

InterVA

Samuel Clark

Bloomberg
Philanthropies



DATA FOR
HEALTH INITIATIVE



CDC Foundation

Together our impact is greater



Vital
Strategies



THE OHIO STATE
UNIVERSITY

INSTITUTE FOR
POPULATION RESEARCH

In collaboration with the WHO VA Reference Group
Columbus, Ohio
November, 2018

Outline

- 1 Background
- 2 Derive InterVA
- 3 Describe InterVA-4 Algorithm
- 4 Implications

Start with Naive Bayes Relationship ...

InterVA was developed by Peter Byass and his colleagues over many years [e.g. [9](#), [8](#), [14](#), [16](#), [10](#), [35](#), [18](#), [7](#)]

Start with Naive Bayes Relationship ...

InterVA was developed by Peter Byass and his colleagues over many years [e.g. 9, 8, 14, 16, 10, 35, 18, 7]

InterVA is widely used and has been validated in a variety of ways – many substantive publications that rely on InterVA [e.g.

4, 30, 14, 3, 35, 11, 17, 18, 27, 12, 20, 36, 2, 24, 15, 32, 31, 7, 37, 34, 22, 33, 1, 28, 21, 13, 23, 5, 19, 29, 26]

1 Background

2 Derive InterVA

3 Describe InterVA-4 Algorithm

4 Implications

Derive InterVA

Start with Naive Bayes Relationship,

$$\Pr(c|\mathbf{s}) = \frac{\Pr(c) \prod_s \Pr(s|c)^s (1 - \Pr(s|c))^{1-s}}{\sum_c \Pr(c) \prod_s \Pr(s|c)^s (1 - \Pr(s|c))^{1-s}} \quad (1)$$

$\Pr(c|\mathbf{s})$ is conditional probability of cause of death c given vector of symptoms \mathbf{s}

$\Pr(c)$ is probability of cause of death c regardless of symptoms – e.g. the **cause-specific mortality fractions**

$\prod_s \Pr(s|c)^s (1 - \Pr(s|c))^{1-s}$ is probability of vector of symptoms \mathbf{s} given cause of death c *assuming* independence among symptoms

$\sum_c \Pr(c) \prod_s \Pr(s|c)^s (1 - \Pr(s|c))^{1-s}$ is probability of symptom vector \mathbf{s} regardless of cause of death

Confronting the Reality of VA Data

The pragmatic reality of VA data is that each dataset includes a different subset of the complete set of possible symptoms, and some do not distinguish between the confirmed absence of a symptom and missing data.

Confronting the Reality of VA Data

The pragmatic reality of VA data is that each dataset includes a different subset of the complete set of possible symptoms, and some do not distinguish between the confirmed absence of a symptom and missing data.

To accommodate this common limitation of the data, **InterVA-4 only utilizes the confirmed presence of a symptom [6]**.

Naive Bayes without Absent Symptoms

Absent symptoms contribute to the factors raised to the $(1 - s)$ power in the naive Bayes relationship for VA,

$$\Pr(c|\mathbf{s}) = \frac{\Pr(c) \prod_s \Pr(s|c)^s (1 - \Pr(s|c))^{1-s}}{\sum_c \Pr(c) \prod_s \Pr(s|c)^s (1 - \Pr(s|c))^{1-s}}$$

Naive Bayes without Absent Symptoms

Absent symptoms contribute to the factors raised to the $(1 - s)$ power in the naive Bayes relationship for VA,

$$\Pr(c|\mathbf{s}) = \frac{\Pr(c) \prod_s \Pr(s|c)^s (1 - \Pr(s|c))^{1-s}}{\sum_c \Pr(c) \prod_s \Pr(s|c)^s (1 - \Pr(s|c))^{1-s}}$$

Ignoring the absent symptoms (or those for which no information exists) drops those factors and produces the InterVA-4 relationship [6]:

$$\Pr(c|\mathbf{s}) = \frac{\Pr(c) \prod_s \Pr(s|c)^s}{\sum_c \Pr(c) \prod_s \Pr(s|c)^s} \quad (2)$$

Final InterVA-4 Relationship

The full naive Bayes relationship in (Equation 1) uses all the information available for the death, i.e. all the elements of \mathbf{s} , whereas the InterVA-4 expression in Equation 2 uses only information on symptoms that are present, i.e. \mathbf{s}' , a subset of the elements of \mathbf{s} where $S' < S$.

