addhaz: Contribution of chronic diseases to the disability burden in R

Renata Yokota
Caspar Looman, Wilma Nusselder, Herman Van Oyen, Geert Molenberghs

UseR!2017, Brussels
Motivation

Why is it important to better understand the disability burden?

- Population ageing worldwide
  - Life expectancy in Brazil: 51 years (1950); 75 years (2013)
- Burden of chronic diseases: main causes of disability
- Limitations in instrumental activities of daily living (IADL)

Social burden: ↓ quality of life and ↑ health care use and costs
Motivation

Mortality

- No longer sufficient to measure population health

Morbidity (chronic diseases, disability)

- Lack of standard assessment method

  Longitudinal studies

  - “Gold standard”
  - Expensive and limited sample size

  Cross-sectional studies

  1. Population attributable fraction (PAF)
  2. Average attributable fraction (AAF)
  3. Years lived with disability (YLD) — Global Burden of Diseases Study
  4. Attribution method
Objectives

To briefly introduce the attribution method and to illustrate the use of addhaz with examples, using the data from the Brazilian National Health Survey, 2013.
Attribution method

Rationale

- **Mortality**: one disease is assigned as underlying cause of death by the physician who fills in the death certificate.

- **Disability (attribution method)**: to attribute the disability cases reported in a survey to a single cause or combination of causes (diseases).

- Partition of disability into additive contributions of causes.

- Takes into account:
  - multimorbidity: individuals can have more than one disease.
  - disability can be present in individuals without chronic conditions (“background”).

---

Attribution method

Background

- Even if a person has only one disease: not necessarily the cause of the disability
- Disability can occur without any disease
  - Physiological changes due to ageing
- Underreported and underdiagnosed diseases in the survey
- Important causes of disability not included in the survey
Methods

Assumptions

- Causal relationship between diseases and disability
- The estimated cross-sectional cumulative rates reflect the transitions rates that would have been estimated with longitudinal data (stationary assumption)
- The recovery rate is zero
- The ratio of the cause-specific cumulative rates to the overall cumulative rate is constant of over time (proportionality assumption)
- Diseases and background act as independent competing causes of disability
Attribution method

Binomial additive hazard model

\[ Y_i \sim \text{Bernoulli}(\pi_i) \]
\[ \pi_i = 1 - \exp(-\eta_i) \]
\[ \eta_i = \alpha_a + \sum_{d=1}^{m} \beta_{ad}(X_{di}X_{ai}) \]

- **\( Y_i \)**: binary response variable (disability) for each individual \( i \)
- **\( \pi_i \)**: estimated probability that individual \( i \) has disability
- **\( \eta_i \)**: linear predictor for each individual \( i \) (overall cumulative disability hazard rate)
- **\( \alpha_a \)**: cumulative hazard of disability for background by age group \( a (a = 1, \ldots, k) \)
- **\( \beta_{ad} \)**: cumulative hazard of disability for disease \( d (d = 1, \ldots, m) \) (disabling impacts) by age group \( a \)
- **\( X_{ai} \)**: indicator variable for age group \( a \) and individual \( i \)
- **\( X_{di} \)**: indicator variable for disease \( d \) and individual \( i \)
Attribution method

Multinomial additive hazard model

\[ Y_{ij} \sim \text{Multinomial}(n_i, \pi_{ij}) \]

\[ \pi_{ij} = \left[ 1 - \exp\left(- \sum_{q=1}^{c} \eta_{iq}\right) \right] \left( \frac{\eta_{ij}}{\sum_{q=1}^{c} \eta_{iq}} \right) \]

\[ \eta_{ij} = \alpha_{aj} + \sum_{d=1}^{m} \beta_{adj}(X_{di}X_{ai}) \]

- **\( Y_{ij} \):** multinomial response variable (disability) for each individual \( i \)
- **\( \pi_{ij} \):** probability that individual \( i \) has disability for each \( j \) category of the outcome
- **\( \eta_{ij} \):** overall cumulative hazard rate of disability for each individual \( i \) for each \( j \) category of the outcome (linear predictor)
- **\( \alpha_{aj} \):** background cumulative rate for each age group \( a(a = 1, \ldots, k) \) for disability category \( j \)
- **\( \beta_{adj} \):** cumulative rate for disability category \( j \) for each diseases \( d(d = 1, \ldots, m) \) (disabling impact) and age group \( a \)
- **\( X_{di} \):** indicator variable for each condition \( d \) and each individual \( i \)
- **\( X_{ai} \):** indicator variable for each age group \( a \) and individual \( i \)
Methods

