

RESEARCH

Open Access



# Spatio-temporal trends of the age-at-menarche percentiles among Portuguese women since 1920

Vitor Rodrigues<sup>1,2</sup>, Rui Martins<sup>3\*</sup> and Bruno de Sousa<sup>4</sup>

## Abstract

**Background** This work aims to study the spatio-temporal evolution of a woman's age at menarche in the central region of Portugal. One of the concerns of the study is early or late menarches; thus, we consider percentile regression to build the respective curves as opposed to the more traditional mean regression approach.

**Methods** We analysed the data from  $N = 452\,348$  women born in the period 1920–1973 who attended a free breast cancer screening program between 1990 and 2019. Distributional regression models inside the package GAMLSS in R were considered. These methods allowed us not only to model the location (mean) of the specific probability distribution of the age at menarche, but also allowed for the scale (variance) parameter of this distribution to depend on covariates. Additionally, a spatial random-effect was considered in order to capture the correlation at the regional level. The obtained clustered spatial effects were analysed to assess geographical differences among the percentiles of the age at menarche by year of birth.

**Results** A decreasing trend in the age at menarche (about 1.5 years in 5 decades) and regional differences for all the considered percentiles were found. Women living in the north-central areas of the central region of Portugal tend to have menarche at older ages.

**Conclusion** We obtained percentile estimates for the age at menarche by year of birth and region of residence and demonstrated that these two explanatory variables have an impact on the explanation about the decreasing trend in age at a woman's first menstruation.

**Keywords** Menarche, GAMLSS, Distributional regression, Spatial, Temporal, Portugal

## Content

Menarche, the term to denote the onset of menstruation, marks the beginning of a girl's fertility. Later age at menarche (AaM) began to occur in the Modern Era, especially after the Industrial Revolution (1760–1840) possibly due to the deterioration of certain living conditions. Estimates for the second half of the 18th century point to an AaM between 15 and 16 years old [1]. This behaviour was found to undergo a reversal in the 20th century, with most studies showing a trend for earlier AaM in both developed and developing countries [2–8]. This feature has been associated with life-style,

\*Correspondence:

Rui Martins

rmmartins@fc.ul.pt

<sup>1</sup> Faculty of Medicine, University of Coimbra, Rua Larga, 3004-504 Coimbra, Portugal

<sup>2</sup> Liga Portuguesa Contra o Cancro, Núcleo Regional do Centro, Rua Dr. António José de Almeida, 329 - piso 2 - Sala 56, 3000-045 Coimbra, Portugal

<sup>3</sup> Departamento de Estatística e Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa, Portugal; Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL), Lisboa, Portugal

<sup>4</sup> Faculty of Psychology and Education Sciences (FPCE); Center for Research in Neuropsychology and Cognitive and Behavioral Intervention (CINEICC), University of Coimbra, Coimbra, Portugal



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

environment, socio-economic status, access to health care and higher levels of literacy [9]. Additionally, several studies point to an association with the consumption of energy-rich food (i.e. with an increased body mass index) [10, 11]. The work of Cabanes et al. 2009 [12], which studied women born from 1925 to 1962, along with Sohn et al. 2017 [13], studying women born in the period 1941 – 1992, noted some rates of decline in AaM. The European Prospective Investigation into Cancer & Nutrition (EPIC) study found that mean AaM decreased among female participants born from 1912 to 1964 in nine European countries [14]. However, in the past few decades, this trend is disappearing, with some studies showing a cessation of the decline [1]. In France, Lally et al. [15] studied a group of school girls born between 1979 and 1994, and in Netherlands Talma et al. [16] studied girls born before the year 2000. Both studies concluded that there are signs of slowing down or stabilization. Finer and Philbin (2014) [17] report a mean AaM among US girls born in 1993 of 12.3 years of age – similar to those born in 1980.

The Portuguese panorama is identical to those already described above. A work [18] that analyses a Portuguese community of students at the University of Coimbra born in the period 1972–1983 claims that the mean AaM for girls born in 1983 was 12.03 years and that the place of residence during childhood and adolescence had a significant effect on the mean AaM. Girls from rural areas had a later menarche when compared to those who spent their childhood/adolescence in urban areas. Despite being focused on a specific-group, this work offers an advantage in our study, in analysing a time-period immediately following that considered in our study (1920 – 1973), thus becoming an interesting point of comparison.

