

Tariff

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 Tariff
- 3 Similarities between Tariff and InterVA

Background

Tariff

- Developed by IHME [1] and is relatively new
- Was not used extensively in its original form, but is beginning to be used in an evolved form called 'Tariff 2.0' [2]
- IHME distributes propriety software called 'SmartVA-Analyze' that implements the Tariff 2.0 algorithm

1 Background

2 Derive Tariff

3 Similarities between Tariff and InterVA

Notation

- I causes of death indexed by i
- J binary-coded symptoms indexed by j
- K deaths indexed by k
- R binary-coded VA responses, one for each symptom j , indexed by r

NB: staying close-ish to the paper by James et al. [1]

PHRMC 'Gold-standard' data and Tariff Values

PHMRC 'gold-standard' data have both VA and medically certified cause of death

This contains information about the relationship between VA symptoms and medically certified causes

PHRMC 'Gold-standard' data and Tariff Values

PHMRC 'gold-standard' data have both VA and medically certified cause of death

This contains information about the relationship between VA symptoms and medically certified causes

Using the PHMRC data, calculate x_{ij} as the fraction of deaths with cause i and symptom j

$$x_{ij} = \frac{\sum_{k=1}^K r_{ij}}{K} \quad (1)$$

x_{ij} is proportional to the strength of relationship between cause i and symptom j

PHRMC *Gold-standard* data and Tariff Values

Calculate Tariff T_{ij} for each cause/symptom combination as x_{ij} standardized across cause

This ensures the T_{ij} are on the same scale within cause

$$T_{ij} = \frac{x_{ij} - \text{median}_i(x_{ij})}{IQR_i(x_{ij})} \quad (2)$$

Tariff Score for each Death

Identification of a cause for each death k begins by calculating a 'Tariff Score' for each cause i

$$S_{ki} = \sum_{j=1}^J T_{ij} \times r_{kj} \quad (3)$$

The Tariff Score is the sum of Tariffs for each death, one for each existing symptom

The Tariff Score reflects the joint relationship between existing symptoms and cause for each death, larger values indicating a stronger cumulative relationship across symptoms

Choosing a Cause

For each death, the most likely cause is the one with the highest Tariff Score, however ...

Choosing a Cause

For each death, the most likely cause is the one with the highest Tariff Score, however ...

Because each cause has a set of typical symptoms and those symptoms are more predictive for some causes than others, the Tariff Score that is “big” for one cause might not be very large for another. Thus, Tariff scores cannot be compared without standardizing or normalizing them somehow

The authors of Tariff chose to do this in a slightly involved way

Standardizing Tariff Scores

There is a distribution of Tariff Scores for each cause in the PHMRC data set, and within each cause, the relationship between Tariff Score and cause is related to where the Tariff Score is in this distribution - a measure of rank in this distribution is comparable across causes

Standardizing Tariff Scores

There is a distribution of Tariff Scores for each cause in the PHMRC data set, and within each cause, the relationship between Tariff Score and cause is related to where the Tariff Score is in this distribution - a measure of rank in this distribution is comparable across causes

Causes occur at different frequencies in the PHMRC dataset, so it is necessary to resample the PHMRC data have sufficient deaths for each cause to see a sufficient number of deaths for each cause

Standardizing Tariff Scores

There is a distribution of Tariff Scores for each cause in the PHMRC data set, and within each cause, the relationship between Tariff Score and cause is related to where the Tariff Score is in this distribution - a measure of rank in this distribution is comparable across causes

Causes occur at different frequencies in the PHMRC dataset, so it is necessary to resample the PHMRC data have sufficient deaths for each cause to see a sufficient number of deaths for each cause

The standardized Tariff Score Q_{ki} for death k and cause i is the quantile value of S_{ki} in the distribution of S_{*i} formed by a set of deaths sampled with replacement from the PHMRC data set so that they have a uniform cause distribution

Standardizing Tariff Scores

There is a distribution of Tariff Scores for each cause in the PHMRC data set, and within each cause, the relationship between Tariff Score and cause is related to where the Tariff Score is in this distribution - a measure of rank in this distribution is comparable across causes

Causes occur at different frequencies in the PHMRC dataset, so it is necessary to resample the PHMRC data have sufficient deaths for each cause to see a sufficient number of deaths for each cause