Contribution of diseases to the disability prevalence - Binomial

1. Probability of having disability by cause

\[
\hat{D}_i = \left[ \frac{\beta_{ad}(x_{di}x_{ai})}{\eta_i} \right] \pi_i \quad \hat{B}_i = \left( \frac{\alpha_a}{\eta_i} \right) \pi_i
\]

2. Number of individuals with disability by cause

\[
\hat{N}_d = \sum_{i=1}^{n} \hat{D}_i \quad \hat{N}_b = \sum_{i=1}^{n} \hat{B}_i
\]

3. Prevalence of disability by cause

\[
\hat{\text{Prev}}_d = \frac{\hat{N}_d}{n} \quad \hat{\text{Prev}}_b = \frac{\hat{N}_b}{n}
\]
**R package addhaz**

**Constrained optimization**

- Convergence problems: non-canonical link function
- Linear inequality constraint: R function `constrOptim`
  - Binomial: \( B = \{ \beta : x_i' \beta \geq 0, \forall i = 1, \ldots, n \} \)
  - Multinomial: \( B = \{ \beta : x_i' \beta_j \geq 0, \forall i = 1, \ldots, n \} \)
- Requirement: initial values inside the parameter space

**Bootstrap confidence intervals**

- Cumulative hazards rates of disability
- Contributions
- Call to `boot` package
  - Parallel option: speeds up computation
Application Data

- Brazilian National Health Survey, 2013
- N (women ≥ 60 years) = 6,294
- Disability: Instrumental Activities of Daily Living (IADL)

“How much difficulty do you have in …?”

- going shopping
- handling finances
- taking own medications
- going to the doctor
- using transportation

Response options

1. Unable
2. A lot of difficulty
3. Some difficulty
4. No difficulty
Application

Data

- Brazilian National Health Survey, 2013
- N (women ≥ 60 years) = 6,294
- Disability: Instrumental Activities of Daily Living (IADL)
  
  “How much difficulty do you have in ...?”
  
  - going shopping
  - handling finances
  - taking own medications
  - using transportation
  - going to the doctor

Response options (Binomial)

1. Unable
2. A lot of difficulty \( \rightarrow \) 1 - Disability
3. Some difficulty
4. No difficulty \( \rightarrow \) 0 - No disability
Application

Data

- Brazilian National Health Survey, 2013
- N (women $\geq 60$ years) = 6,294
- Disability: Instrumental Activities of Daily Living (IADL)

“How much difficulty do you have in ...?”

- going shopping
- handling finances
- taking own medications
- using transportation
- going to the doctor

Response options (Multinominal)

1. Unable
2. A lot of difficulty
3. Some difficulty
4. No difficulty

$1 -$ Mild disability
$2 -$ Severe disability
$0 -$ No disability
Application

Covariates

- Age: 0 (60-79 years); 1 (≥ 80 years)
- 5 chronic conditions: diabetes, arthritis, stroke
Application

Binomial model 1

```r
R> model1 <- BinAddHaz(dis.bin ~ diab + arth + stro, data = disbData,
                      weights = wgt, attrib = TRUE, type.attrib = "both")
R> summary(model1)

$call
BinAddHaz(formula = dis.bin ~ diab + arth + stro, data = disbData, weights = wgt,
           attrib = TRUE, type.attrib = "both")

$bootstrap
[1] FALSE

$coefficients
[1] " Estimate  StdErr  t.value  p.value "
[2] "(Intercept) 0.2970833 0.0094264  31.5161 < 2.2e-16 ***"
[3] "diab     0.1331831 0.0238217   5.5908  2.355e-08 ***"
[4] "arth     0.1306445 0.0222037   5.8839  4.213e-09 ***"
[5] "stro    0.5927519 0.0756976   7.8305  5.663e-15 ***"
[6] "---"
[7] "Signif. codes:  0 '***'  0.001 '**'  0.01 '*'  0.05 '.'  0.1 ' '  1"
attr("class")
[1] "summary.binaddhazmod"
```
Binomial model 1 - Contributions

\[ R \texttt{ model1$contribution} \]

\[
\texttt{\$att.rel}
\]

