Learning from Positive Deviance to Improve Colorectal Cancer Prevention in High-risk Communities: A Project Proposal

Johnie Rose, MD, PhD
CWRU Center for Community Health Integration
Population Cancer Analytics Shared Resource
Goals for Today

1. To provide an update on ongoing cancer informatics work occurring at CWRU/Case Comprehensive Cancer Center
2. To outline (and receive feedback on) a related recent initial R01 submission
1. A Multi-level Cancer Data Infrastructure
Johnie Rose
Siran Koroukian
Jackie Curtis
Weichuan Dong
Joseph Hnanth
Becky Karns
Uriel Kim
Abby Statler

Mustafa Ascha
Mark Beno
Wanda Lattimore
Harry Menegay
Paola Saroufim
Sunah Song
Devin Tian

Brian Christian
Adrienne Love
Jared Joslin

Funding Sources

Case Center for Reducing Health Disparities
Cleveland Translational Science Collaborative (CTSC; Parent UL1TR000439)
Susan G Komen of Northeast Ohio
Why Develop a Multi-level Cancer Data Infrastructure

• To identify variation in care processes and outcomes, and to understand what may be driving it, we need multi-level data
  • Molecular
  • Individual demographics
  • Clinical
  • Health services resources
  • Socioeconomic environment
  • Retail environment
  • Physical environment

• Efforts to bring these data sources together are often
  • *ad hoc*
  • Limited to research efforts
  • Performed with minimal input from the stakeholders who can help ask the right questions and provide the right interpretation.
Why Develop a Multi-level Cancer Data Infrastructure?

• To turn data into actionable knowledge about
  • Variation in cancer health services and outcomes
  • The disease, individual, health system, and community features that drive this variation

• To broaden the user base for valuable cancer and population data resources beyond the research community

• To facilitate translational and disparities research by enabling controlled linkage and sharing of data sources from multiple levels: from “cell to society”
The Cancer Informatics and Data Sharing Infrastructure (CIDaSh)

• A platform for secure integration, as well as controlled and efficient sharing of disparate sources of genomic, clinical, population, and health system data, thus creating a vast, evolving cancer data network across Ohio.

• Intended Users
  • Researchers
  • Cancer Centers
  • Advocacy and other community organizations
  • Public health

• The pilot project, NEO-CASE (Northeast Ohio Cancer Assessment and Surveillance Engine), is underway and in early use.
  • Funded through CTSA, institutional, and Komen Northeast Ohio pilot awards
NEO-CASE v1.0 Structure

**Individual-level Data**
- OCISS Cancer Registry
  - **Linkage:** Zip code, municipality, county

**Community data**
- Mammography Facilities
- Health professional shortage area (HRSA)
- Area Health Resource File (HRSA)
- American Community Survey Census Estimates
Available Geographies in NEO-CASE

State
Available Geographies in NEO-CASE

County
Available Geographies in NEO-CASE

Zip Code
Available Geographies in NEO-CASE

Municipality
Select population of interest

Describe how to aggregate your data.

Start by describing how to group the data (i.e. the categories over which to report numbers of cases):

**Geographic level**

- One group: everyone in Ohio

Next, choose how you want to filter the data to keep only certain patients:

**Date Range:**

- 01/01/06 to 31/12/15

**Rural Urban Continuum**

- [ ]

**County**

- [ ]

**Age at Diagnosis**

- [ ]

**Sex**

- [ ]

**Race**

- [ ]

**Ethnicity**

- [ ]

**Primary Site**

- [ ]

**SEER Summary Stage**

- [ ]

