Estimating adult mortality in countries with defective data: some alternative approaches

Bernardo Lanza Queiroz

Universidade Federal de Minas Gerais – Dept. of Demography and Cedeplar

lanza@cedeplar.ufmg.br

Based on research projects developed since 2011 with funding from CNPq and collaboration with Marcos Gonzaga (UFRN), Flavio Freire (UFRN), Everton Lima (Unicamp), Emerson Baptista (ColMex), Tim Riffe (University of the Basque Country) and several graduate and under-graduate students from UFMG, Unicamp and UFRN.

Formal Demography Working Group, November 29th, 2022

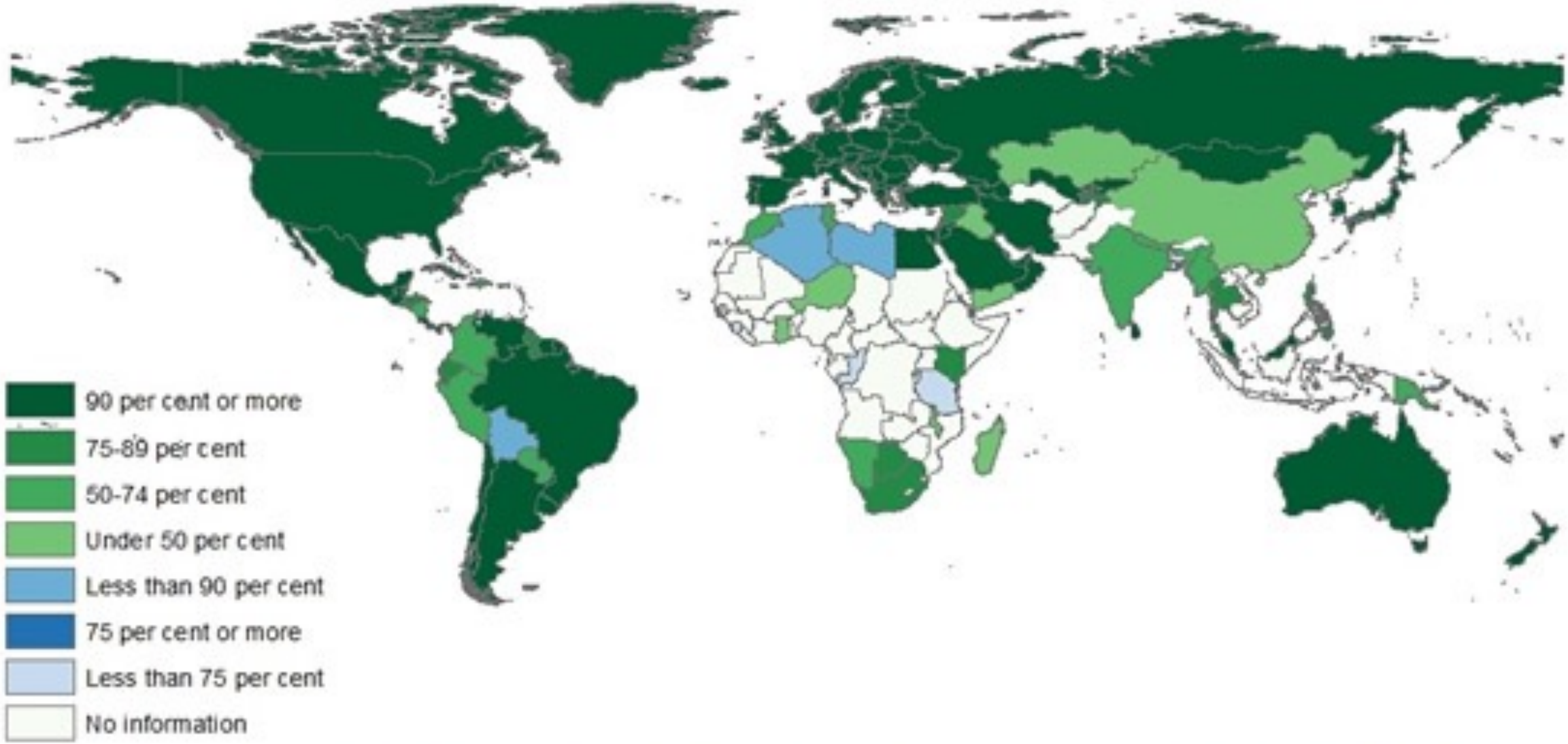
Why is there a problem?

- Incomplete vital registration
- Inaccurate censuses
- Age misreporting

Why does it matter?

- Produce proper mortality estimates
- Mortality forecasting
- Analysis of burden of diseases
- **What is the death toll of the Covid-19 pandemic??**

Coverage of Death Registration (last update: December 2017, United Nations Statistics Division)



The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations.

Important to have in mind

- **Coverage** is a measure of the population that the registry system serves.
  - This is primarily, although not exclusively, an issue of access to the reporting system

- **Completeness is a measure of how well we capture all the events in that population.**
  - **Completeness is also frequently used in relation to the individual data fields collected for each record**

Motivation and relevance

Ken Hill: Demographic Detective Tracking Population Health Mysterie

The art of demography "is a judgment—a feel for the culture, a feel for the data. One tries to avoid being too naïve, accepting everything at face value."

Population Studies, **54** (2000), 129–131
Printed in Great Britain

William Brass 1921–1999

William Brass, who died on 11 November 1999, is admired throughout the world of population science as one of its most creative and influential exponents. To mark his death, we publish an obituary written by John Cleland for a British newspaper, *The Independent*[1], and an appreciation by Basia Zaba. As a tribute to him, a number of William Brass's former students and colleagues have contributed to a volume of essays and reports of studies in the field of medical demography. Entitled *Brass Tacks*, the book has been edited by Basia Zaba and John Blacker and has a foreword by Griffith Feeney. It will be published later this year in London by The Athlone Press.

AN OBITUARY BY JOHN CLELAND

WILLIAM BRASS 1921–1999

WILLIAM BRASS was one of Britain's most distinguished population scientists since the Reverend Thomas Malthus. Over a long career he almost single-handedly created a new and important strand of demography. The central problem that he addressed was this. In developing countries where most births and deaths are not recorded, how can reliable estimates be made of death rates, birth rates and trends in the size and structure of a population?

Over the last 40 years, he devised a series of ingenious answers to this question. For instance he showed how child mortality levels and totals could be derived from simple questions in censuses or surveys to mothers on numbers of children born and still surviving; and how adult mortality could be estimated by asking people whether their father or mother had died.

He was a genius at transforming sow's ears into silk purses. The phrase 'Brass estimates' has become part of the everyday vocabulary of demographers. Much of what we know about the populations of the developing world stems from the battery of methods that he devised.

Brass's work always had a strong methodological focus but the applications ranged widely. He made major contributions to the study of the bio-social determinants of fertility and mortality; population forecasting; and evaluation of family planning programmes. One of his most important studies, from a policy stance, was the demonstration that the high birth rates in the 1970s of some immigrant groups in this country would not be sustained.

Bill Brass studied at the University of Edinburgh from 1940 to 1943 and from 1946 to 1947. From 1948 to 1955 he was Statistician, and later Deputy Director, in the East African Statistical Department. There he worked on the East African Medical Survey and some of the early colonial censuses and it was during this time that he developed many of his ideas on the collection of demographic data and techniques for their analysis.

In 1955 he returned to academic life in the Department of Statistics in the University of Aberdeen. He remained there for nine years, and was appointed Senior Lecturer in 1963. During this period, a year's leave of absence was taken at the Office of Population Research at Princeton University. From this collaboration with Ansley

[1] Reprinted by permission from *The Independent*, Obituaries, 19 November 1999.

Short personal note:
First contact I had with the methods I am talking today was in the late 1990s with Profa. Diana Sawyer and José Alberto Carvalho at Cedeplar/UFMG. In the early 2000´s, I had the chance to be a RA for Ken Hill – we were not in the same university – in a project to evaluate data quality and estimate adult mortality in Latin America
They are such great teachers and mentors.

Overview of Methodology – Death Distribution Methods

Methods that compare the distribution of deaths by age with the age distribution of the living are attractive:

- Provide age pattern of mortality
- Defined reference period
- Often can use existing data

But,

- Strong assumptions required
- No clear knowledge of sensitivity to the assumptions

What type of data do we need?

- 1) Two population age structures
  - Normally from censuses

- 2) Death counts by age
  - CRVS systems
  - Household deaths in the previous 12 months
  - Other household surveys

- This is na advantage compared to Other methods, such as sibiling survival, because we use available data and have a specific period of reference.

Growth Balance and General Growth Balance

- The Growth Balance Identity simply states that the (*rate of)* change in population between two time points is equal to the difference between the *(rates of)* entries and the *(rates of)* exits during the interval (Brass 1975; Hill, 1987).

- In a population with no migration, entries *e* are births and exits *d* are deaths. For an open-ended age group *a*+, entries are a^{th} birthdays and exits are deaths *a* and over. Thus in terms of *rates*:

$$e(a+) - r(a+) = d(a+)$$

where

$e(a+) = N(a)/N(a+)$,

$r(a+)$ is the growth rate of the population *a*+, and

$d(a+) = D(a+)/N(a+)$

If we can estimate $e(a+)$ and $r(a+)$ from age distributions, their difference provides a consistency check for $d(a+)$.