Final InterVA-4 Relationship

The full naive Bayes relationship in (Equation 1) uses all the information available for the death, i.e. all the elements of \mathbf{s} , whereas the InterVA-4 expression in Equation 2 uses only information on symptoms that are present, i.e. \mathbf{s}' , a subset of the elements of \mathbf{s} where $S' < S$.

S' is different for each death because each death exhibits a different subset of the possible symptoms. InterVA-4 calculates the following death-specific 'propensity' (not a probability) for each cause of death, for each death d

$$\Pr(c|\mathbf{s}'_d) = \frac{\Pr(c) \prod_{s=1 \rightarrow S'} \Pr(s|c)^s}{\sum_c \Pr(c) \prod_{s=1 \rightarrow S'} \Pr(s|c)^s}, \quad (3)$$

where $s \in S'_d$.

1 Background

2 Derive InterVA

3 Describe InterVA-4 Algorithm

4 Implications

InterVA-4 Software

Two software implementations of InterVA-4

- 1 Peter Byass' compiled Windows code in versions 4.02 and 4.03 (**IVA4-B**):
<http://www.interva.net>
- 2 Our R package named **InterVA4** [25, 38] that replicates both versions 4.02 and 4.03 and is available as open source package on CRAN (**IVA4-R**):
<https://cran.r-project.org/package=InterVA4>

InterVA-4 Inputs

Input:

- \mathbf{S}_{ij} = data matrix consisting of binary values 'no'={no data or absent} and 'yes'=present for $i = 1$ to I deaths by $j = 1$ to 245 symptoms; corresponds to the IVA4-B input file **batchin.csv**
- \mathbf{c}'_k = vector of initial cause-specific probabilities $\Pr(\text{cause } k)$ for $k = 1$ to 60 causes, **built-in values**
- \mathbf{P}_{jk} = matrix of probabilities $\Pr(\text{symptom } j \mid \text{cause } k)$ for $j = 1$ to 245 symptoms by $k = 1$ to 60 causes, **built-in values**
- Malaria and HIV adjustment factors for 'very low', 'low', and 'high' prevalence of each

InterVA-4 Outputs

Output:

- \mathbf{C}_{ik} = matrix of probabilities $\Pr(\text{death } i \text{ caused by cause } k)$ for $i = 1$ to I deaths by $k = 1$ to 60 causes; corresponds to IVA4-B output file **memvars.csv**
- List of top 3 causes and corresponding likelihoods $(\gamma_{1i}, \gamma_{2i}, \gamma_{3i})$ for each death; corresponds to IVA4-B output file **valog.txt**
- Warnings generated by the algorithm; corresponds to IVA4-B output file **warnings.txt**

Sketch of InterVA-4 Algorithm I

```
1: for each death  $i = 1$  to  $I$  do
2:   Check that sex has a value
3:   Check consistency of  $S_{i*}$ 
4:   for each cause  $k = 1$  to 60 do
5:      $C_{ik} \leftarrow c'_k$ 
6:   end for
7:   for each symptom  $j = 1$  to 245 do
8:     if  $S_{ij} = \text{'yes'}$  then
9:       for each cause  $k = 1$  to 60 do
10:         $C_{ik} \leftarrow C_{ik} \times P_{jk}$ 
11:      end for
12:      Normalize  $C_{ik}$ 
13:      Truncate  $C_{ik}$  below 0.000001
14:    end if
15:  end for
16:  Identify top 3 causes or 'indeterimate'
17: end for
```

- 1 Background
- 2 Derive InterVA
- 3 Describe InterVA-4 Algorithm
- 4 Implications

InterVA-4 version 4.02 and 4.03

- Version 4.02 distributed by Byass has software bugs in the consistency check and normalization/truncation steps
- Version 4.03 fixes these bugs and releases source code in the form of a dBASE '.prg' program text file
- Our R version replicates both 4.02 and 4.03 exactly and **outputs propensities associated with all causes for each death, not just top three**

Symptom-Cause Information in InterVA-4

SCI for InterVA-4:

- \mathbf{c}' = vector of initial cause-specific probabilities $\Pr(\text{cause})$
- **Most important:** \mathbf{P} = matrix of $\Pr(\text{symptom}|\text{cause})$
- Malaria and HIV adjustment factors

Symptom-Cause Information in InterVA-4

SCI for InterVA-4:

- \mathbf{c}' = vector of initial cause-specific probabilities $\Pr(\text{cause})$
- **Most important:** \mathbf{P} = matrix of $\Pr(\text{symptom}|\text{cause})$
- Malaria and HIV adjustment factors

Consequently, no training data necessary

Symptom-Cause Information in InterVA-4

SCI for InterVA-4:

