New Analysis Method Incorporating Bayesian Networks for HIV Surveillance Data in Investigating Factors Related to 'Out of Care' Status

Authors: Xichun Wang, Lei Zhang, Lourdes Yapjoco, Michael Johnson

(Steve) Xichun Wang, PhD, MPH

Public Health Informatics Scientist I

Southern Nevada Health District (SNHD)

Email: wangs@snhd.org

BACKGROUND

Step 1
- Diagnosis of HIV infection

Step 2
- Linkage to HIV medical care

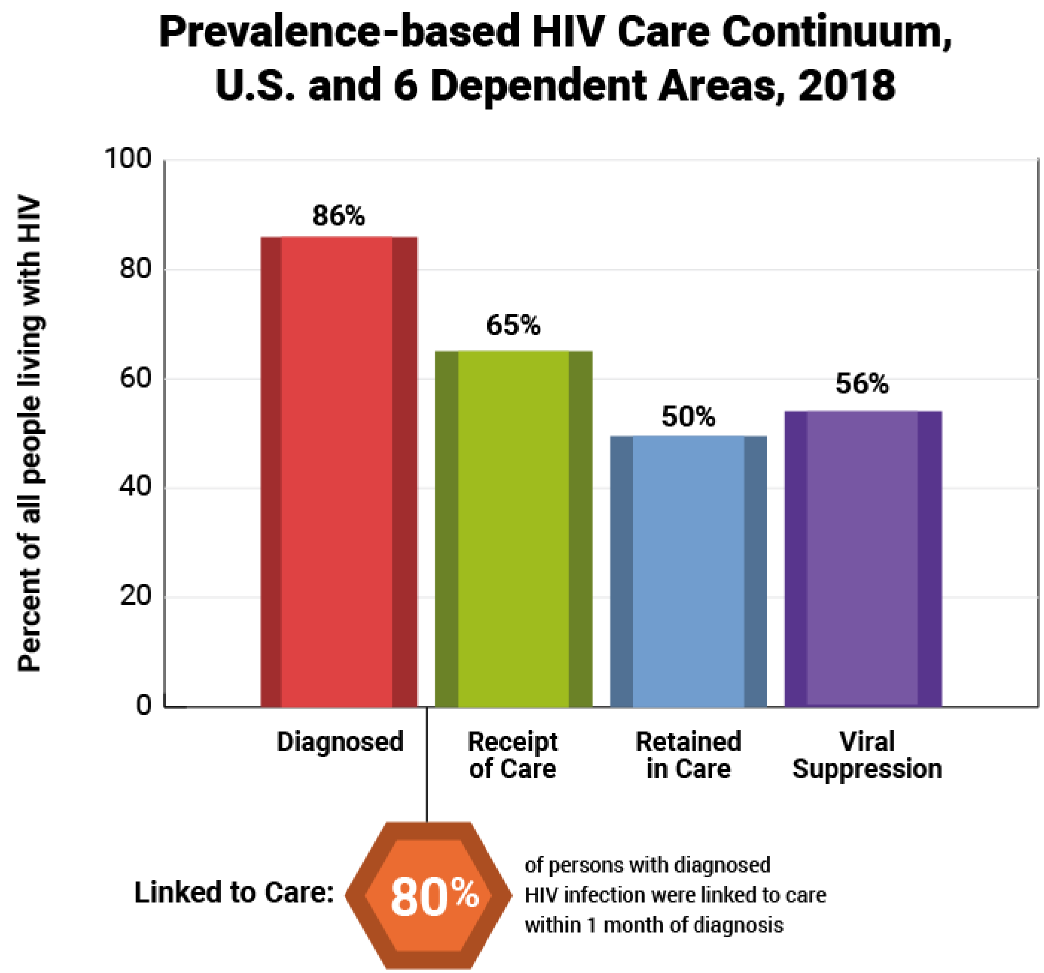
Step 3
- Receipt of HIV medical care

Step 4
- Retention in medical care

Step 5
- Achievement and maintenance of viral suppression

Source https://www.hiv.gov/federal-response/policies-issues/hiv-aids-care-continuum