Typical data errors include incomplete death recording and changes in coverage from one census to another. If deaths are registered with completeness c (at all ages),

$$D(a+) = \{1/c\}D^{o}(a+)$$

If we assume constant exponential growth between two censuses,

$$r(a+) = \{1/t\}\ln\{N2(a+)/N1(a+)\}$$

If coverage of the second census differs from that of the first by a factor k (also assumed constant at all ages),

$$r(a+) = \{1/t\}\ln\{[N2_{o}(a+)/N1_{o}(a+)]*[1/k]\} = r_{o}(a+) - \{1/t\}\ln(k)$$

Replacing true values with observed values,

$$\frac{N(a)}{N(a+)} - r^{o}(a+) = \frac{1}{t}\ln(k) + \frac{1}{c} * \frac{D^{o}(a+)}{N(a+)} \quad (1)$$

- the residual estimate (LHS) of the death rate in each open-ended age group should be linearly related to the observed death rate (RHS) in the age group, **with slope equal to the reciprocal of completeness of death recording *c* and intercept a function of any change in completeness of population recording and the length of the interval.**

Left-hand terms are derived from census age distributions

Right-hand term comes from deaths by age and census age distributions.

Synthetic Extinct Generations Method

Vincent (1951) proposed that (with perfect recording of deaths) the population age *a* at time *t* could be estimated *ex post facto* by cumulating the deaths to that cohort after time *t* until the cohort was *extinct*.

On period basis, in a stationary (life table) population, the survivors to age *a* are equal to the deaths above age *a*:

$$\ell(a) = \sum_{x=a}^{\omega} {}_1d_x$$

In a stable population, the deaths above age *a* need to be expanded by the stable growth rate to reflect change in birth cohort size (Preston et al. 1980):

$$N(a) = \int_a^{\varpi} D(x)e^{r(x-a)}dx$$

If death registration is c complete, constant by age, the ratio of the estimated population age *a* based on the expanded deaths *a*+ to the observed population age *a* estimates the completeness of death recording relative to population recording:

$$\frac{\hat{N}(a)}{N(a)} = \frac{\int_a^{\varpi} D^{o}(x)e^{r(x-a)}dx}{\int_a^{\varpi} D(x)e^{r(x-a)}dx} = \frac{c * \int_a^{\varpi} D(x)e^{r(x-a)}dx}{\int_a^{\varpi} D(x)e^{r(x-a)}dx} = c$$

Where *c* is the completeness of death recording (relative to population)

Bennett and Horiuchi (1981;1984) generalized the method to any closed population by using the **variable growth rates above age *a*:**

$$N(a) = \int_a^{\varpi} D(x) e^{\int_a^x r(y)dy} dx$$

That is, the population age *a* can be estimated from the deaths above that age by applying summed age-specific growth rates to allow for the demographic history of the population. **The ratio of the population age *a* estimated in this way from the deaths to the observed population age *a* estimates the completeness of death recording relative to population recording.**

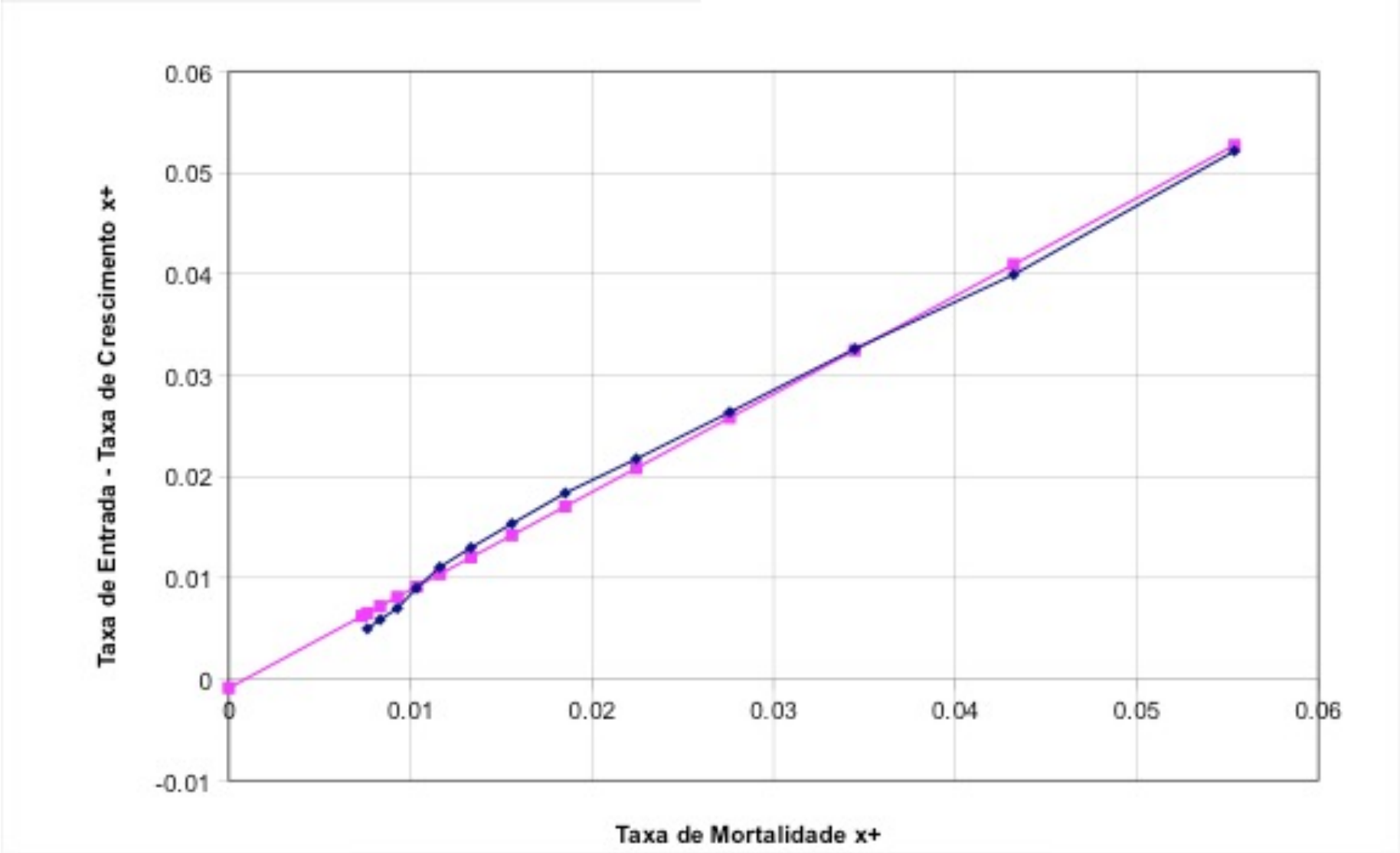
GGB-SEG (or hybrid)

- Hill, You and Choi (2009) proposed a combination of both methods based on a series of simulation
  - 1) use General Growth Balance to adjust the level of coverage of both censuses
    - It does not mean that censos data are good or bad, we are justing adjusting so they have the same level of coverage
  - 2) Use Bennett-Horiuchi to estimate completeness of death counts
  - Rob Dorrington and Tom Moultrie have an alternative approach (SEG-delta) with similar results (and it is very elegant).