**Tumor Behavior**

- [ ]
<table>
<thead>
<tr>
<th>County</th>
<th>Crude Case Count</th>
<th>Cases: African American</th>
<th>%</th>
<th>Cases: Non-African American</th>
<th>%</th>
<th>Cases: Hispanic</th>
<th>%</th>
<th>Cases: Non-Hispanic</th>
<th>%</th>
<th>Cases: 15-29</th>
<th>%</th>
<th>Cases: 30-44</th>
<th>%</th>
<th>Cases: 45-54</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuyahoga</td>
<td>11516</td>
<td>3016</td>
<td>26.2</td>
<td>8500</td>
<td>73.8</td>
<td>132</td>
<td>1.1</td>
<td>11157</td>
<td>96.9</td>
<td>44</td>
<td>0.4</td>
<td>922</td>
<td>8.0</td>
<td>2168</td>
<td>18.8</td>
</tr>
<tr>
<td>Summit</td>
<td>4363</td>
<td>505</td>
<td>11.6</td>
<td>3858</td>
<td>88.4</td>
<td>18</td>
<td>0.4</td>
<td>3835</td>
<td>87.9</td>
<td>17</td>
<td>0.4</td>
<td>393</td>
<td>9.0</td>
<td>885</td>
<td>20.3</td>
</tr>
<tr>
<td>Stark</td>
<td>3020</td>
<td>181</td>
<td>6.0</td>
<td>2839</td>
<td>94.0</td>
<td></td>
<td></td>
<td>2966</td>
<td>98.2</td>
<td></td>
<td></td>
<td>252</td>
<td>8.3</td>
<td>506</td>
<td>16.8</td>
</tr>
<tr>
<td>Lorain</td>
<td>2356</td>
<td>143</td>
<td>6.1</td>
<td>2213</td>
<td>93.9</td>
<td>50</td>
<td>2.1</td>
<td>2204</td>
<td>93.5</td>
<td></td>
<td></td>
<td>231</td>
<td>9.8</td>
<td>423</td>
<td>18.0</td>
</tr>
<tr>
<td>Lake</td>
<td>2154</td>
<td>47</td>
<td>2.2</td>
<td>2107</td>
<td>97.8</td>
<td></td>
<td></td>
<td>2080</td>
<td>96.6</td>
<td></td>
<td></td>
<td>193</td>
<td>9.0</td>
<td>415</td>
<td>19.3</td>
</tr>
<tr>
<td>Mahoning</td>
<td>2151</td>
<td>285</td>
<td>13.2</td>
<td>1866</td>
<td>86.8</td>
<td>19</td>
<td>0.9</td>
<td>2015</td>
<td>93.7</td>
<td></td>
<td></td>
<td>143</td>
<td>6.6</td>
<td>371</td>
<td>17.2</td>
</tr>
<tr>
<td>Trumbull</td>
<td>1798</td>
<td>114</td>
<td>6.3</td>
<td>1684</td>
<td>93.7</td>
<td></td>
<td></td>
<td>1725</td>
<td>95.9</td>
<td></td>
<td></td>
<td>141</td>
<td>7.8</td>
<td>303</td>
<td>16.9</td>
</tr>
<tr>
<td>Medina</td>
<td>1425</td>
<td></td>
<td></td>
<td>1411</td>
<td>99.0</td>
<td>11</td>
<td>0.8</td>
<td>1364</td>
<td>95.7</td>
<td></td>
<td></td>
<td>143</td>
<td>10.0</td>
<td>323</td>
<td>22.7</td>
</tr>
<tr>
<td>Portage</td>
<td>1112</td>
<td>37</td>
<td>3.3</td>
<td>1075</td>
<td>96.7</td>
<td></td>
<td></td>
<td>1055</td>
<td>94.9</td>
<td></td>
<td></td>
<td>107</td>
<td>9.6</td>
<td>228</td>
<td>20.5</td>
</tr>
<tr>
<td>Wayne</td>
<td>850</td>
<td></td>
<td></td>
<td>840</td>
<td>98.8</td>
<td></td>
<td></td>
<td>830</td>
<td>97.6</td>
<td></td>
<td></td>
<td>67</td>
<td>7.9</td>
<td>150</td>
<td>17.6</td>
</tr>
</tbody>
</table>