BACKGROUND



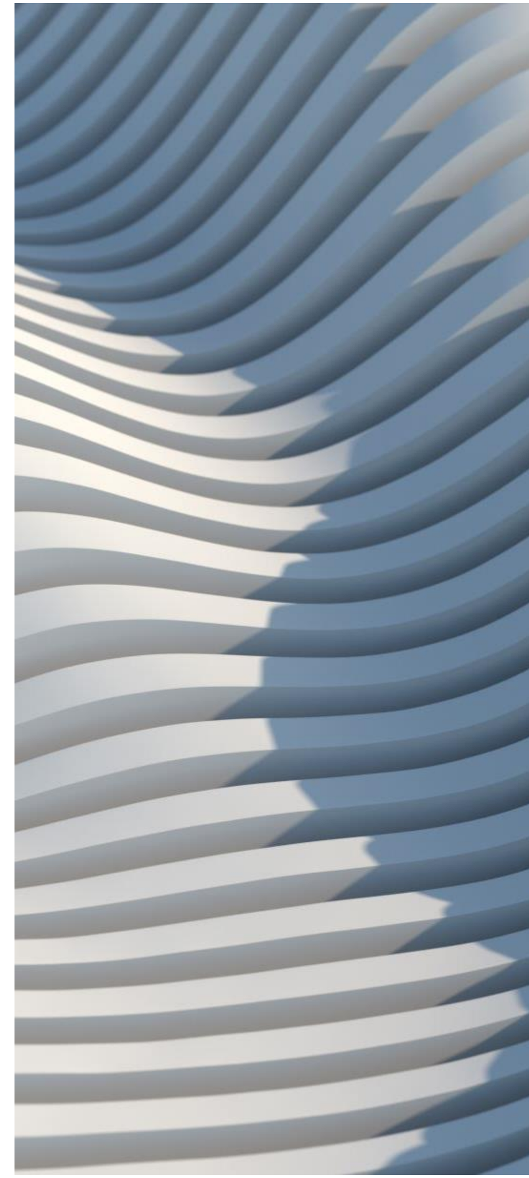
Note: Receipt of medical care was defined as ≥1 test (CD4 or VL) in 2016. Retained in medical care was defined as ≥ 2 tests (CD4 or VL) ≥ 3 months apart in 2016. Viral suppression was defined as < 200 copies/mL on the most recent test in 2016. Linkage to care is defined as having ≥ one CD4 or VL test within 30 days (1 month) of diagnosis. (Linkage is calculated differently from the other steps in the continuum, and cannot be directly compared to other steps.)

Source https://www.hiv.gov/federal-response/policies-issues/hiv-aids-care-continuum

Importance of Retention in Care

- Significantly lower viral loads
- Higher CD4 cell counts
- Reduced morbidity and mortality
- Increased safer sexual behaviors
- Lower rates of progression to AIDS
- Decreased rates of hospitalization
- Improved overall health

Source *https://www.hiv.gov/federal-response/policies-issues/hiv-aids-care-continuum*



Factors May Predict Patient Out of Care from Previous Studies

- Age
- Gender
- Race or ethnicity,
- Socioeconomic status
- Health insurance
- Medically insured via public health services
- Newly changed their insurance carrier
- Drug or alcohol dependence
- Mental Health Status

https://www.cdc.gov/hiv/clinicians/treatment/care-retention.html

Methods

Definition of Out of Care (OOC)

Out of Care

- ✓ Clients who currently reside in Clark County and have not had lab results reported in a 12-month period after a full public health case investigation