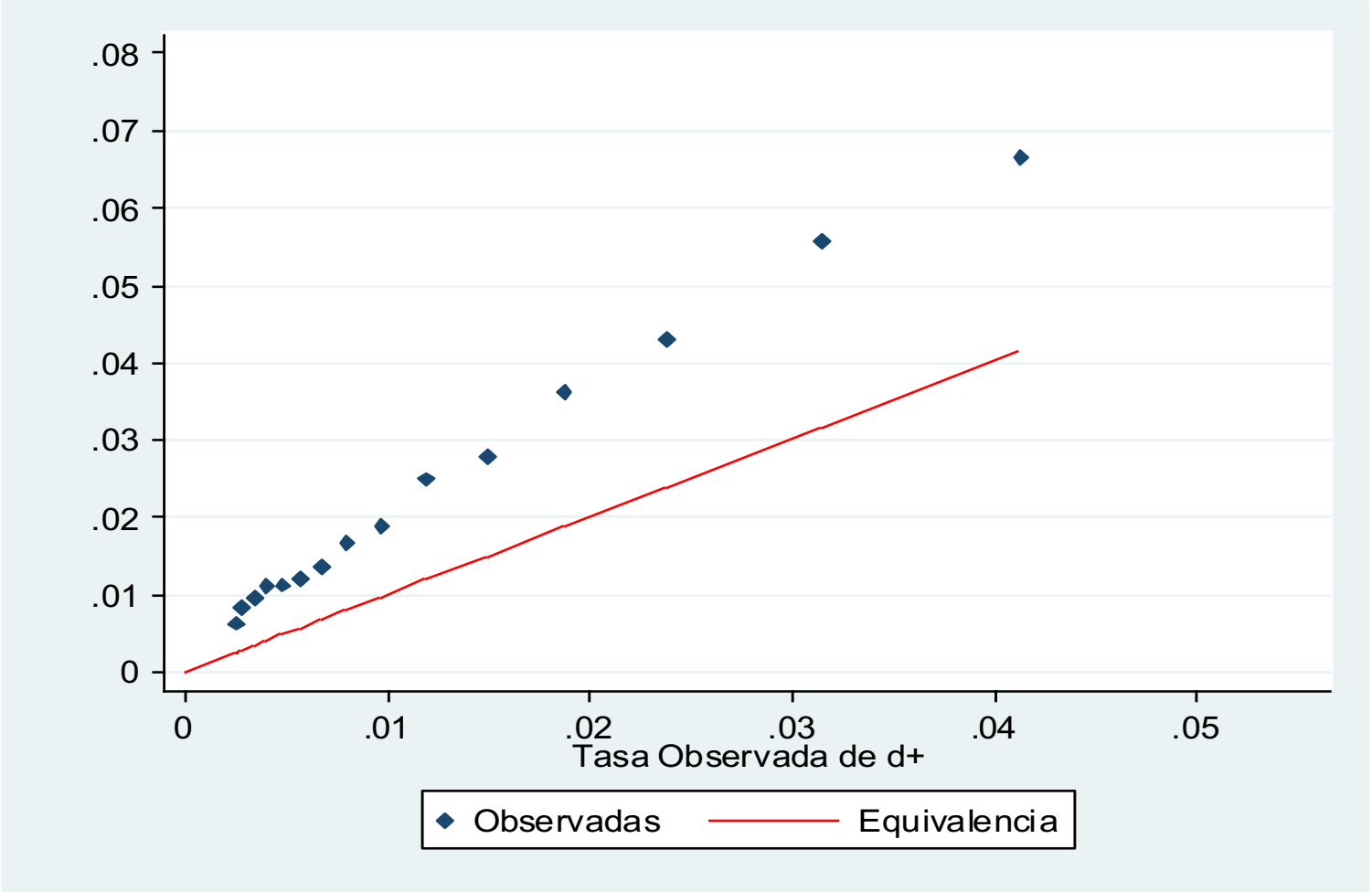
But, how we define the age range to estimate completeness?

- It is very important to take a close look at the diagnostic plots (more soon) and evaluate possible problems related to the assumptions;
- Hill, You and Choi (2009) perfomed a simulation analysis to investigate how each method is affected by the assumptions and data limitation and suggest the use of ages 15+ to 55+ or 15+ to 65+, showing that the results are quite similar;
- Murray et.al (2010) performed a diferente simulation analysis and suggested similar age ranges
- Hill (2017) suggested that avoiding ages with more migration might be a better alternative.

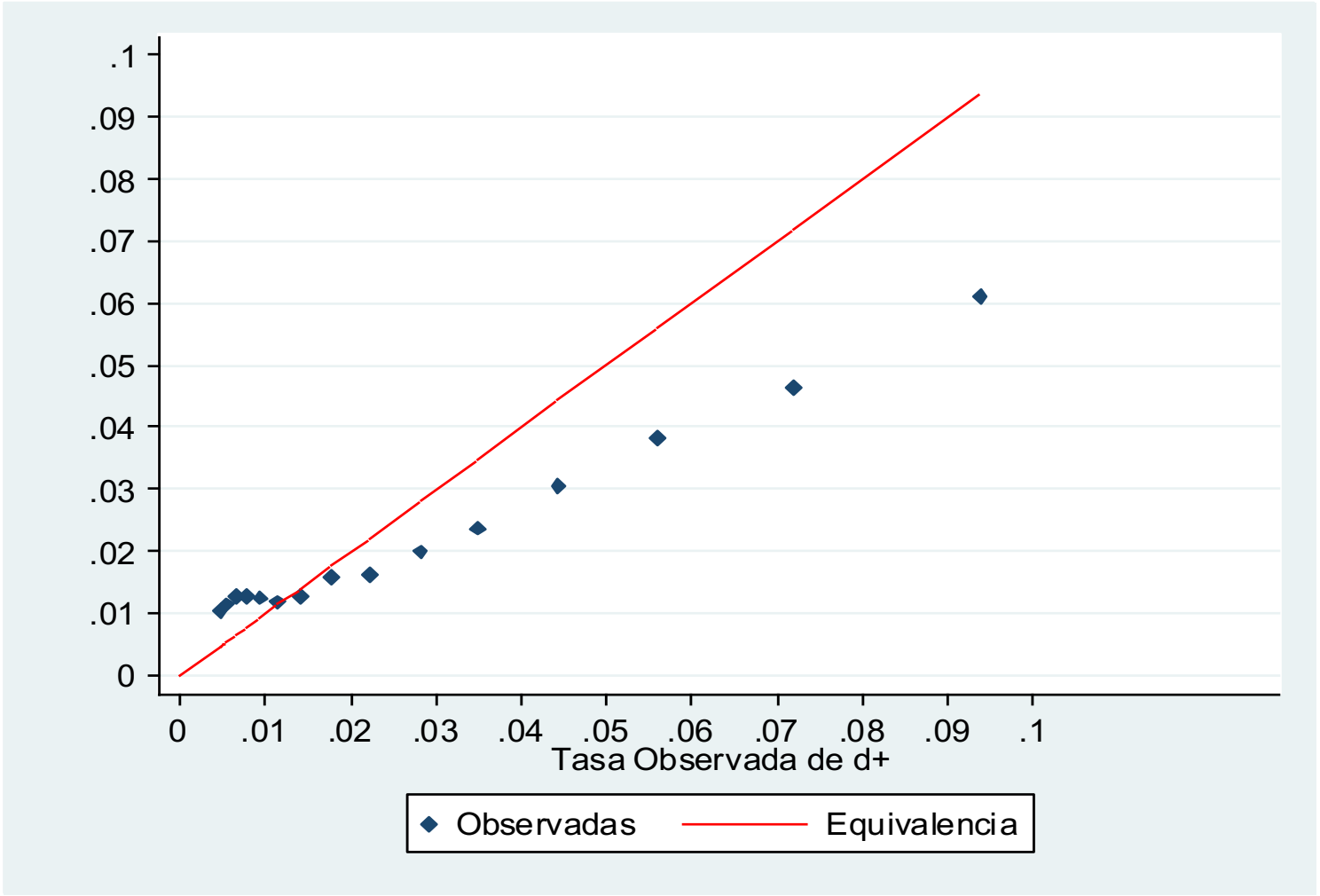
Diagnostic Plots, GGB, Sao Paulo, Males, 2010