Showing 1 to 10 of 22 entries
<table>
<thead>
<tr>
<th>County</th>
<th>Cases: 55-64</th>
<th>%</th>
<th>Cases: 65-74</th>
<th>%</th>
<th>Cases: 75-84</th>
<th>%</th>
<th>Cases: 85+</th>
<th>%</th>
<th>Cases: Local Stage</th>
<th>%</th>
<th>Cases: Regional Stage</th>
<th>%</th>
<th>Cases: Distant Stage</th>
<th>%</th>
<th>Unknown Stage</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuyahoga</td>
<td>2840</td>
<td>24.7</td>
<td>2568</td>
<td>22.3</td>
<td>2008</td>
<td>17.4</td>
<td>966</td>
<td>8.4</td>
<td>7025</td>
<td>61.0</td>
<td>3335</td>
<td>29.0</td>
<td>774</td>
<td>6.7</td>
<td>338</td>
<td>2.9</td>
</tr>
<tr>
<td>Summit</td>
<td>1059</td>
<td>24.3</td>
<td>998</td>
<td>22.9</td>
<td>697</td>
<td>16.0</td>
<td>314</td>
<td>7.2</td>
<td>2726</td>
<td>62.5</td>
<td>1251</td>
<td>28.7</td>
<td>247</td>
<td>5.7</td>
<td>117</td>
<td>2.7</td>
</tr>
<tr>
<td>Stark</td>
<td>742</td>
<td>24.6</td>
<td>748</td>
<td>24.8</td>
<td>521</td>
<td>17.3</td>
<td>242</td>
<td>8.0</td>
<td>1918</td>
<td>63.5</td>
<td>839</td>
<td>27.8</td>
<td>151</td>
<td>5.0</td>
<td>100</td>
<td>3.3</td>
</tr>
<tr>
<td>Lorain</td>
<td>630</td>
<td>26.7</td>
<td>539</td>
<td>22.9</td>
<td>379</td>
<td>16.1</td>
<td>146</td>
<td>6.2</td>
<td>1470</td>
<td>62.4</td>
<td>681</td>
<td>28.9</td>
<td>138</td>
<td>5.9</td>
<td>58</td>
<td>2.5</td>
</tr>
<tr>
<td>Lake</td>
<td>497</td>
<td>23.1</td>
<td>541</td>
<td>25.1</td>
<td>359</td>
<td>16.7</td>
<td>142</td>
<td>6.6</td>
<td>1391</td>
<td>64.6</td>
<td>567</td>
<td>26.3</td>
<td>131</td>
<td>6.1</td>
<td>53</td>
<td>2.5</td>
</tr>
<tr>
<td>Mahoning</td>
<td>546</td>
<td>25.4</td>
<td>491</td>
<td>22.8</td>
<td>428</td>
<td>19.9</td>
<td>166</td>
<td>7.7</td>
<td>1250</td>
<td>58.1</td>
<td>630</td>
<td>29.3</td>
<td>107</td>
<td>5.0</td>
<td>144</td>
<td>6.7</td>
</tr>
<tr>
<td>Trumbull</td>
<td>447</td>
<td>24.9</td>
<td>434</td>
<td>24.1</td>
<td>337</td>
<td>18.7</td>
<td>132</td>
<td>7.3</td>
<td>1108</td>
<td>61.6</td>
<td>520</td>
<td>28.9</td>
<td>103</td>
<td>5.7</td>
<td>59</td>
<td>3.3</td>
</tr>
<tr>
<td>Medina</td>
<td>347</td>
<td>24.4</td>
<td>334</td>
<td>23.4</td>
<td>179</td>
<td>12.6</td>
<td>92</td>
<td>6.5</td>
<td>903</td>
<td>63.4</td>
<td>420</td>
<td>29.5</td>
<td>62</td>
<td>4.4</td>
<td>37</td>
<td>2.6</td>
</tr>
<tr>
<td>Portage</td>
<td>306</td>
<td>27.5</td>
<td>265</td>
<td>23.8</td>
<td>156</td>
<td>14.0</td>
<td>49</td>
<td>4.4</td>
<td>690</td>
<td>62.1</td>
<td>342</td>
<td>30.8</td>
<td>61</td>
<td>5.5</td>
<td>14</td>
<td>1.3</td>
</tr>
<tr>
<td>Wayne</td>
<td>241</td>
<td>28.4</td>
<td>201</td>
<td>23.6</td>
<td>135</td>
<td>15.9</td>
<td>52</td>
<td>6.1</td>
<td>572</td>
<td>67.3</td>
<td>205</td>
<td>24.1</td>
<td>28</td>
<td>3.3</td>
<td>34</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Showing 1 to 10 of 22 entries
<table>
<thead>
<tr>
<th>County</th>
<th>Median Time to Treatment (Days)</th>
<th>Deceased</th>
<th>%</th>
<th>Median Follow-up Time (Months)</th>
<th>Crude Incidence</th>