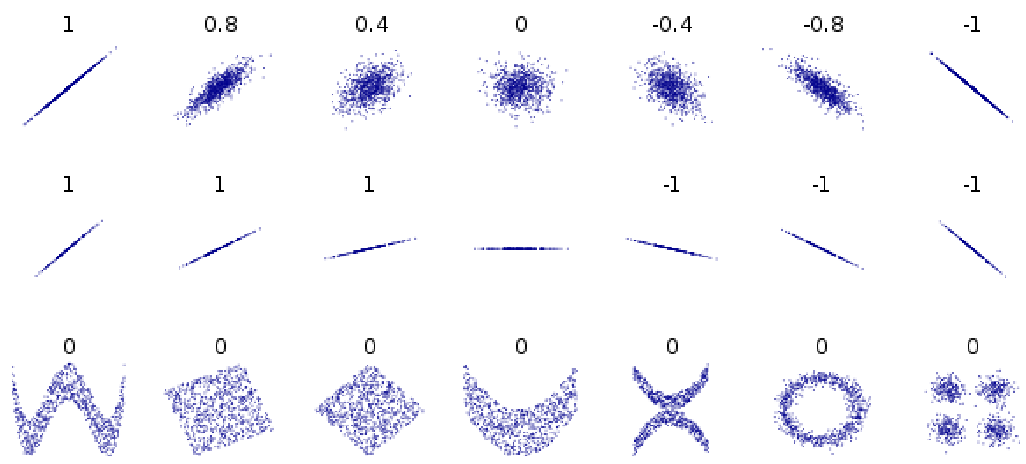
Data Source

CAREWare

- ✓ Data from CAREWare from 2018 to 2020 for people living with HIV (PLWHA) residing in Clark County

Linear model V.S. non-linear model

- **Correlation**: Statistical relationships involving dependence, commonly, two variables have a linear relationship with each other.
- **Dependence:** Any statistical relationship, whether causal or not, between two random variables.



Several sets of (x, y) points, with the Pearson correlation coefficient of x and y for each set.

Source: https://en.wikipedia.org/wiki/Correlation_and_dependence

Bayesian network for Non-Linear Modeling

A probabilistic graphical model that represents a set of variables and their conditional dependencies

Nodes correspond to random variables from the dataset

Edges represent conditional dependencies

nodes that are not connected represent variables that are conditionally independent of each other

Source: https://en.wikipedia.org/wiki/Bayesian_network

Methods

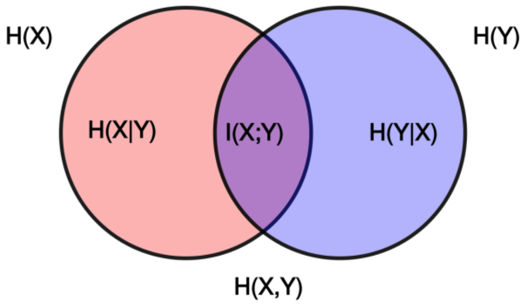
Apply Mutual Information to Relationship Scoring

1. For Discrete data:

$$I(X;Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right)$$

2. Continuous data — k-nearest neighbor methods:

$$\widehat{I}_{KSG}(X;Y) = \frac{1}{n}\sum_{i=1}^{n}\left\{\varphi\left(\kappa\right) + \log\left(n\right) + \log\left(\kappa_{x,i}\right) - \log\left(\kappa_{y,i}\right)\right\}$$