The standardized Tariff Score Q_{ki} for death k and cause i is the quantile value of S_{ki} in the distribution of S_{*i} formed by a set of deaths sampled with replacement from the PHMRC data set so that they have a uniform cause distribution

Values for Q_{ki} are bounded between 0 and 1 so they have the same scale and can be compared

The cause chosen for each death k is the one with the largest Q_{ki}

1 Background

2 Derive Tariff

3 Similarities between Tariff and InterVA

Relationship between InterVA and Tariff

The central relationship in InterVA is

$$\Pr(c|\mathbf{s}) = \frac{\Pr(c) \prod_s \Pr(s|c)^s}{\sum_c \Pr(c) \prod_s \Pr(s|c)^s} \propto \prod_s \Pr(c) \Pr(s|c)^s \quad (4)$$

The central relationship in Tariff is

$$S_{ki} = \sum_{j=1}^J T_{ij} \times r_{kj} \quad (5)$$

Relationship between InterVA and Tariff

The logic of InterVA and Tariff are effectively identical net of a log transform

- There is a conceptual equivalence between $\Pr(c|\mathbf{s})$ and S_{ki} – call them simply ‘the score’.

Relationship between InterVA and Tariff

The logic of InterVA and Tariff are effectively identical net of a log transform

- There is a conceptual equivalence between $\Pr(c|\mathbf{s})$ and S_{ki} – call them simply ‘the score’.
- Likewise for $\Pr(s|c)$ and T_{ij} – call them symptom-cause-information or SCI

Relationship between InterVA and Tariff

The logic of InterVA and Tariff are effectively identical net of a log transform

- There is a conceptual equivalence between $\Pr(c|\mathbf{s})$ and S_{ki} – call them simply ‘the score’.
- Likewise for $\Pr(s|c)$ and T_{ij} – call them symptom-cause-information or SCI
- InterVA calculates the score by multiplying SCI for existing symptoms

Relationship between InterVA and Tariff

The logic of InterVA and Tariff are effectively identical net of a log transform

- There is a conceptual equivalence between $\Pr(c|\mathbf{s})$ and S_{ki} – call them simply ‘the score’.
- Likewise for $\Pr(s|c)$ and T_{ij} – call them symptom-cause-information or SCI
- InterVA calculates the score by multiplying SCI for existing symptoms
- Tariff calculates the score by summing SCI for existing symptoms

Difference between InterVA and Tariff

The most important difference between the two is the SCI they use

- InterVA uses physician-derived $\Pr(s|c)$ obtained through an 'oracle' approach – doesn't require 'gold-standard' deaths, but relies on opinions; subjective and not representative

Difference between InterVA and Tariff

The most important difference between the two is the SCI they use

- InterVA uses physician-derived $\Pr(s|c)$ obtained through an 'oracle' approach – doesn't require 'gold-standard' deaths, but relies on opinions; subjective and not representative
- Tariff uses x_{ij} from 'gold-standard' deaths from PHMRC obtained through an empirical approach – doesn't rely on opinion but does require 'gold-standard' deaths; hard to come by and not representative

SmartVA-Analyze

SmartVA-Analyze implements Tariff 2.0 algorithm (fewer inputs, PHMRC 'shortform' instrument)

SmartVA-Analyze

SmartVA-Analyze implements Tariff 2.0 algorithm (fewer inputs, PHMRC 'shortform' instrument)

An R package for Smart-VA Analyze that will work with openVA is in development

SmartVA-Analyze

SmartVA-Analyze implements Tariff 2.0 algorithm (fewer inputs, PHMRC 'shortform' instrument)

An R package for Smart-VA Analyze that will work with openVA is in development

Should be ready during 2018 at some point

References I

- [1] Spencer L. James, Abraham D. Flaxman, Christopher J. L. Murray, and PHMRC. Performance of the Tariff Method: validation of a simple additive algorithm for analysis of verbal autopsies. *Population Health Metrics*, 9, AUG 4 2011.
- [2] Peter Serina, Ian Riley, Andrea Stewart, Spencer L James, Abraham D Flaxman, Rafael Lozano, Bernardo Hernandez, Meghan D Mooney, Richard Luning, Robert Black, et al. Improving performance of the tariff method for assigning causes of death to verbal autopsies. *BMC medicine*, 13(1):291, 2015.