<table>
<thead>
<tr>
<th>Contribution</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>backgrnd</td>
<td>0.80405374</td>
</tr>
<tr>
<td>diab</td>
<td>0.06938567</td>
</tr>
<tr>
<td>arth</td>
<td>0.07451155</td>
</tr>
<tr>
<td>stro</td>
<td>0.05204903</td>
</tr>
</tbody>
</table>

\[
\texttt{\$att.abs}
\]

<table>
<thead>
<tr>
<th>Contribution</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>disab</td>
<td>0.30853338</td>
</tr>
<tr>
<td>backgrnd</td>
<td>0.24807742</td>
</tr>
<tr>
<td>diab</td>
<td>0.02140780</td>
</tr>
<tr>
<td>arth</td>
<td>0.02298930</td>
</tr>
<tr>
<td>stro</td>
<td>0.01605886</td>
</tr>
</tbody>
</table>
Application

Multinomial model

```r
R> disease <- as.matrix(disabData[, c("diab", "arth", "stro")])
R> head(disease)

diab  arth  stro
36 1 0 0
98 0 0 0
113 0 1 1
347 1 0 0
352 1 0 0
436 0 0 0

R> model2 <- MultAddHaz(dis.mult ~ factor(age) - 1 +
+  diseases:factor(age), data = disablingData, weights = wgt,
+  start = TRUE, start.val = c(rep(0.5, 8), rep(0.3, 8)),
+  attrib.var = age, attrib = TRUE, type.attrib = "both",
+  set.seed = TRUE, seed = 111, bootstrap = TRUE,
+  nbootstrap = 1000, parallel = TRUE, type.parallel = "snow",
+  ncpus = 4)
```
Application

Multinomial model (Cont.)

R> cbind(model2$coefficients, model2$ci)