Pipeline Design

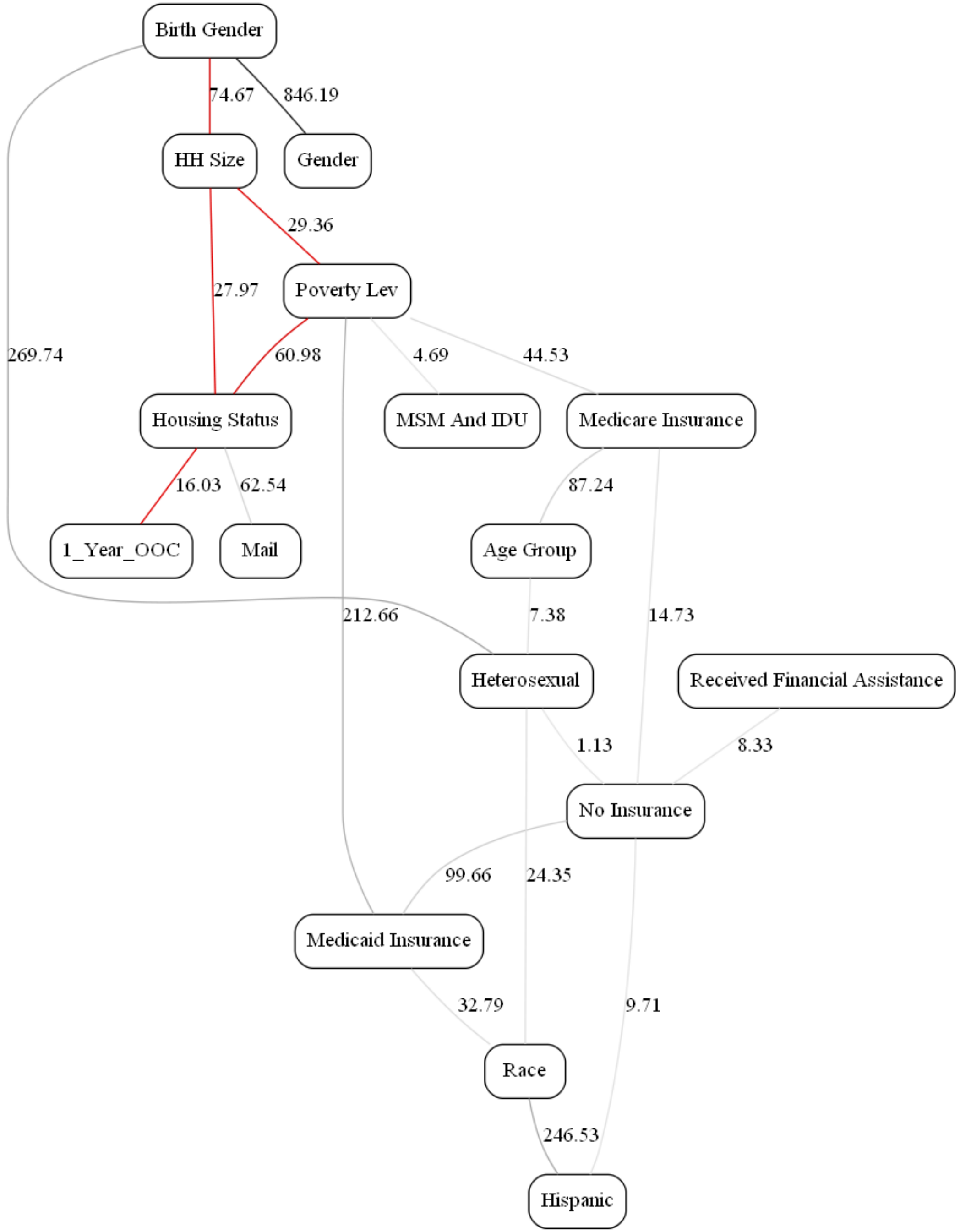
- STEP1: variable selection
- STEP2: construct Bayesian Networks (BNs)
- STEP3: sub-networks analysis