<th>Age-Adjusted Incidence</th>
<th>CI: Low</th>
<th>CI: High</th>
<th>Total Population</th>
<th>Population: 14 and younger</th>
<th>%</th>
<th>Population: 15-29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuyahoga</td>
<td>34</td>
<td>2714</td>
<td>23.6</td>
<td>8.3</td>
<td>172</td>
<td>165</td>
<td>156</td>
<td>175</td>
<td>1272533</td>
<td>230847</td>
<td>18.1</td>
<td>90526</td>
</tr>
<tr>
<td>Summit</td>
<td>31</td>
<td>949</td>
<td>21.8</td>
<td>6.3</td>
<td>156</td>
<td>151</td>
<td>138</td>
<td>166</td>
<td>541592</td>
<td>99032</td>
<td>18.3</td>
<td>39616</td>
</tr>
<tr>
<td>Stark</td>
<td>33</td>
<td>674</td>
<td>22.3</td>
<td>5.4</td>
<td>156</td>
<td>146</td>
<td>130</td>
<td>163</td>
<td>375348</td>
<td>69215</td>
<td>18.4</td>
<td>27920</td>
</tr>
<tr>
<td>Lorain</td>
<td>32</td>
<td>458</td>
<td>19.4</td>
<td>8.4</td>
<td>154</td>
<td>149</td>
<td>132</td>
<td>170</td>
<td>301720</td>
<td>58191</td>
<td>19.3</td>
<td>22893</td>
</tr>
<tr>
<td>Lake</td>
<td>33</td>
<td>442</td>
<td>20.5</td>
<td>8.5</td>
<td>183</td>
<td>165</td>
<td>145</td>
<td>189</td>
<td>229850</td>
<td>40860</td>
<td>17.8</td>
<td>16144</td>
</tr>
<tr>
<td>Mahoning</td>
<td>35</td>
<td>540</td>
<td>25.1</td>
<td>6.1</td>
<td>176</td>
<td>155</td>
<td>136</td>
<td>177</td>
<td>237033</td>
<td>40549</td>
<td>17.1</td>
<td>17363</td>
</tr>
<tr>
<td>Trumbull</td>
<td>32</td>
<td>398</td>
<td>22.1</td>
<td>6.2</td>
<td>168</td>
<td>148</td>
<td>128</td>
<td>172</td>
<td>208792</td>
<td>36973</td>
<td>17.7</td>
<td>14827</td>
</tr>
<tr>
<td>Medina</td>
<td>27</td>
<td>275</td>
<td>19.3</td>
<td>7.8</td>
<td>162</td>
<td>161</td>
<td>137</td>
<td>190</td>
<td>173252</td>
<td>35132</td>
<td>20.3</td>
<td>13015</td>
</tr>
<tr>
<td>Portage</td>
<td>30</td>
<td>220</td>
<td>19.8</td>
<td>6.5</td>
<td>132</td>
<td>145</td>
<td>120</td>
<td>174</td>
<td>163387</td>
<td>26446</td>
<td>16.2</td>
<td>18028</td>
</tr>
<tr>
<td>Wayne</td>
<td>28</td>
<td>162</td>
<td>19.1</td>
<td>7.6</td>
<td>146</td>
<td>148</td>
<td>119</td>
<td>183</td>
<td>114750</td>
<td>23941</td>
<td>20.9</td>
<td>9824</td>
</tr>
<tr>
<td>County</td>
<td>Population: 30-44</td>
<td>%</td>
<td>Population: 45-54</td>
<td>%</td>
<td>Population: 55-64</td>
<td>%</td>
<td>Population: 65-74</td>
<td>%</td>
<td>Population: 75-84</td>
<td>%</td>
<td>Population: 85 and older</td>
<td>%</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------</td>
<td>----</td>
<td>-------------------</td>
<td>----</td>
<td>-------------------</td>
<td>----</td>
<td>-------------------</td>
<td>----</td>
<td>-------------------</td>
<td>----</td>
<td>----------------------</td>
<td>----</td>
</tr>
<tr>
<td>Cuyahoga</td>
<td>232446</td>
<td>18.3</td>
<td>190856</td>
<td>15.0</td>
<td>168959</td>
<td>13.3</td>
<td>98657</td>
<td>7.8</td>
<td>67851</td>
<td>5.3</td>
<td>33627</td>
<td>2.6</td>
</tr>
<tr>
<td>Summit</td>
<td>99754</td>
<td>18.4</td>
<td>82225</td>
<td>15.2</td>
<td>73631</td>
<td>13.6</td>
<td>41772</td>
<td>7.7</td>
<td>27153</td>