<table>
<thead>
<tr>
<th></th>
<th>Coefficients</th>
<th>CI2.5</th>
<th>CI97.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor(age)01</td>
<td>0.118777413</td>
<td>0.10084729</td>
<td>0.13757177</td>
</tr>
<tr>
<td>factor(age)11</td>
<td>0.278112342</td>
<td>0.20586839</td>
<td>0.37416105</td>
</tr>
<tr>
<td>factor(age)0:diseasesdiab1</td>
<td>0.003884386</td>
<td>-0.03457998</td>
<td>0.04912954</td>
</tr>
<tr>
<td>factor(age)1:diseasesdiab1</td>
<td>-0.024085843</td>
<td>-0.17326054</td>
<td>0.15932837</td>
</tr>
<tr>
<td>factor(age)0:diseasesarth1</td>
<td>0.011934258</td>
<td>-0.02439410</td>
<td>0.05150632</td>
</tr>
<tr>
<td>factor(age)1:diseasesarth1</td>
<td>0.117057591</td>
<td>-0.08191648</td>
<td>0.31068100</td>
</tr>
<tr>
<td>factor(age)0:diseasesstro1</td>
<td>0.035815272</td>
<td>-0.03703729</td>
<td>0.14437896</td>
</tr>
<tr>
<td>factor(age)1:diseasesstro1</td>
<td>-0.025366501</td>
<td>-0.22643583</td>
<td>0.23889701</td>
</tr>
<tr>
<td>factor(age)02</td>
<td>0.109223363</td>
<td>0.09134694</td>
<td>0.13203584</td>
</tr>
<tr>
<td>factor(age)12</td>
<td>0.668467722</td>
<td>0.54173157</td>
<td>0.83466951</td>
</tr>
<tr>
<td>factor(age)0:diseasesdiab2</td>
<td>0.120686329</td>
<td>0.06711944</td>
<td>0.17812519</td>
</tr>
<tr>
<td>factor(age)1:diseasesdiab2</td>
<td>0.298938314</td>
<td>-0.09581495</td>
<td>0.81820286</td>
</tr>
<tr>
<td>factor(age)0:diseasesarth2</td>
<td>0.055391778</td>
<td>0.01137376</td>
<td>0.10035500</td>
</tr>
<tr>
<td>factor(age)1:diseasesarth2</td>
<td>0.657383862</td>
<td>0.31908734</td>
<td>1.05974398</td>
</tr>
<tr>
<td>factor(age)0:diseasesstro2</td>
<td>0.461533638</td>
<td>0.25233461</td>
<td>0.72703221</td>
</tr>
<tr>
<td>factor(age)1:diseasesstro2</td>
<td>1.211655291</td>
<td>0.49043852</td>
<td>2.27159856</td>
</tr>
</tbody>
</table>
### Multinomial model (Cont.)

```
R> fit3$contribution[[2]]

<table>
<thead>
<tr>
<th></th>
<th>Contribution</th>
<th>CI2.5</th>
<th>CI97.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>disab1</td>
<td>0.1155774093</td>
<td>0.1152298272</td>
<td>0.1159763662</td>
</tr>
<tr>
<td>backgrnd11</td>
<td>0.1117723939</td>
<td>0.1117475557</td>
<td>0.1117939601</td>
</tr>
<tr>
<td>factor(age)0:diseasdiab11</td>
<td>0.0006712916</td>
<td>0.0006051811</td>
<td>0.0007454994</td>
</tr>
<tr>
<td>factor(age)0:diseasarth11</td>
<td>0.0022903903</td>
<td>0.0020812734</td>
<td>0.0025196930</td>
</tr>
<tr>
<td>factor(age)0:diseasstro11</td>
<td>0.0008433334</td>
<td>0.0005840998</td>
<td>0.0011509320</td>
</tr>
<tr>
<td>disab12</td>
<td>0.2547833194</td>
<td>0.2504305042</td>
<td>0.2597823421</td>
</tr>
<tr>
<td>backgrnd12</td>
<td>0.2409059079</td>
<td>0.2401282149</td>
<td>0.2415867590</td>
</tr>
<tr>
<td>factor(age)1:diseasdiab12</td>
<td>-0.0035318634</td>
<td>-0.0047468931</td>
<td>-0.0023745757</td>
</tr>
<tr>
<td>factor(age)1:diseasarth12</td>
<td>0.0179766315</td>
<td>0.0132721585</td>
<td>0.0232856447</td>
</tr>
<tr>
<td>factor(age)1:diseasstro12</td>
<td>-0.000673566</td>
<td>-0.0009474658</td>
<td>-0.0002662531</td>
</tr>
<tr>
<td>disab11</td>
<td>0.1377962224</td>
<td>0.1340952549</td>
<td>0.1417504047</td>
</tr>
<tr>
<td>backgrnd11</td>
<td>0.0951301736</td>
<td>0.0948755326</td>
<td>0.0953621675</td>
</tr>
<tr>
<td>factor(age)0:diseasdiab11</td>
<td>0.0203187966</td>
<td>0.0184200024</td>
<td>0.0224523101</td>
</tr>
<tr>
<td>factor(age)0:diseasarth11</td>
<td>0.0100280692</td>
<td>0.0091522835</td>
<td>0.0109775362</td>
</tr>
<tr>
<td>factor(age)0:diseasstro11</td>
<td>0.0122625830</td>
<td>0.0092980735</td>
<td>0.0158258748</td>
</tr>
<tr>
<td>disab12</td>
<td>0.5277177920</td>
<td>0.5142920672</td>
<td>0.5409109020</td>
</tr>
<tr>
<td>backgrnd12</td>
<td>0.3842856363</td>
<td>0.3776051135</td>
<td>0.3910465118</td>
</tr>
<tr>
<td>factor(age)1:diseasdiab12</td>
<td>0.0323951992</td>
<td>0.0250947889</td>
<td>0.0419637842</td>
</tr>
<tr>
<td>factor(age)1:diseasarth12</td>
<td>0.0776952995</td>
<td>0.0639938909</td>
<td>0.0924141151</td>
</tr>
<tr>
<td>factor(age)1:diseasstro12</td>
<td>0.0324416570</td>
<td>0.0205413063</td>
<td>0.0466561612</td>
</tr>
</tbody>
</table>
```
Limitations

- Computation time
  - Problem with high dimensional data

- Causality assumption
  - Plausible: disease $\rightarrow$ disability
  - Cross-sectional data: disability incorrectly attributed to disease when disability $\rightarrow$ disease

Future research

- Multinomial model with ordinal responses
- Alternatives to reduce computation time
Thank you!
renata.yokota@wiv-isp.be