- \mathbf{c}' = vector of initial cause-specific probabilities $\Pr(\text{cause})$
- **Most important:** \mathbf{P} = matrix of $\Pr(\text{symptom}|\text{cause})$
- Malaria and HIV adjustment factors

Consequently, no training data necessary

However, training data *can* be used to calculate the conditional probabilities $\Pr(\text{symptom}|\text{cause})$ [25]

- 'gold standard' data, or
- De-biased physician-coded VAs

References I

- [1] Raymond A. Aborigo, Pascale Allotey, Paulina Tindana, Daniel Azongo, and Cornelius Debpuur. Cultural imperatives and the ethics of verbal autopsies in rural Ghana. *Global Health Action*, 6:1–11, 2013.
- [2] Evasius Bauni, Carolyn Ndila, George Mochamah, Gideon Nyutu, Lena Matata, Charles Ondieki, Barbara Mambo, Maureen Mutinda, Benjamin Tsofa, Eric Maitha, Anthony Etyang, and Thomas N. Williams. Validating physician-certified verbal autopsy and probabilistic modeling (InterVA) approaches to verbal autopsy interpretation using hospital causes of adult deaths. *Population Health Metrics*, 9, AUG 5 2011.
- [3] Jacqueline S. Bell, Moctar Ouedraogo, Rasmane Ganaba, Issiaka Sombie, Peter Byass, Rebecca F. Baggaley, Veronique Filippi, Ann E. Fitzmaurice, and Wendy J. Graham. The epidemiology of pregnancy outcomes in rural Burkina Faso. *Tropical Medicine & International Health*, 13(1):31–43, JUL 2008.
- [4] P Byass, DL Huong, and HV Minh. A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. *Scandinavian Journal of Public Health*, 31(62):32–37, 2003.
- [5] Peter Byass. Usefulness of the Population Health Metrics Research Consortium gold standard verbal autopsy data for general verbal autopsy methods. *BMC Medicine*, 12, FEB 4 2014.

References II

- [6] Peter Byass, Daniel Chandramohan, Samuel J Clark, Lucia D'Ambruoso, Edward Fottrell, Wendy J Graham, Abraham J Herbst, Abraham Hodgson, Sennen Hounton, Kathleen Kahn, et al. Strengthening standardised interpretation of verbal autopsy data: The new InterVA-4 tool. *Global Health Action*, 5, 2012.
- [7] Peter Byass, Daniel Chandramohan, Samuel J. Clark, Lucia D'Ambruoso, Edward Fottrell, Wendy J. Graham, Abraham J. Herbst, Abraham Hodgson, Sennen Hounton, Kathleen Kahn, Anand Krishnan, Jordana Leita, Frank Odhiambo, Osman A. Sankoh, and Stephen M. Tollman. Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool. *Global Health Action*, 5:1–8, 2012.
- [8] Peter Byass, Edward Fottrell, Dao Lan Huong, Yemane Berhane, Tumani Corrah, Kathleen Kahn, Lulu Muhe, et al. Refining a probabilistic model for interpreting verbal autopsy data. *Scandinavian journal of public health*, 34(1):26–31, 2006.
- [9] Peter Byass, Dao Lan Huong, and Hoang Van Minh. A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in vietnam. *Scandinavian Journal of Public Health*, 31(62 suppl):32–37, 2003.

References III

- [10] Peter Byass, Kathleen Kahn, Edward Fottrell, Mark A Collinson, and Stephen M Tollman. Moving from data on deaths to public health policy in agincourt, south africa: approaches to analysing and understanding verbal autopsy findings. *PLoS medicine*, 7(8):1073, 2010.
- [11] Peter Byass, Kathleen Kahn, Edward Fottrell, Mark A. Collinson, and Stephen M. Tollman. Moving from Data on Deaths to Public Health Policy in Agincourt, South Africa: Approaches to Analysing and Understanding Verbal Autopsy Findings. *PLoS Medicine*, 7(8), AUG 2010.
- [12] Peter Byass, Kathleen Kahn, Edward Fottrell, Paul Mee, Mark A. Collinson, and Stephen M. Tollman. Using verbal autopsy to track epidemic dynamics: the case of HIV-related mortality in South Africa. *Population Health Metrics*, 9, AUG 5 2011.
- [13] Nikita Desai, Lukasz Aleksandrowicz, Pierre Miasnikof, Ying Lu, Jordana Leita, Peter Byass, Stephen Tollman, Paul Mee, Dewan Alam, Suresh Kumar Rathi, Abhishek Singh, Rajesh Kumar, Faujdar Ram, and Prabhat Jha. Performance of four computer-coded verbal autopsy methods for cause of death assignment compared with physician coding on 24,000 deaths in low- and middle-income countries. *BMC Medicine*, 12, FEB 4 2014.