<td>5.0</td>
<td>12321</td>
<td>2.3</td>
</tr>
<tr>
<td>Stark</td>
<td>66467</td>
<td>17.7</td>
<td>55659</td>
<td>14.8</td>
<td>52052</td>
<td>13.9</td>
<td>32243</td>
<td>8.6</td>
<td>20317</td>
<td>5.4</td>
<td>9617</td>
<td>2.6</td>
</tr>
<tr>
<td>Lorain</td>
<td>56932</td>
<td>18.9</td>
<td>46174</td>
<td>15.3</td>
<td>40785</td>
<td>13.5</td>
<td>24422</td>
<td>8.1</td>
<td>14349</td>
<td>4.8</td>
<td>6050</td>
<td>2.0</td>
</tr>
<tr>
<td>Lake</td>
<td>42209</td>
<td>18.4</td>
<td>36686</td>
<td>16.0</td>
<td>32529</td>
<td>14.2</td>
<td>20292</td>
<td>8.8</td>
<td>12727</td>
<td>5.5</td>
<td>5016</td>
<td>2.2</td>
</tr>
<tr>
<td>Mahoning</td>
<td>40834</td>
<td>17.2</td>
<td>35160</td>
<td>14.8</td>
<td>34742</td>
<td>14.7</td>
<td>20629</td>
<td>8.7</td>
<td>14717</td>
<td>6.2</td>
<td>7589</td>
<td>3.2</td>
</tr>
<tr>
<td>Trumbull</td>
<td>36154</td>
<td>17.3</td>
<td>31303</td>
<td>15.0</td>
<td>30750</td>
<td>14.7</td>
<td>19486</td>
<td>9.3</td>
<td>12488</td>
<td>6.0</td>
<td>5297</td>
<td>2.5</td>
</tr>
<tr>
<td>Medina</td>
<td>33998</td>
<td>19.6</td>
<td>28385</td>
<td>16.4</td>
<td>23419</td>
<td>13.5</td>
<td>13631</td>
<td>7.9</td>
<td>7044</td>
<td>4.1</td>
<td>3070</td>
<td>1.8</td>
</tr>
<tr>
<td>Portage</td>
<td>27403</td>
<td>16.8</td>
<td>23974</td>
<td>14.7</td>
<td>20468</td>
<td>12.5</td>
<td>12329</td>
<td>7.5</td>
<td>6706</td>
<td>4.1</td>
<td>2485</td>
<td>1.5</td>
</tr>
<tr>
<td>Wayne</td>
<td>19933</td>
<td>17.4</td>
<td>16158</td>
<td>14.1</td>
<td>14674</td>
<td>12.8</td>
<td>9403</td>
<td>8.2</td>
<td>5404</td>
<td>4.7</td>
<td>2453</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Showing 1 to 10 of 22 entries
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuyahoga</td>
<td>33627</td>
<td>2.6</td>
<td>394639</td>
<td>31.0</td>
<td>877894</td>
<td>69.0</td>
<td>62778</td>
</tr>
<tr>
<td>Summit</td>
<td>12321</td>
<td>2.3</td>
<td>85331</td>
<td>15.8</td>
<td>456261</td>
<td>84.2</td>
<td>9194</td>
</tr>
<tr>
<td>Stark</td>
<td>9617</td>
<td>2.6</td>
<td>34293</td>
<td>9.1</td>
<td>341055</td>
<td>90.9</td>
<td>6295</td>
</tr>
<tr>
<td>Lorain</td>
<td>6050</td>
<td>2.0</td>
<td>31384</td>
<td>10.4</td>
<td>270336</td>
<td>89.6</td>
<td>25890</td>
</tr>
<tr>
<td>Lake</td>
<td>5016</td>
<td>2.2</td>
<td>10070</td>
<td>4.4</td>
<td>219780</td>
<td>95.6</td>
<td>8144</td>
</tr>
<tr>
<td>Mahoning</td>
<td>7589</td>
<td>3.2</td>
<td>40231</td>
<td>17.0</td>
<td>196802</td>
<td>83.0</td>
<td>11482</td>
</tr>
<tr>
<td>Trumbull</td>
<td>5297</td>
<td>2.5</td>
<td>19682</td>
<td>9.4</td>
<td>189110</td>
<td>90.6</td>
<td>2984</td>
</tr>
<tr>
<td>Medina</td>
<td>3070</td>
<td>1.8</td>
<td>3056</td>
<td>1.8</td>
<td>170196</td>
<td>98.2</td>
<td>2964</td>
</tr>
<tr>
<td>Portage</td>
<td>2485</td>
<td>1.5</td>
<td>8646</td>
<td>5.3</td>
<td>154741</td>
<td>94.7</td>
<td>2308</td>
</tr>
<tr>
<td>Wayne</td>
<td>2453</td>
<td>2.1</td>
<td>2589</td>
<td>2.3</td>
<td>112161</td>
<td>97.7</td>
<td>1814</td>
</tr>
</tbody>
</table>