STEP1: variable selection

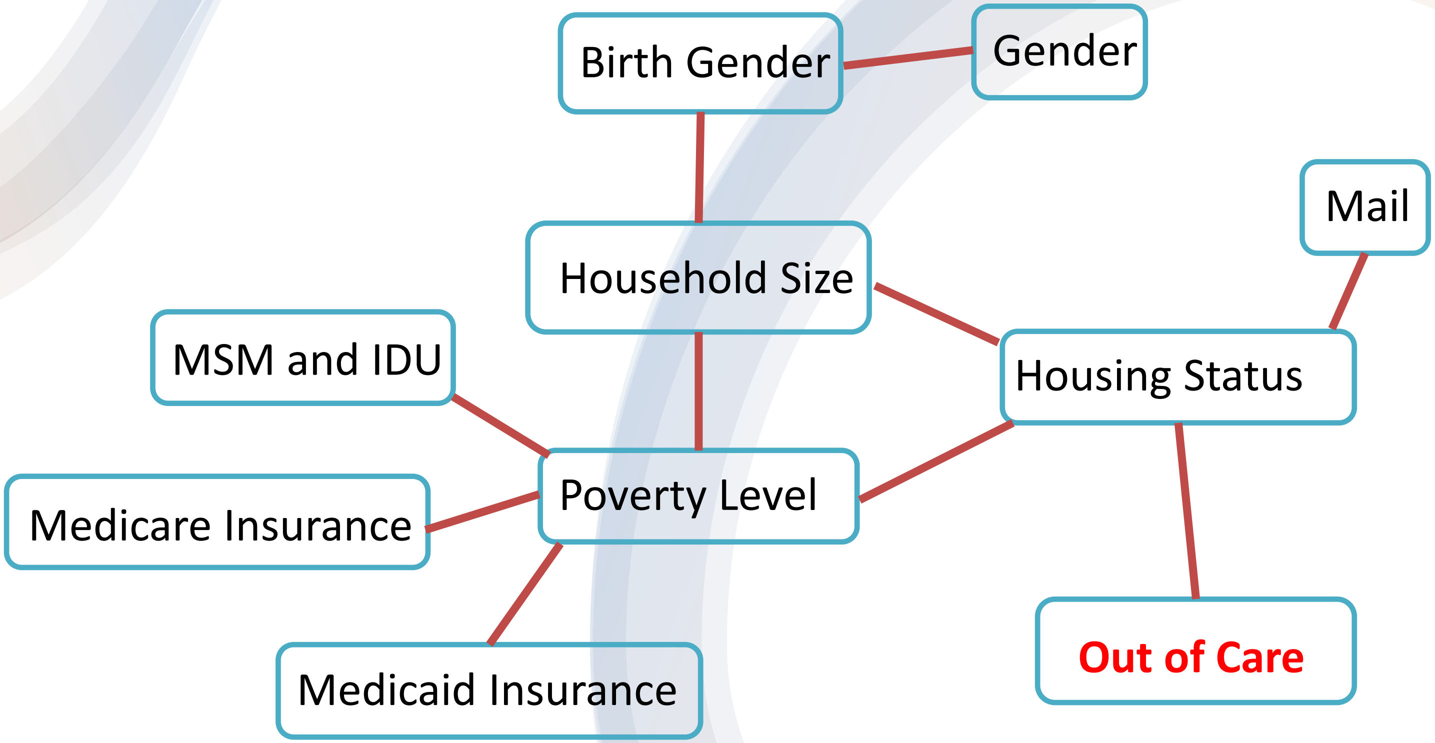
- Race
- Current Gender
- Age Group
- Household Size
- Poverty Level
- No Insurance
- Medicare Insurance
- Medicaid Insurance
- Hispanic
- Hemophiliac Risk Factor
- Heterosexual Risk Factor
- Transfusion Risk Factor
- Housing Status
- MSM And IDU Risk Factor
- Mail
- Birth Gender
- Received Financial Assistant

Race: Am. Indian/Alaska Native, Asian, Black or African-American, More than one race, Not Specified, Pacific Islander, White
Current Gender: Male, Female, Transgender MtF, Transgender FtM, Transgender Other
Age Group: Less than 2, 2 to 12, 13 - 24, 25 - 44, 45 - 64, Greater than 65
Housing Status: Unstable, Stable/Permanent, Temporary

Fig 1. Full Bayesian Networks derived from the 1-year OOC data



Number of patient In Care: 1017
Number of patient OOC: 999



Nearest Markov Neighborhood Around Variables of Interest

Validation

P-values from the Chi-Square test

Variables	Edge Value	P-Value
Medicaid Insurance V.S. Race	2.79	4.88E-23
Poverty Lev V.S. MSM And IDU	4.69	0.025
Age Group V.S. Heterosexual	7.38	5.45E-13
Heterosexual V.S. No Insurance	7.48	3.79E-08
No Insurance V.S. Hispanic	9.71	3.04E-11
Housing Status V.S. 2_Years_OOC	11.99	1.11E-11
No Insurance V.S. Rcvd Fin Ass Service	18.16	5.61E-10
Medicare Insurance V.S. No Insurance	21.3	2.90E-13
Heterosexual V.S. Race	24.35	4.11E-15
HH Size V.S. Housing Status	27.97	3.42E-229
HH Size V.S. Poverty Lev	29.36	0
Poverty Lev V.S. Medicare Insurance	44.53	1.25E-27
Poverty Lev V.S. Housing Status	60.98	1.47E-259
Housing Status V.S. Mail	62.54	2.81E-27
Sex At Birth V.S. HH Size	74.67	3.31E-47
Medicare Insurance V.S. Age Group	87.24	1.86E-72
No Insurance V.S. Medicaid Insurance	99.66	5.58E-14
Poverty Lev V.S. Medicaid Insurance	212.66	2.42E-65
Race V.S. Hispanic	246.53	2.69E-100
Sex At Birth V.S. Heterosexual	269.74	1.18E-123
2_Years_OOC V.S. 1_Year_OOC	295.24	6.11E-101
Sex At Birth V.S. Gender	846.19	0