References IV

- [14] M Fantahun, E Fottrell, Y Berhane, S Wall, U Hogberg, and P Byass. Assessing a new approach to verbal autopsy interpretation in a rural Ethiopian community: the InterVA model. *Bulletin of the World Health Organization*, 84(3):204–210, MAR 2006.
- [15] Cleusa P. Ferri, Daisy Acosta, Mariella Guerra, Yueqin Huang, Juan J. Llibre-Rodriguez, Aquiles Salas, Ana Luisa Sosa, Joseph D. Williams, Ciro Gaona, Zhaorui Liu, Lisbeth Noriega-Fernandez, A. T. Jotheeswaran, and Martin J. Prince. Socioeconomic Factors and All Cause and Cause-Specific Mortality among Older People in Latin America, India, and China: A Population-Based Cohort Study. *PLoS Medicine*, 9(2), FEB 2012.
- [16] Edward Fottrell, Peter Byass, Thomas W Ouedraogo, Cecile Tamini, Adjima Gbangou, Issiaka Sombié, Ulf Högberg, Karen H Witten, Sohinee Bhattacharya, Teklay Desta, et al. Revealing the burden of maternal mortality: a probabilistic model for determining pregnancy-related causes of death from verbal autopsies. *Popul Health Metr*, 5(1):10–1186, 2007.
- [17] Edward Fottrell, Kathleen Kahn, Nawi Ng, Benn Sartorius, Dao Lan Huong, Hoang Van Minh, Mesganaw Fantahun, and Peter Byass. Mortality measurement in transition: proof of principle for standardised multi-country comparisons. *Tropical Medicine & International Health*, 15(10):1256–1265, OCT 2010.

References V

- [18] Edward Fottrell, Kathleen Kahn, Stephen Tollman, and Peter Byass. Probabilistic Methods for Verbal Autopsy Interpretation: InterVA Robustness in Relation to Variations in A Priori Probabilities. *PLoS ONE*, 6(11), NOV 3 2011.
- [19] Judith R. Glynn, Clara Calvert, Alison Price, Menard Chihana, Lackson Kachiwanda, Sebastian Mboma, Basia Zaba, and Amelia C. Crampin. Measuring causes of adult mortality in rural northern Malawi over a decade of change. *Global Health Action*, 7, 2014.
- [20] Abraham J. Herbst, Tshepiso Mafojane, and Marie-Louise Newell. Verbal autopsy-based cause-specific mortality trends in rural KwaZulu-Natal, South Africa, 2000-2009. *Population Health Metrics*, 9, AUG 5 2011.
- [21] Chifundo Kanjala, Denna Michael, Jim Todd, Emma Slaymaker, Clara Calvert, Raphael Isingo, Alison Wringe, Basia Zaba, and Mark Urassa. Using HIV-attributable mortality to assess the impact of antiretroviral therapy on adult mortality in rural Tanzania. *Global Health Action*, 7:1–8, 2014.
- [22] JC Leita, D Chandramohan, P Byass, R Jakob, K Bundhamcharoen, C Choprapowan, et al. Revising the WHO verbal autopsy instrument to facilitate routine cause-of-death monitoring. *Global Health Action*, 6(21518), 2013.

References VI

- [23] Jordana Leitaó, Nikita Desai, Lukasz Aleksandrowicz, Peter Byass, Pierre Miasnikof, Stephen Tollman, Dewan Alam, Ying Lu, Suresh Kumar Rathi, Abhishek Singh, Wilson Suraweera, Faujdar Ram, and Prabhat Jha. Comparison of physician-certified verbal autopsy with computer-coded verbal autopsy for cause of death assignment in hospitalized patients in low- and middle-income countries: systematic review. *BMC Medicine*, 12, FEB 4 2014.
- [24] Rafael Lozano, Michael K. Freeman, Spencer L. James, Benjamin Campbell, Alan D. Lopez, Abraham D. Flaxman, Christopher J. L. Murray, and PHMRC. Performance of InterVA for assigning causes of death to verbal autopsies: multisite validation study using clinical diagnostic gold standards. *Population Health Metrics*, 9, AUG 5 2011.
- [25] Tyler H McCormick, Zehang Richard Li, Clara Calvert, Amelia C Crampin, Kathleen Kahn, and Samuel J Clark. Probabilistic cause-of-death assignment using verbal autopsies. *Journal of the American Statistical Association*, 111(515):1036–1049, 2016.
- [26] Yohannes Adama Melaku, Berhe Weldearegawi Sahle, Fisaha Haile Tesfay, Afework Mulugeta Bezabih, Alemseged Aregay, Semaw Ferede Abera, Loko Abreha, and Gordon Alexander Zello. Causes of Death among Adults in Northern Ethiopia: Evidence from Verbal Autopsy Data in Health and Demographic Surveillance System. *PLoS ONE*, 9(9), SEP 4 2014.