Showing 1 to 10 of 22 entries
Some ongoing work using NEO-CASE in Cuyahoga County...
A Different View of Risk Factors: “Phenotyping” Approach

• Our knowledge of risk factors is usually regression-based: mostly independent effects

• We define phenotypes as sets of commonly co-occurring traits that can be used to categorize individuals

• Some sets can serve as markers of higher risk

• How to identify these phenotypes?
Clustering Algorithms

• Group of unsupervised machine learning methods allowing segmentation of observations into groups of greatest similarity to each other along multiple dimensions

• Simple Example: 2-deminsional clustering of individuals on a university campus
  • Older, high income
  • Younger, low income

• Use some measure of mathematical distance between observations to assign each observation to a cluster of similar observations.
  • Euclidian Distance – continuous variables
  • Gower’s Distance – categorical variables
  • …
Identifying “Phenotypes” associated with delayed treatment in Breast Cancer

These clusters associated with >60 days from biopsy to first treatment

<table>
<thead>
<tr>
<th>Phenotype (n=189)</th>
<th>Phenotype 2 (n=107)</th>
<th>Phenotype 3 (n=344)</th>
<th>Phenotype 4 (n=268)</th>
<th>Phenotype 5 (n=229)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>AA</td>
<td>AA</td>
<td>AA</td>
<td>AA</td>
</tr>
<tr>
<td>Young Age at Dx</td>
<td>Not Married</td>
<td>Young Age at Dx</td>
<td>Young Age at Dx</td>
<td>Married</td>
</tr>
<tr>
<td>Medicaid Insurance</td>
<td>Advanced cancer</td>
<td>Not Married</td>
<td>Not Married</td>
<td>Private insurance</td>
</tr>
<tr>
<td>High residential mobility</td>
<td>Medicare insurance</td>
<td>Non-advanced cancer</td>
<td>Private Insurance</td>
<td>Higher than average median income</td>
</tr>
<tr>
<td>High rental burden</td>
<td>High residential mobility</td>
<td>Low median household income</td>
<td>Higher than average median income</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No HH car</td>
<td>High rental burden</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High vacant housing units</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High unemployment rate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Identifying Phenotype 3 ‘Positive Deviants’
Next Steps
NEO-CASE v2.0 Structure

**Individual-level Data**
- Medicare Claims
- OCISS Cancer Registry
- Death Certs

**Community data**
- Colonoscopy Facilities
- Mammography Facilities
- Health professional shortage area (HRSA)
- Area Health Resource File (HRSA)
- BRFSS-based small area estimates for CRC screening
- American Community Survey Census Estimates

**Linkage:**
- Distance
- Block group, census tract, zip code, municipality, county
Our longer-term vision for the Cancer Informatics and Data Sharing (CIDaSh) data infrastructure

OCISS integrated with publicly available databases

Subset A
- Administrative claims data

Subset B
- Biospecimen data

Subset C
- Data from the Breast and Cervical Cancer Early Detection Program (BCCEDP)

Subset D
- Data from cancer patients navigation programs*

Subset E
- EHR Data
- Family Registry for people with Lynch Syndrome*

Subset F
- Data from the Breast and Cervical Cancer Early Detection Program (BCCEDP)

* Examples of future data sources

¶ e.g., U.S. Census, Area Health Resource File
2. Learning from Positive Deviance to Improve Colorectal Cancer Prevention in High-risk Communities
If successful...