Validation

P-values from the Chi-Square test

Variables	P-Value
Housing Status V.S. 2_Years_OOC	1.11E-11
Age Group V.S. 2_Years_OOC	8.63E-08
Poverty Lev V.S. 2_Years_OOC	1.17E-05
Race V.S. 2_Years_OOC	0.0001
Hispanic V.S. 2_Years_OOC	0.0139
HH Size V.S. 2_Years_OOC	0.0380
Mail V.S. 2_Years_OOC	0.0383
Sex At Birth V.S. 2_Years_OOC	0.0387
Medicaid Insurance V.S. 2_Years_OOC	0.1655
Gender V.S. 2_Years_OOC	0.5359
Rcvd Fin Ass Service V.S. 2_Years_OOC	0.6250
Medicare Insurance V.S. 2_Years_OOC	0.6379
Heterosexual V.S. 2_Years_OOC	0.7004
Hemophiliac V.S. 2_Years_OOC	0.7703
Transfusion V.S. 2_Years_OOC	0.8329
No Insurance V.S. 2_Years_OOC	0.8899
MSM And IDU V.S. 2_Years_OOC	0.9405

Conclusion

Housing Status is the KEY

Remedies for Patients with unstable housing (Examples)

National Alliance to End Homelessness

✓ **Housing First**

A homeless assistance approach that prioritizes providing permanent housing to people experiencing homelessness

assurance wireless

✓ **Assurance Wireless**

Provides eligible low-income free monthly data, unlimited texting, and free monthly minutes + a free phone.

Conclusion

Housing Status is the KEY

Remedies for Patients with unstable housing (Examples)

✓ **Nevada Homeless Alliance**

Program to end homelessness in Southern Nevada through advocacy, public awareness, education, and the coordination of services.

✓ **Social Worker Referral**

✓ **Routing Survey Question in Patient Visits**

Future Studies

Build a bigger network with more variables

Investigate other variable of Interest

Try other data system, e.g. Trisano or eHARS

Include discrete and continues Variables

Methods of building a network with Thousands of Variables

Vector X:

$V_1, \quad V_2 \quad \ldots \quad V_N$

p_1 $[V_{11},V_{11},V_{11}]$, $[V_{12},V_{12},V_{12}]\ldots[V_{1N},V_{1N},V_{1N}]$

p_2 $[V_{21},V_{21},V_{21}]$, $[V_{22},V_{22},V_{22}]\ldots[V_{2N},V_{2N},V_{2N}]$

…

p_T $[V_{T1},V_{T1},V_{T1}]$, $[V_{T2},V_{T2},V_{T2}]$, $[V_{TN},V_{TN},V_{TN}]$

Vector Y:

	Target
p_1	T_1
p_2	T_2
…	
p_T	T_T

Variable Selection

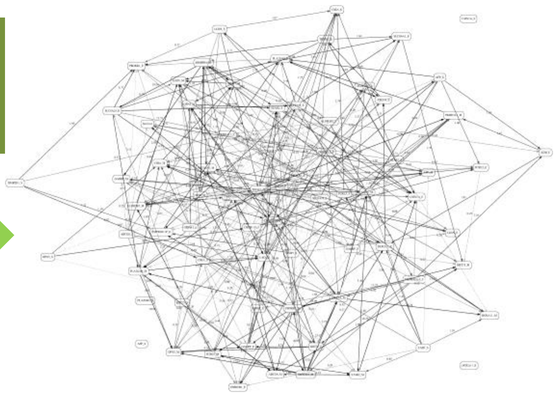
Combine table ($\geq$ 99.5% MI score) :