References VII

- [27] Christopher J. L. Murray, Alan D. Lopez, Robert Black, Ramesh Ahuja, Said Mohd Ali, et al. Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets. *Population Health Metrics*S, 9, AUG 4 2011.
- [28] Christopher J. L. Murray, Rafael Lozano, Abraham D. Flaxman, Peter Serina, David Phillips, Andrea Stewart, et al. Using verbal autopsy to measure causes of death: the comparative performance of existing methods. *BMC Medicine*, 12, JAN 9 2014.
- [29] Carolyn Ndila, Evasius Bauni, Vysaul Nyirongo, George Mochamah, Alex Makazi, Patrick Kosgei, Gideon Nyutu, Alex Macharia, Sailoki Kapesa, Peter Byass, and Thomas N. Williams. Verbal autopsy as a tool for identifying children dying of sickle cell disease: a validation study conducted in Kilifi district, Kenya. *BMC Medicine*, 12, APR 22 2014.
- [30] Samuel O. Oti and Catherine Kyobutungi. Verbal autopsy interpretation: a comparative analysis of the InterVA model versus physician review in determining causes of death in the Nairobi DSS. *Population Health Metrics*S, 8, 2010.

References VIII

- [31] Heribert Ramroth, Eva Lorenz, Johanna C. Rankin, Edward Fottrell, Maurice Ye, Florian Neuhaus, Mark Ssennono, Ali Sie, Peter Byass, and Heiko Becher. Cause of death distribution with InterVA and physician coding in a rural area of Burkina Faso. *Tropical Medicine & International Health*, 17(7):904–913, JUL 2012.
- [32] Johanna C. Rankin, Eva Lorenz, Florian Neuhaus, Maurice Ye, Ali Sie, Heiko Becher, and Heribert Ramroth. Exploring the role narrative free-text plays in discrepancies between physician coding and the InterVA regarding determination of malaria as cause of death, in a malaria holo-endemic region. *Malaria Journal*, 11, FEB 21 2012.
- [33] Sebsibe Tadesse. Validating the InterVA Model to Estimate the Burden of Mortality from Verbal Autopsy Data: A Population-Based Cross-Sectional Study. *PLoS ONE*, 8(9), SEP 13 2013.
- [34] Sebsibe Tadesse and Takele Tadesse. Evaluating the performance of interpreting Verbal Autopsy 3.2 model for establishing pulmonary tuberculosis as a cause of death in Ethiopia: a population-based cross-sectional study. *BMC Public Health*, 12, NOV 29 2012.

References IX

- [35] Biruk Tensou, Tekebash Araya, Daniel S. Telake, Peter Byass, Yemane Berhane, Tolcha Kebebew, Eduard J. Sanders, and Georges Reniers. Evaluating the InterVA model for determining AIDS mortality from verbal autopsies in the adult population of Addis Ababa. *Tropical Medicine & International Health*, 15(5):547–553, MAY 2010.
- [36] Stefania Vergnano, Edward Fottrell, David Osrin, Peter N. Kazembe, Charles Mwansambo, Dharma S. Manandhar, Stephan P. Munjanja, Peter Byass, Sonia Lewycka, and Anthony Costello. Adaptation of a probabilistic method (InterVA) of verbal autopsy to improve the interpretation of cause of stillbirth and neonatal death in Malawi, Nepal, and Zimbabwe. *Population Health Metrics* S, 9, AUG 5 2011.
- [37] Yazoume Ye, Marilyn Wamukoya, Alex Ezeh, Jacques Emina, and Osman Sankoh. Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-Sahara Africa? *BMC public health*, 12(1):741, September 2012.
- [38] Zehang Richard Li and Tyler H. McCormick and Samuel J. Clark. Interva4: An R package to analyze verbal autopsy data. Technical report, Working paper no. 146, Center for Statistics and the Social Sciences, University of Washington, 2014.