• A model for deriving more locally actionable insight from cancer registry data
• A better understanding of barriers to and facilitators of CRC screening in higher-risk communities
• A community informed set of intervention/implementation strategies for increasing CRC screening uptake which better considers context
Significance

- Second leading cause of U.S. cancer death
- Mortality disparities across dimensions of
  - Race
  - Rurality
  - SES
- 7,690 deaths could be avoided annually if all Americans experienced the same CRC mortality rate as the most educated whites*
- Single largest determinant of cancer-specific survival: **Stage**
- Strong associations: Stage at diagnosis, screening uptake, subgroup membership

Screening Disparities

- **Race/Ethnicity***
  - NHW: 65.4%
  - NHB: 61.8%
  - AI/AN: 54.3%
  - Hispanic: 49.9%
  - Asian: 49.4%

- **Rurality†**
  - Urban: 54%
  - Remote rural: 45%

- **Education†**
  - College graduates: 71.3%
  - High school dropouts: 47.4%

- **Income‡‡**
  - >400% FPL: 70.0%
  - <138% FPL: 46.9%

* American Cancer Society. Colorectal Cancer Facts & Figures 2017
Interventions to Improve Screening Uptake

• Often in low-income and/or minority communities
• Tend to be clinic-based, targeting the patient, provider, or system levels
• Few have examined community-based interventions
• Few address context-specific implementation factors
• **Aim 1** – *Who, what, and where:* What groups are most at risk for diagnosis with metastatic colorectal cancer (CRC)? Where do they, cluster spatially, and which clusters represent positive deviants?

• **Aim 2** – *Why:* Why do patients in certain type of communities face barriers to screening completion? Why do subsets of these communities experience better outcomes?

• **Aim 3** – *How:* Tapping, the wisdom residing in higher-risk communities, how can we sustainably improve screening uptake in the real world?
Why the outcome of distant stage CRC?

• Difficult to obtain screening estimates below the county level for all areas
• Even if possible, those estimates tend to be based on political/census geographies
• Self-reported screening may overestimate completed screening
• 1/3 of positive stool-based screens are not followed up by diagnostic colonoscopy*, AND AA’s and other underserved groups are more likely to choose stool-based methods†.

Aim 1 - Identify spatial clusters of high-risk individuals with and without distant stage CRC

1. Sort all Northeast Ohio patients with CRC into sociodemographic phenotypes based on economic, demographic, and healthcare access characteristics

<table>
<thead>
<tr>
<th>Variable Category</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient demographics</td>
<td>Age, gender, race, ethnicity, marital status</td>
</tr>
<tr>
<td>Patient insurance status</td>
<td>Private insurance, Medicaid, Medicare, Military/VA, uninsured</td>
</tr>
<tr>
<td>Family history</td>
<td>Whether family history of cancer is known</td>
</tr>
<tr>
<td>Healthcare access indicators</td>
<td>Residence in a health professional shortage area, Volume-weighted(78) driving distance to nearest colonoscopy facility</td>
</tr>
<tr>
<td>Census block group socioeconomic indicators</td>
<td>Median income, education, employment (by industry), labor force participation, rural/urban indicators</td>
</tr>
<tr>
<td>Census block group household indicators</td>
<td>Family composition, household overcrowding, residential mobility, home ownership, rental burden, vacant housing</td>
</tr>
<tr>
<td>Census block group Transportation indicators</td>
<td>Mode of transport to job, automobile access, public transit access</td>
</tr>
</tbody>
</table>
Aim 1 - *Identify spatial clusters of high-risk individuals with and without distant stage CRC*

2. Identify three of the highest risk (for presenting with metastatic CRC) phenotypes and map the location of all patients belonging to each
   • At least one urban phenotype; at least one rural

3. Find spatial clusters of non-distant and distant stage patients for three sociodemographic phenotypes

<table>
<thead>
<tr>
<th>Sociodemographic Phenotype (A, B, or C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distant (D) or Non-distant (N) Stage</td>
</tr>
<tr>
<td>AD</td>
</tr>
<tr>
<td>AN</td>
</tr>
</tbody>
</table>
Aim 1 Methods Summary

1. Define sociodemographic phenotypes of CRC patients using non-spatial clustering algorithms

2a. Identify three sociodemographic phenotypes associated with high risk of distant stage CRC (non-spatial clustering)

2b. Select 3 high-risk sociodemographic phenotypes

3. Map separate spatial clusters of distant and non-distant stage patients from each phenotype

4. Sample two communities from each phenotype/stage group

- AD1
- AD2
- AN1
- AN2
- BD1
- BD2
- BN1
- BN2
- CD1
- CD2
- CN1
- CN2
Aim 2 – Identify the community characteristics and resources which promote or inhibit earlier detection of colorectal cancer in high-risk settings