	G_1_V,	G_1_V,	G_1_V	…	G_N_V,	G_N_V,	G_N_V,	T
p_1	V_{11},	V_{12},	V_{13}	…	V_{1N},	V_{1N},	V_{1N}	T_1
p_2	V_{21},	V_{22},	V_{23}	…	V_{2N},	V_{2N},	V_{2N}	T_2
…								
p_T	V_{T1},	V_{T2},	V_{T3}	…	V_{TN},	V_{TN},	V_{TN}	T_T

Build BNs

For mixed data

Source: Wang, Xichun et al. "New Analysis Framework Incorporating Mixed Mutual Information and Scalable Bayesian Networks for Multimodal High Dimensional Genomic and Epigenomic Cancer Data." *Frontiers in genetics* vol. 11 648. 18 Jun. 2020, doi:10.3389/fgene.2020.00648



$$\widehat{I}(X;Y) = \frac{1}{n}\sum_{i=1}^{n}\left\{\varphi\left(\widetilde{\kappa}_i\right) + \log\left(n\right) + \log\left(\kappa_{x,i}\right) - \log\left(\kappa_{y,i}\right)\right\}$$

Reference: Publications about Retention in Care and Out of Care

- ✓ Missed Visits and Mortality in Patients Establishing Initial Outpatient HIV Treatment
  Mugavero, Michael J et al. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America vol. 48,2 (2009): 248-56. doi:10.1086/595705

- ✓ Retention in HIV Care: What the Clinician Needs to Know
  Giordano TP. Retention in HIV care: what the clinician needs to know. Top Antivir Med. 2011 Feb-Mar;19(1):12-6. PMID: 21852711; PMCID: PMC6148858.

- ✓ Out of Care" HIV Case Investigations: A Collaborative Analysis Across 6 States in the Northwest US.
  Dombrowski, Julia C et al. Journal of acquired immune deficiency syndromes (1999) vol. 74 Suppl 2,Suppl 2 (2017): S81-S87. doi:10.1097/QAI.0000000000001237

- ✓ Project Engage: Snowball Sampling and Direct Recruitment to Identify and Link Hard-to-Reach HIV-Infected Persons Who Are Out of Care.
  Wohl AR et al., J Acquir Immune Defic Syndr. 2017 Jun 1;75(2):190-197. doi: 10.1097/QAI.0000000000001312. PMID: 28169872

- ✓ HIV Care and Viral Load Suppression After Sexual Health Clinic Visits by Out-of-Care HIV-Positive Persons.
  Tymejczyk, Olga et al., AIDS patient care and STDs vol. 32,10 (2018): 390-398. doi:10.1089/apc.2018.00972020

- ✓ Retention in Care and Viral Load Improvement After Discharge Among Hospitalized Out-of-Care People With HIV Infection: A Post Hoc Analysis of a Randomized Controlled Trial.
  English, Kellee et al. Open forum infectious diseases vol. 7,6 ofaa193. 26 May. 2020, doi:10.1093/ofid/ofaa193

Reference: Publications about Bayesian network modeling

- ✓ Estimating mutual information for discrete-continuous mixtures.
  Gao, Weihao, et al. arXiv preprint arXiv:1709.06212 (2017)

- ✓ New Algorithm and Software (BNOmics) for Inferring and Visualizing Bayesian Networks from Heterogeneous Big Biological and Genetic Data.
  Gogoshin, Grigoriy et al. Journal of computational biology : a journal of computational molecular cell biology vol. 24,4 (2017): 340-356. doi:10.1089/cmb.2016.0100

- ✓ New Analysis Framework Incorporating Mixed Mutual Information and Scalable Bayesian Networks for Multimodal High Dimensional Genomic and Epigenomic Cancer Data.
  Wang, Xichun et al. Frontiers in genetics vol. 11 648. 18 Jun. 2020, doi:10.3389/fgene.2020.00648

Acknowledgment

Xandee Shirley Bernabe
Community Health Nurse II, SNHD
Email: bernabe@SNHD.ORG

Merylyn Yegon
CHN Supervisor, SNHD
Email: yegon@SNHD.ORG

Ronny Soy
Sr. Comm Health Nurse, SNHD
Email: soy@SNHD.ORG


Thank you !