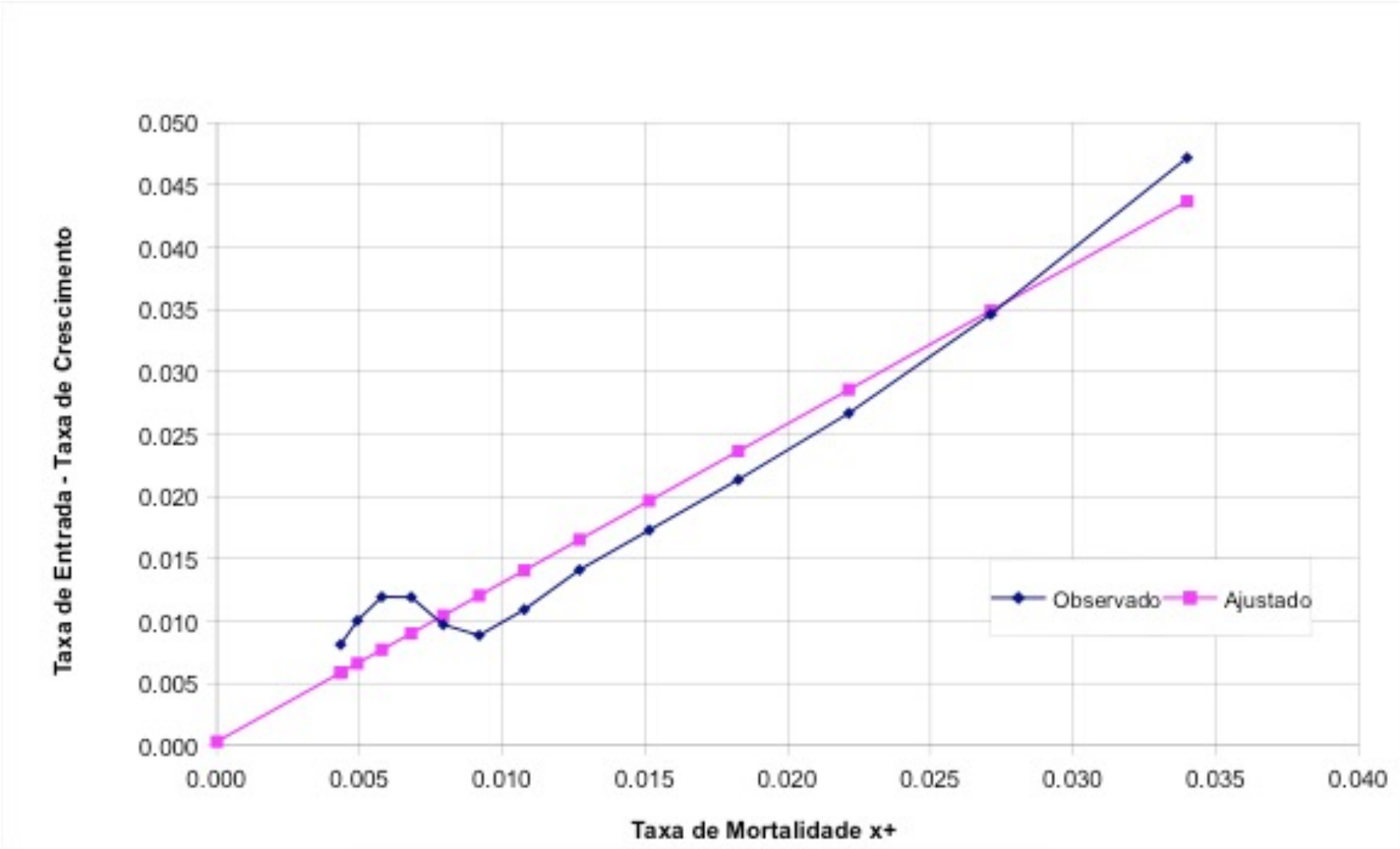
Honduras Females 1988-2001



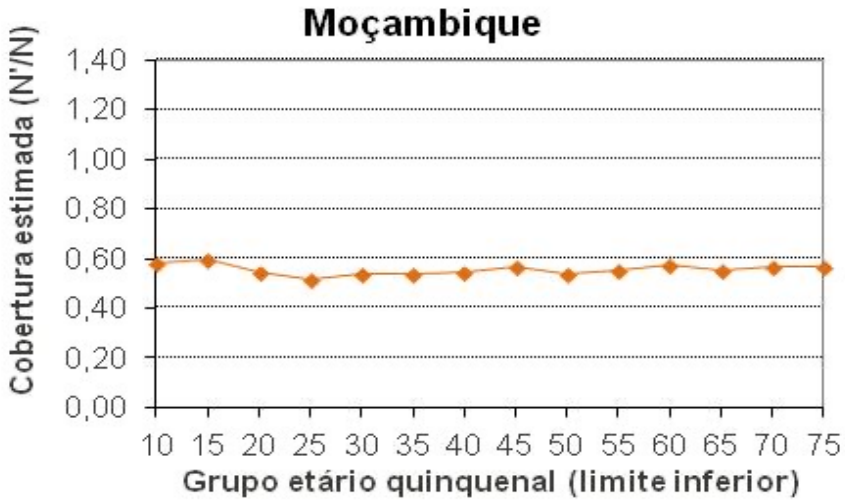
Nicaragua Females 1995-2005

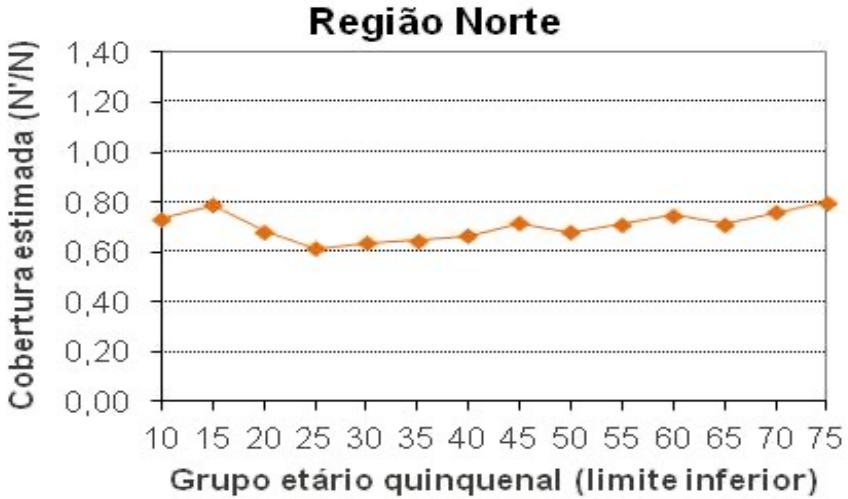


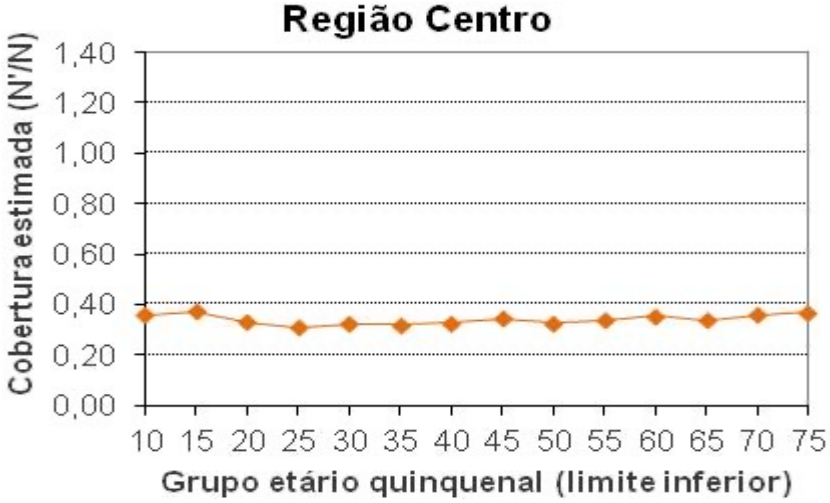
Diagnostic Plots, GGB, Maranhão, Males, 2010