• Spatial Video Geonarrative (SVG)
  • An approach to eliciting information from participants while they are immersed in their environments
  • Records the reflections and insights of an individual as he/she moves through a landscape
  • Memories and informed interpretation of events are triggered by the sights, sounds and smells of the surrounding environment
  • Subject comments captured by digital recorder, while the outside environment is videoed using a global positioning system (GPS) enabled camera
Aim 2 – Identify the community characteristics and resources which promote or inhibit earlier detection of colorectal cancer in high-risk settings

• Videomapper/Wordmapper software for SVG
  • Merges video stream, transcribed audio, and GPS path
  • Enables quantitative or qualitative analysis
  • Can map locations associated with keywords/themes
  • Can identify keywords/themes associated with locations
Aim 2 – *Identify the community characteristics and resources which promote or inhibit earlier detection of colorectal cancer in high-risk settings*

• Recruiting Subjects
  • Partners
    • Case Comprehensive Cancer Center
    • Neighborhood Leadership Institute (NLI)
    • Ohio Association of Community Health Centers (OACHC)
  • 4-5 community members/CRC survivors per spatial cluster
    • Recruitment network or contact through registry
  • 2-3 subjects who work with patients/clients in each area
    • Clinical
    • Social work
    • Advocacy
    • Public health
Aim 2 – *Identify the community characteristics and resources which promote or inhibit earlier detection of colorectal cancer in high-risk settings*

• **Expected topics**
  • Perceived risks to health
  • The extent and type of discussions in their community around health generally and cancer in particular
  • How information about health risks is relayed and what sources of information residents find credible
  • The priorities that compete with healthy behaviors or seeking healthcare
  • Norms for receiving regular medical care from a consistent source or provider
  • Initiatives/resources that have made a positive difference in the health of people in the community
  • Where the residents get their social support, strength, or hope
  • What about their community is good or strong, and what are the biggest problems
Aim 3 — *Partner with stakeholders in designing implementation strategies to improve CRC screening access and uptake in high-risk settings*

- We will adapt/expand a pre-existing system dynamics (SD) model built to examine the role of social marketing interventions on CRC screening uptake.
- First, some background...
System Dynamics

A set of system mapping and modeling tools and methods for using them to improve our understanding of the ways in which complex systems function and to use that understanding to design high leverage policies for success.

(adapted from Sterman, 2006)
System Dynamics

• Thinking about populations in aggregate

• Components
  • Stocks – levels of some important entity (infected people, total spending, degree of satisfaction...)
  • Flows – Rates of change in stocks (can be represented by differential equations)
  • Causal influences

• Feedback is a central concept.
Aim 3 — Partner with stakeholders in designing implementation strategies to improve CRC screening access and uptake in high-risk settings

• Extend the prior model to include:
  • Patient access to primary care
  • Availability of primary care providers and colonoscopy providers in the community
  • A two-stage screening process: stool-based screening followed up by colonoscopy if positive

• Based on the literature and on barriers and facilitators identified through Aim 2 SVGs, add influences affecting flows between stocks
  • Provider level
  • Community level
  • Health system level
  • Policy level
Aim 3 — Partner with stakeholders in designing implementation strategies to improve CRC screening access and uptake in high-risk settings

• Group model-model building (GMB) sessions with stakeholders – a series for each high-risk phenotype
  • Session 1 – Introducing conventions of SD, model review, and estimating parameters and their distribution
  • Session 2 – Review revised model and preliminary analysis of potential intervention strategies; Prioritize a set of intervention strategies based on implementability and impact (informed by simulation results).
  • Session 3 – Review revised model/results; Revised prioritization of strategies based on results; Develop action plan
    • Key outlets
    • Next steps for group members
    • External partners and resources needed (and how to engage)
    • Setting-specific risks
If successful...

• A model for deriving more locally actionable insight from cancer registry data
• A better understanding of barriers to and facilitators of CRC screening in higher-risk communities
• A community informed set of intervention/implementation strategies for increasing CRC screening uptake which better considers context
Questions?
Suggestions?

Johnie Rose, MD, PhD
Center for Community Health Integration
johnie.rose@case.edu
216-368-6860