 (21.6%) | |
| Diagnostic Error (Y) | | | |
| Yes | 900 (58.8%) | 258 (50.7%) | |
| No | 858 (51.2%) | 251(49.3%) | |

## Results

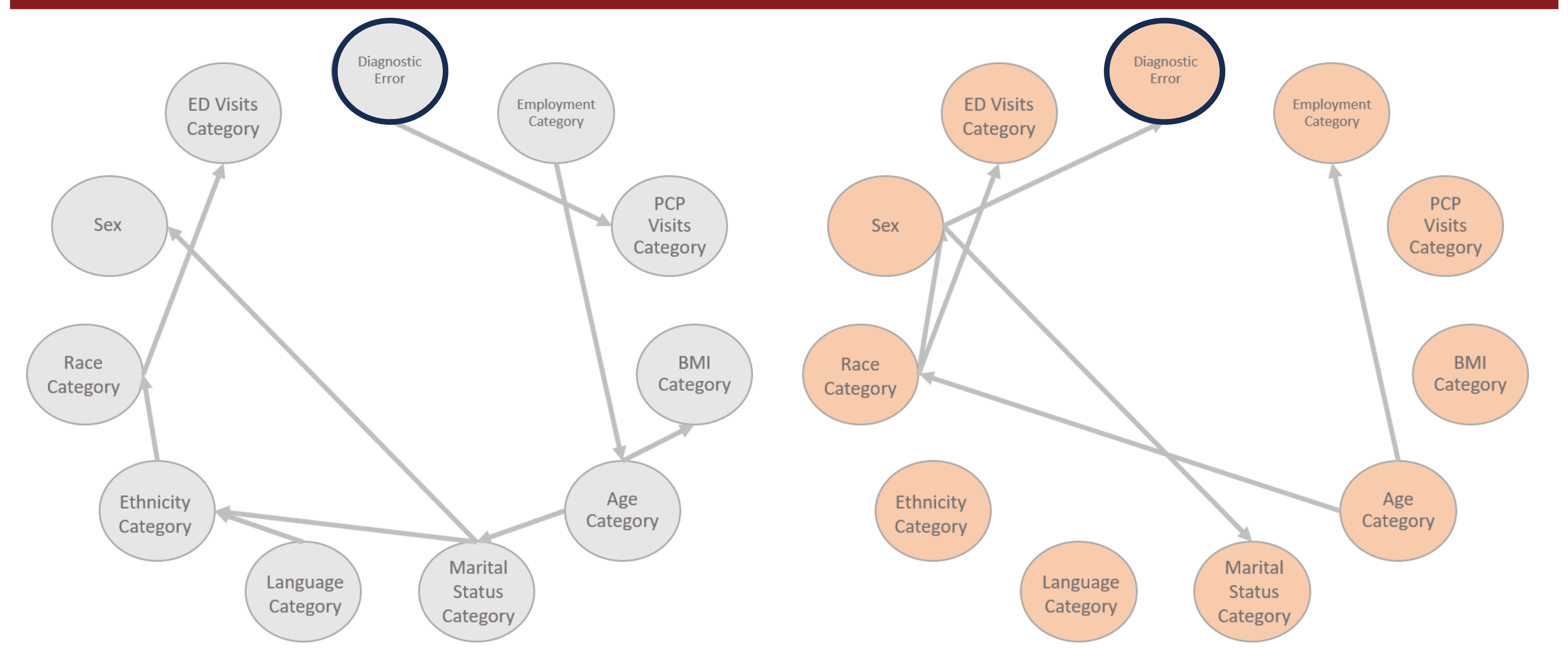


*Figure 2: Learned Bayesian Networks for PCP 1 and PCP 2*

- 55 maximum possible edges
- In PCP1, more edges are observed involving demographic factors (Ethnicity, Marital Status, Language)
- In PCP2, there are direct associations between Sex and Diagnostic Error, which are absent in PCP1

*Table 2: BN Comparison Metrics*

| Metric | BN for PCP1 | BN for PCP2 |
|---|---|---|
| Edges (shared) | 2 | 2 |
| Edges (not shared) | 7 | 4 |
| P(Y=1\| parents (Y)) | 0.5857 | 0.4533 |

## Conclusion & Future Work

- Performed statistical tests to analyze differences between PCP populations
- Developed BN models to highlight structural differences to compare 2 PCP locations
- Future Work:
  - Mathematically quantify BN differences
  - Incorporate continuous variables
  - Introduce hidden nodes for a robust and realistic model


**Contact:**
Yukti Kathuria
PhD Student, Department of Mechanical and Industrial Engineering
ykathuria@umass.edu



**References:**

1. Holmér, S., Nedlund, A.C., Thomas, K. et al. How health care professionals handle limited resources in primary care – an interview study. BMC Health Serv Res 23, 6 (2023). https://doi.org/10.1186/s12913-022-08996-y
2. Holt, T. A., Stokes, T., McKay, J., & Riley, R. D. (2016). Impact of primary care on delays in diagnosis and treatment of Type 2 Diabetes: A cohort study. British Journal of General Practice, 66(643), e114-e122. https://doi.org/10.3399/bjgp16X683485
3. Koller, D., & Friedman, N. (2009). Probabilistic Graphical Models: Principles and Techniques.