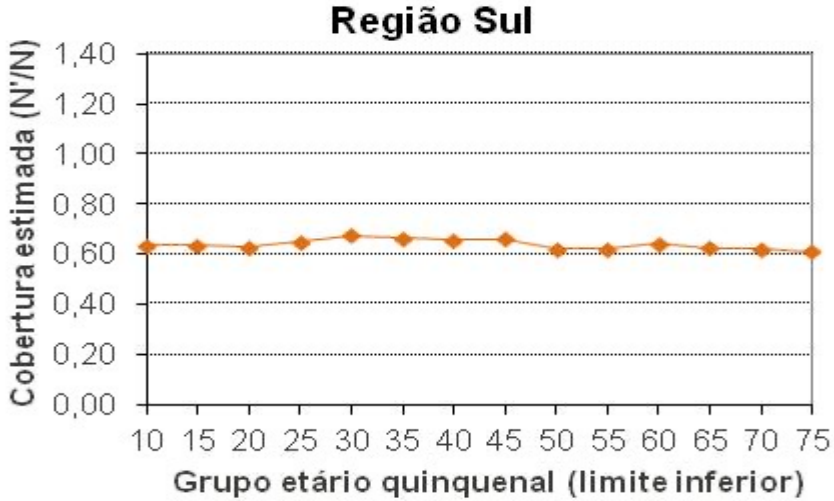


Mozambique and regions – BH-adjusted, Females, 1997/2007



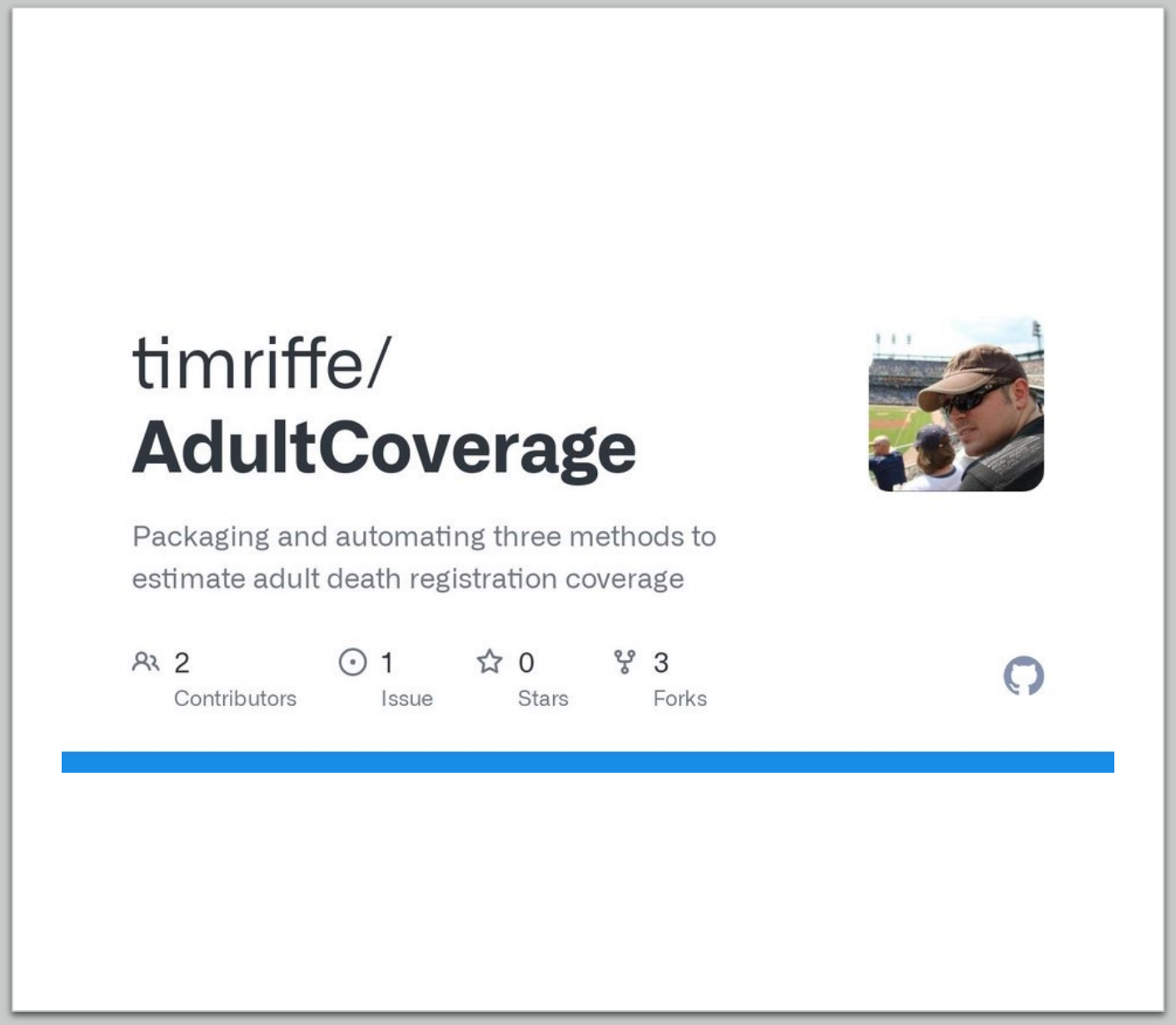






DDM R package

- This project has produced a small R package that implements three methods for estimation of death registration coverage (Generalized Growth Balance, Synthetic Extinct Generations, and a hybrid of the two).
- https://cran.r-project.org/package=DDM
- Collaborative work with Tim Riffe, Everton Lima and Peter Johnson



Additional approaches – I am not talking about them today

1) Schmertmann, C. P., & Gonzaga, M. R. (2018). Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. *Demography*, *55*(4), 1363-1388.
    1) But they also use Topals model (that I will talk about soon)

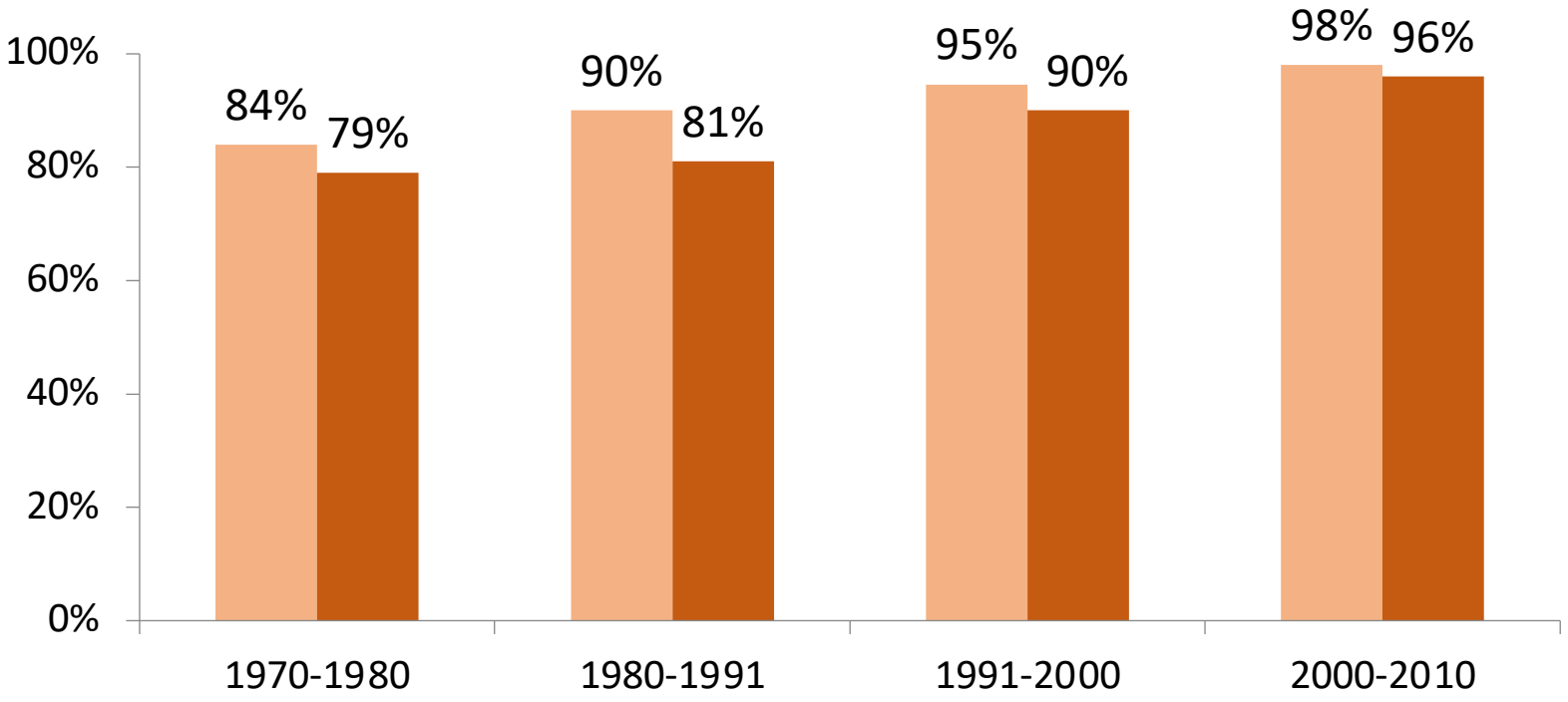
2) Adair, T., & Lopez, A. D. (2018). Estimating the completeness of death registration: an empirical method. *PloS one*, *13*(5), e0197047.

A few examples

Brazil – country level

Estimates of completeness of adult mortality (45q15)

Evaluation of the Completeness of Death Counts Coverage, Brazil, 1980 - 2010

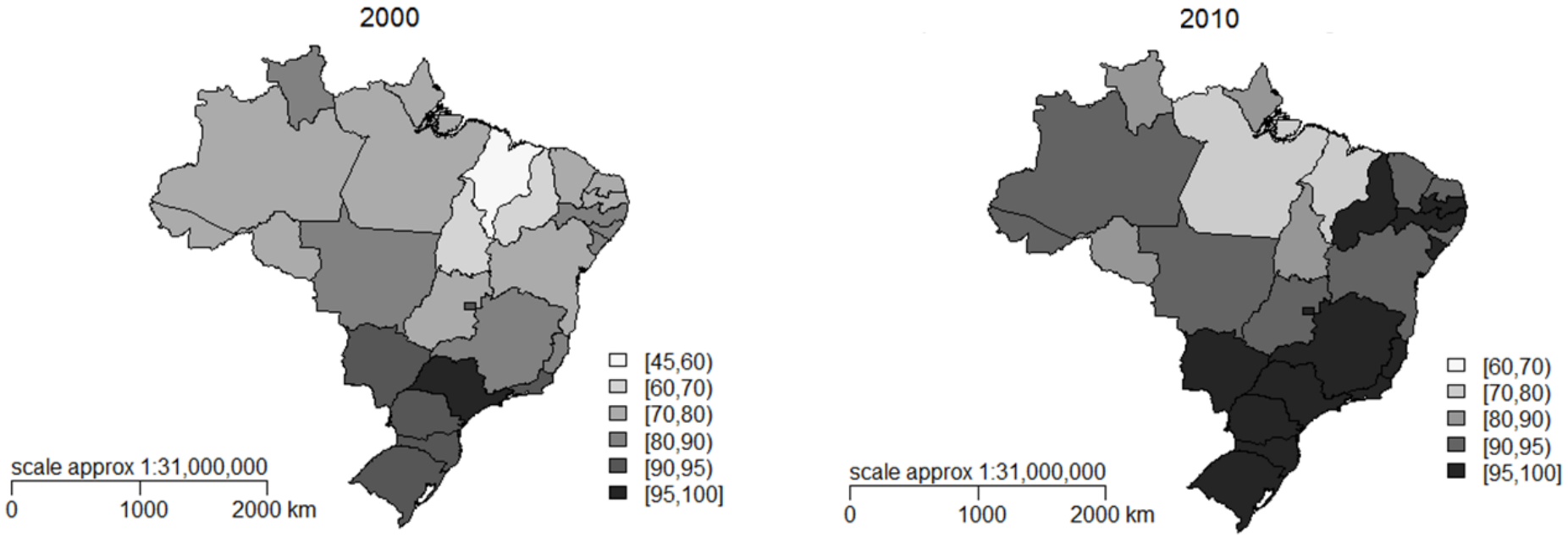


Queiroz, BL; Freire, FH; Lima, EC; Gonzaga, M. Completeness of death-count coverage and adult mortality (45q15) for Brazilian states from 1980 to 2010. **Rev Bras Epidemiol**; v. 20 SUPPL 1: 21-33, 2017.

State level – Brazil

Estimates of completeness of adult mortality (45q15)

Evoluation of the Completeness of Death Counts Coverage, Brazilian States, 1980 - 2010



Queiroz, BL; Freire, FH; Lima, EC; Gonzaga, M. Completeness of death-count coverage and adult mortality (45q15) for Brazilian states from 1980 to 2010. **Rev Bras Epidemiol**; v. 20 SUPPL 1: 21-33, 2017.

But, there is an increase interest in small-areas

1) Two main Problems

    1) **Small number of events**
        1) First alternative: use indirect standardization – full of assumptions

        2) Final approach: Topals method (Gonzaga and Schmertmann mofication)

Gonzaga, M. R., & Schmertmann, C. P. (2016). Estimating age-and sex-specific mortality rates for small areas with TOPALS regression: an application to Brazil in 2010. *Revista Brasileira de Estudos de População*, *33*, 629-652.

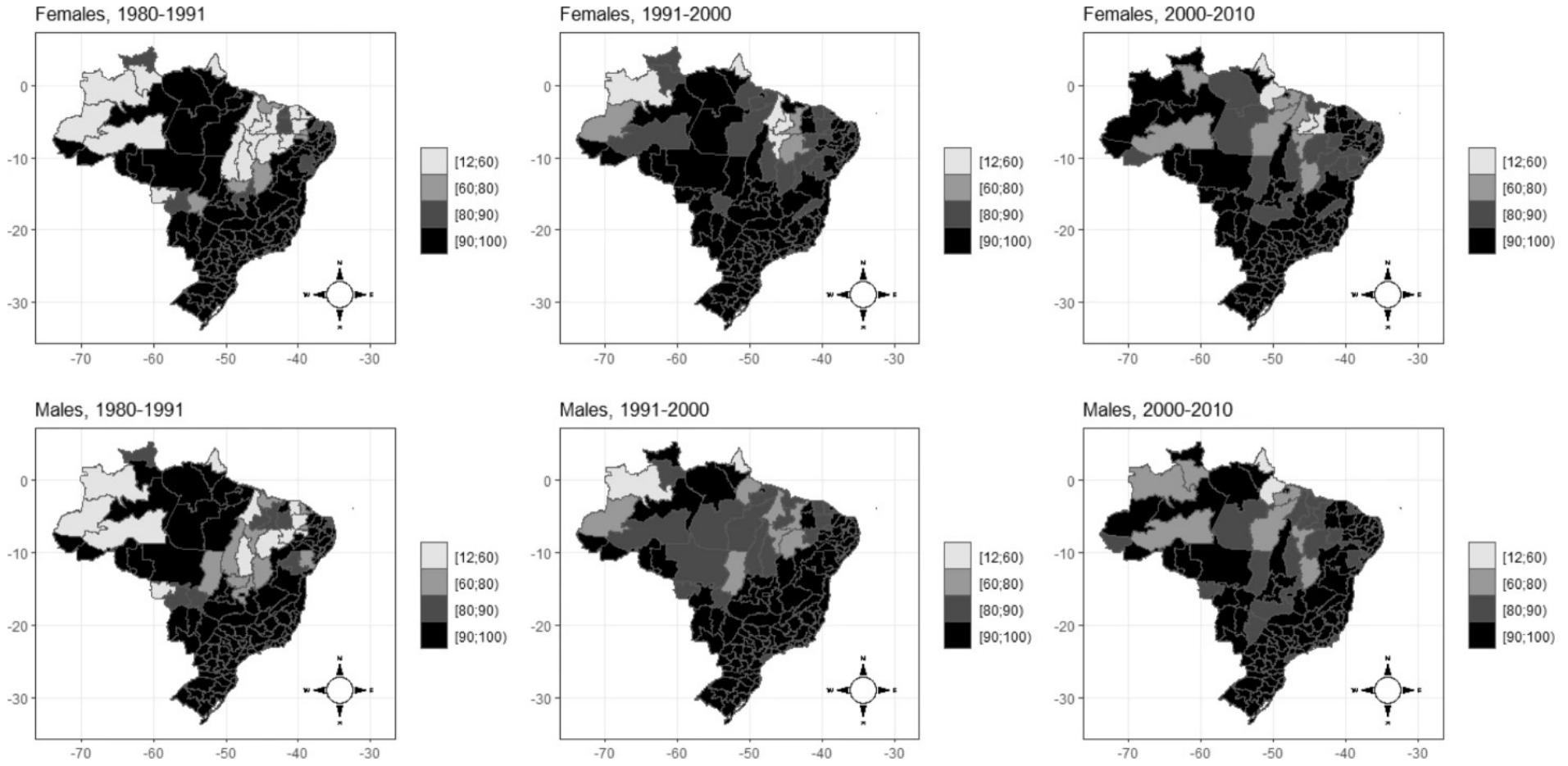
    1) **Closed population or very small migration flows**
        1) Could use Hill and Queiroz (2010)
            1) Based on Rogers and Castro standard
        2) Opted to follow Hill (2017) suggesting and use a smaller age range to produce estimates of completeness

What did we do, in the case of Brazil?

- We produce estimates of data quality (completeness of death counts) and adult mortality for 137 regions in Brazil from 1980 to 2010;
- Relevant to understand the dynamics of health and mortality across time and space in the country
- Procedure:
  - Apply Topals regression model to obtain complete mortality schedules by age and sex
  - Apply DDM to evaluate the completeness of death counts in each region
- Produce estimates of adult mortality (45q15) by sex from 1980 to 2010
    - Results are used in ongoing projects
      - Analysis of mortality by causes of deaths in Brazil
      - Mortality forecasts for mesoregions using a simplified version of the Lee-Carter method
- **All codes and data are available at:**
  - **https://demografiaufrn.net/laboratorios/lepp/paper_genus/**

Smaller areas – 137 mesoregions

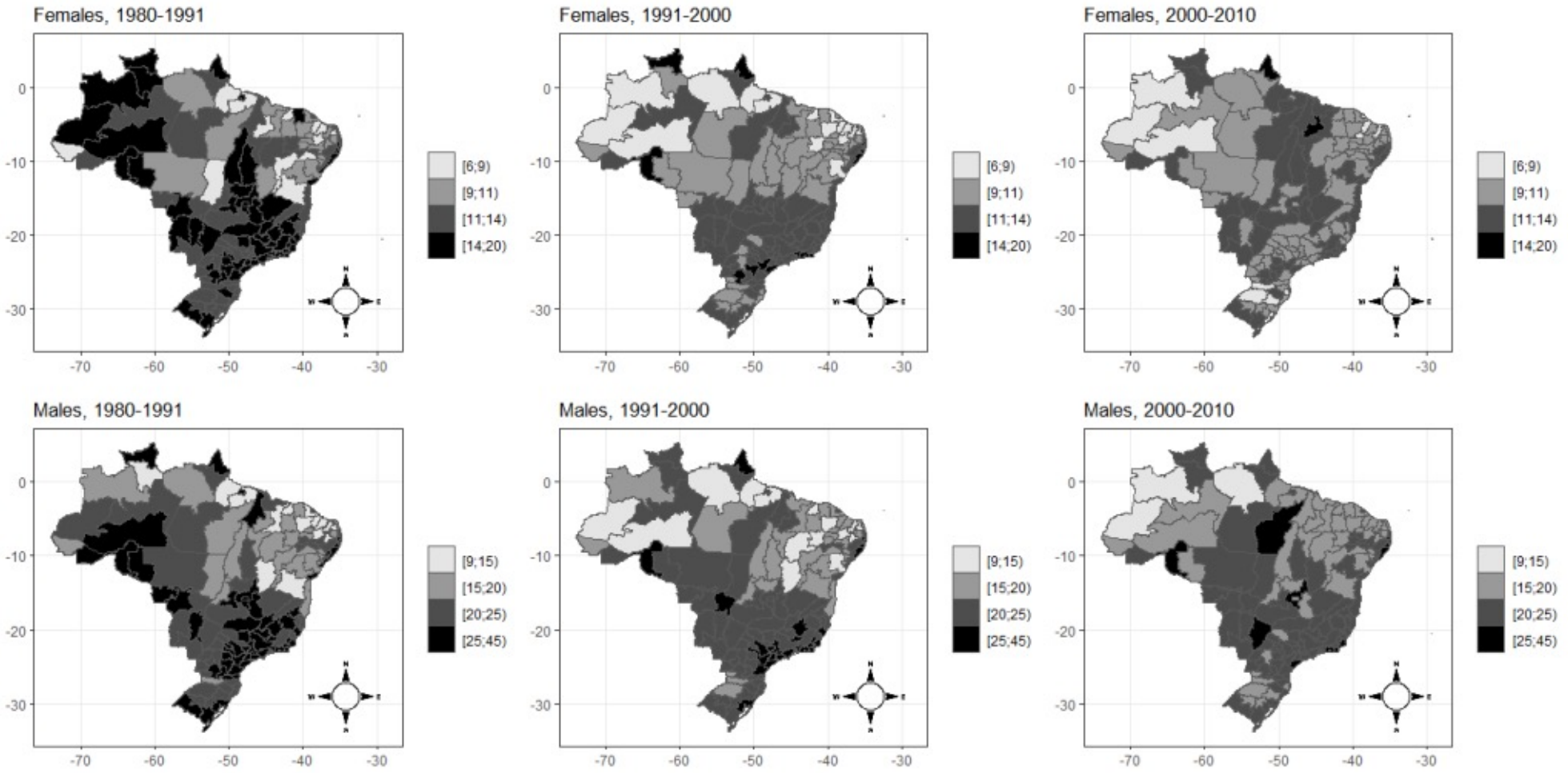
Adult mortality data quality (1980 – 2010)



Queiroz, B. L., Lima, E. E., Freire, F. H., & Gonzaga, M. R. (2020). Temporal and spatial trends of adult mortality in small areas of Brazil, 1980–2010. *Genus*, *76*(1), 1-22.

Smaller areas – 137 mesoregions

Adult mortality data quality (1980 – 2010)

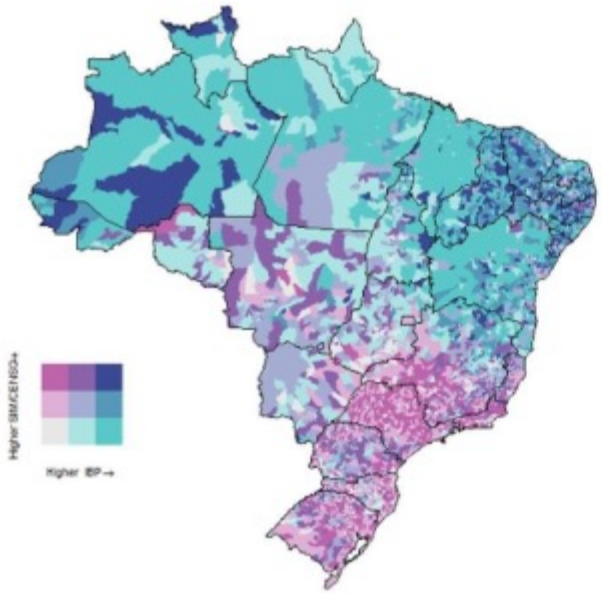


Queiroz, B. L., Lima, E. E., Freire, F. H., & Gonzaga, M. R. (2020). Temporal and spatial trends of adult mortality in small areas of Brazil, 1980–2010. *Genus*, *76*(1), 1-22.

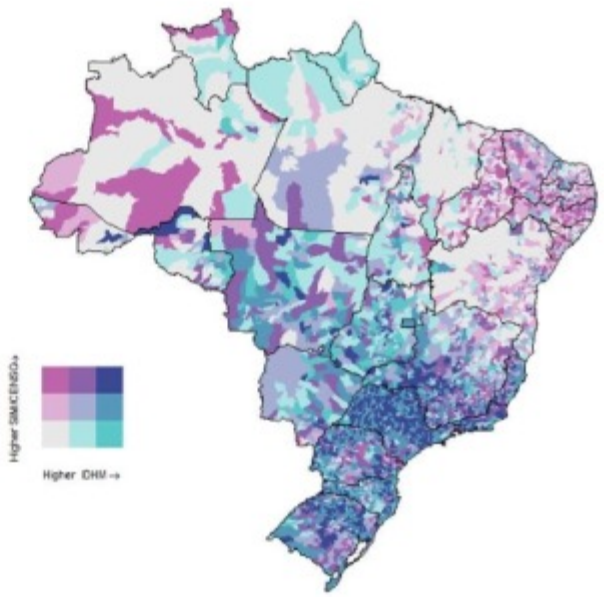
Brazil at the local level, 2010

Figura 2 – Indicadores socioeconômicos (IBP e IDHM) X Razão SIM/CENSO, Brasil, 2010.

2a – IBP x Razão SIM/CENSO



2b – IDHM x Razão SIM/CENSO



Fonte: Elaboração própria a partir dos dados do Censo Demográfico 2010/IBGE, Registro Civil/IBGE e SIM/DATASUS/MS.

DIFERENCIAIS NA COBERTURA DAS BASES DE DADOS DE ÓBITOS NO BRASIL EM 2010

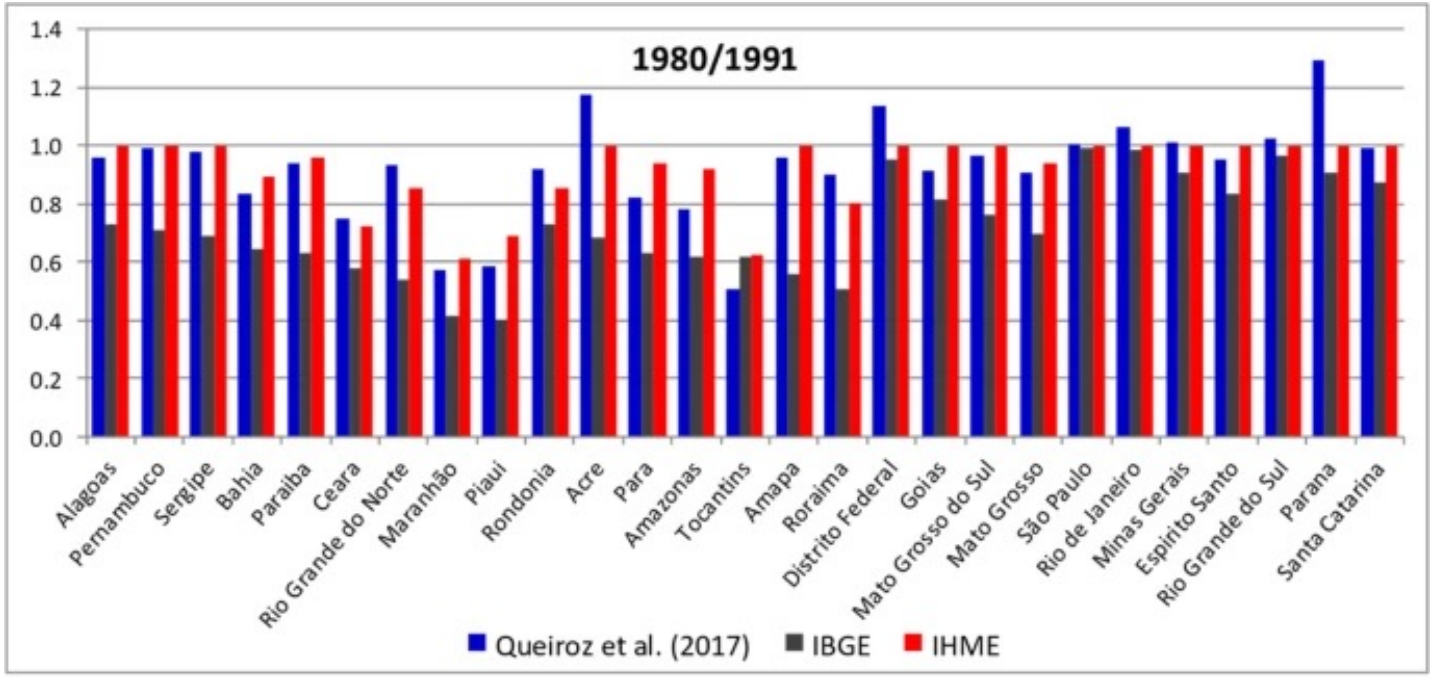
Cobertura dos registros de óbitos no Brasil

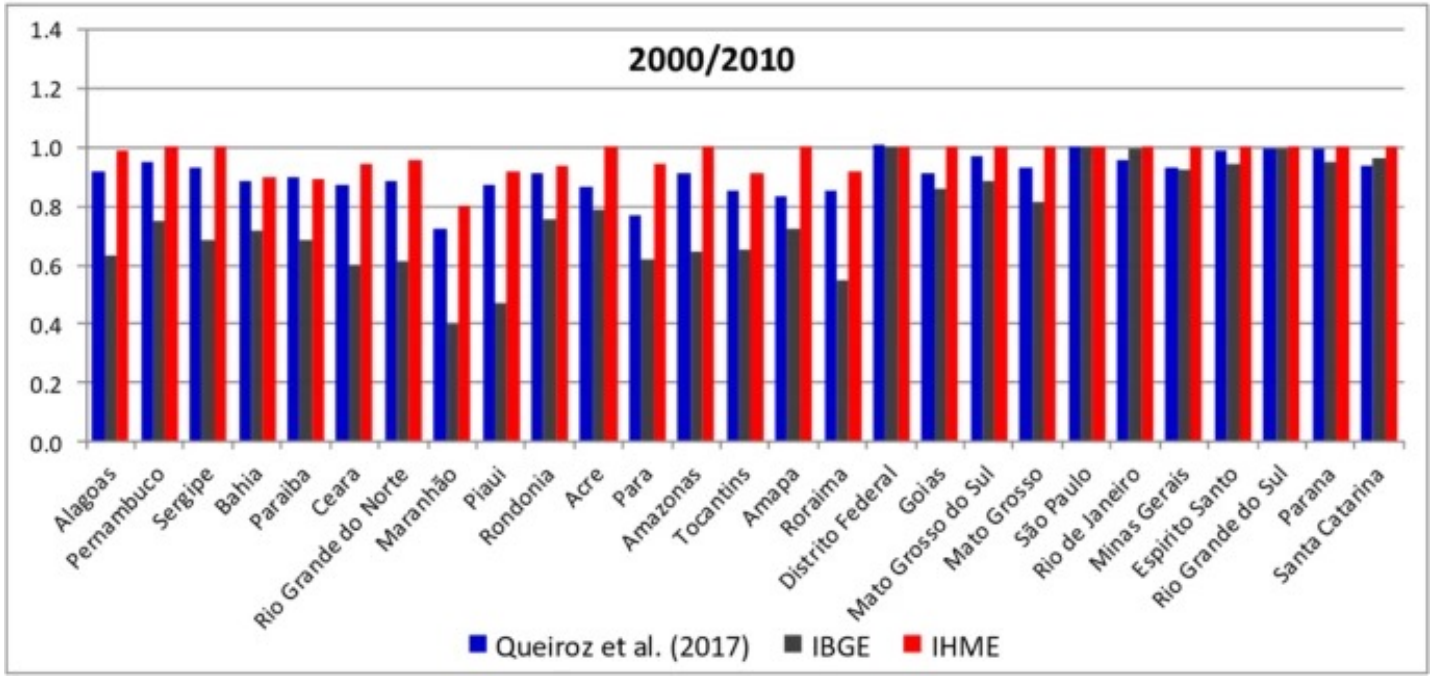
Victor Hugo Dias Diógenes[I,II,III], Elzo Pereira Pinto Júnior[III], Marcos Roberto Gonzaga[I], Maria Yury Travassos Ichihara[III], Bernardo Lanza Queiroz[IV], Everton Emanuel Campos de Lima[V], Lilia Carolina Carneiro da Costa[III,VI] , Aline S. Rocha[III,VI], Andrêa J. F. Ferreira[III], Camila S. S. Teixeira[III], Flávia Jôse O. Alves[III,VI], Leila Rameh[III], Renzo Flores-Ortiz[III], Maurício L. Barreto[III].

How does it compare to other studies and sources?

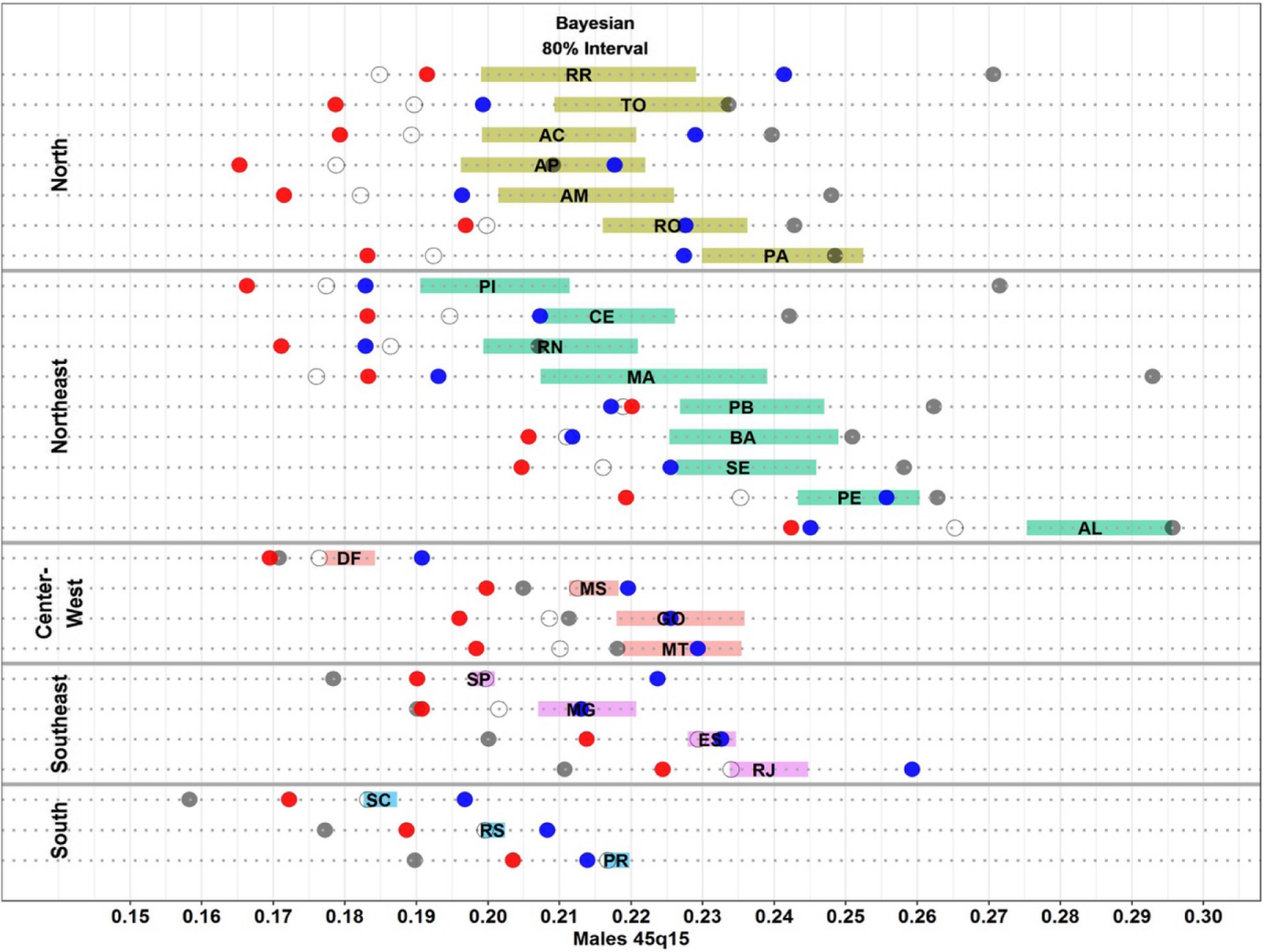
- Main interest is to compare to
  - Brazilian Statistics Office (IBGE)
  - Institute of Health Metrics Evaluation (IHME)
  - Results were published in 2020
    - Queiroz, B. L., Gonzaga, M. R., Vasconcelos, A., Lopes, B. T., & Abreu, D. M. (2020). Comparative analysis of completeness of death registration, adult mortality and life expectancy at birth in Brazil at the subnational level. *Population health metrics*, *18*(1), 1-15.

Comparing completeness for States from different sources, males





Adult mortality – different sources and methods, 2010



- GBD – IHME
- Queiroz, et. al (2017)
- IBGE
- Bayesian model (80% CI)
- Datasus – no adjustment (empty circle)

Conclusion

- important the continuous development of the registration system for deaths and births (CRVS);
- importance of investing in the quality of the data declaration information, individual characteristics,
- develop or use other types of administrative records and combine them with matching methods;
- especially important to better understand mortality at more advanced ages
- In the absence of a vital registry or a perfect mortality information system, the use of the demographic census could be an interesting complementary alternative, but it should not be considered a final solution.
- Statistical modelling – similar to GBD – should not be seen as a substitute to good quality CRVS.

Thank you!

email: lanza@cedeplar.ufmg.br

https://docentes.face.ufmg.br/lanza/

There is no method which functions at all times and in all situations. All are based on certain suppositions about how errors are made; it is very difficult, if not impossible, to anticipate patterns of error. While there is only a single truth, there is an infinite number of ways to make errors. (Bill Brass)

Some results using census data from Mozambique

Questions in each country

- Mozambique (2007)
  - In this household, has anybody passed away in the past 12 months? (from August 1, 2006 to July 31, 2007)
  - Age and sex of the deceased
  - For Women aged 12-50 only
    - Did death occur during pregnancy?
    - Did death occur during delivery?
    - Did death occur within 2 months after pregnancy or delivery?
  - In **2017** it was also included in the census
    - Asked whether the death was registered in the CRVS system
    - the date of death (day, month and year).

Results – Mozambique - Completeness

Tabela 1

Estimativas de cobertura de enumeração de óbitos, segundo métodos, por intervalos etários. Moçambique, 1997/2007.

Métodos	Intervalos etários selecionados							
	15-59	25-59	25-69	30-64	30-74	Mínimo	Máximo	Média
Masculino								
EGB	0,665	0,636	0,621	0,745	0,666	0,621	0,745	0,667
GE	0,592	0,603	0,610	0,615	0,619	0,592	0,619	0,608
GEA	0,682	0,646	0,628	0,761	0,681	0,628	0,761	0,680
Feminino								
EGB	0,782	0,702	0,727	0,728	0,763	0,702	0,782	0,740
GE	0,908	0,900	0,887	0,878	0,857	0,857	0,908	0,884
GEA	0,746	0,658	0,670	0,682	0,714	0,658	0,745	0,694

EGB: método da equação de balanceamento; GE: método das gerações extintas; GEA: método das gerações extintas ajustadas.
Fonte: dados básicos: censos demográficos de 1980, 1997 e 2007 (Departamento Nacional de Estatística, 1983; Instituto Nacional de Estatística, 1999; 2010).

Adult Mortality, 45q15, Mozambique

Probabilidades de morte entre as idades 15 e 60 anos (45q15), por sexo. Moçambique, 1997/2007.

Método	Probabilidade de morte (45q15)							
	Região Norte		Região Centro		Região Sul		Moçambique	
	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres
Dados diretos	0,478	0,433	0,467	0,373	0,577	0,386	0,499	0,399
EGB	0,560	0,475	0,663	0,641	0,703	0,556	0,626	0,539
GE	0,495	0,535	0,571	0,623	0,553	0,533	0,536	0,564
GEA	0,558	0,505	0,670	0,658	0,703	0,566	0,631	0,555

EGB: método da equação de balanceamento; GE: método das gerações extintas; GEA: método das gerações extintas ajustadas.

Fonte: dados básicos: censos demográficos de 1997 e 2007 (Departamento Nacional de Estatística, 1983; Instituto Nacional de Estatística, 1999; 2010).