Age at menarche plays a very important role in the research about breast cancer, hence having important individual and public health implications. Early puberty has been associated with an increased risk of this type of cancer, but also with obesity and diabetes [19]. Burgess et al. [20] suggested a protective effect of later AaM on breast cancer risk.

It is well known that AaM varies between and within populations and is influenced by many factors. It may be delayed under poor socio-economic and health conditions or accelerated by residence in an urban community [21]. The starting point for the present research was the idea to understand the spatio-temporal evolution of menarche in the central region of Portugal. To achieve this goal we analysed the dataset on the Breast Cancer Screening Program provided by the Portuguese Cancer League (Liga Portuguesa Contra o Cancro – LPCC) in the Central region of Portugal [22].

A typical regression analysis, in most contexts, focuses on explaining the expected value dependency of a response variable as a function of a set of explanatory variables. However in some situations, if we want to understand not only the behaviour of the average person, but also the behaviour of those belonging to the extremes of the population, we might consider percentile regression [23, 24] which allows us to go beyond mean regression, enabling building regression curves for the percentiles, instead of the mean of a response variable. Rigby et al. [25] argue that for data with more than 1000 observations, regression models beyond the mean should be the norm, not the exception.

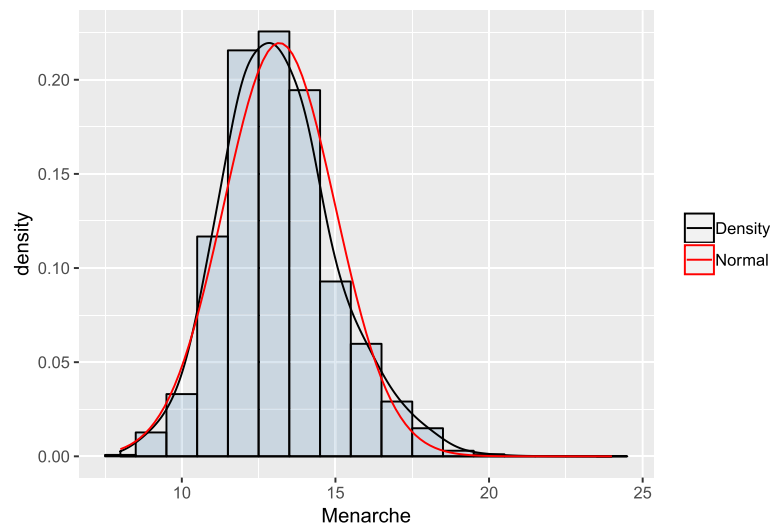
### Data description

The study of the AaM percentiles in the central region of Portugal was carried out considering the full dataset on the Breast Cancer Screening Program provided by the Portuguese Cancer League (LPCC) in the Central region of Portugal in the period 1990 – 2019 which has screened  $N = 452\,348$  women born in the period 1920 – 1973. At the age of 45 (since 2010 the age is 50) all women in each of the 89 municipalities (see Fig. 3) were invited to undergo a free screening mammogram and every two years thereafter until the age of 69. These regions roughly represent 25% of the Portuguese population. Although we must add a caveat here, because this spatial information is regarding only the current place of residence, and no other spatial information such as place of birth or where women grew up as a child is known. More details about the screening program and the inclusion criteria are given elsewhere [22, 26, 27].

Age at menarche (AaM) registered in years at the first interview, was the response variable of interest. The average was 13.18 with a minimum of 8 and a maximum of 24. We could raise the question about the existence of bias in the reported ages, but many studies have demonstrated that this reported age is quite accurate, most likely due to the emotional significance of menarche for a young girl [28]. Figure 1 shows the histogram for the unconditional distribution of this variable with a modal class at 13 years old. Summarized in Table 1 are its descriptive percentiles by decade from 1920 to 1973, and Table 2 depicts the mean and the median by municipality and decade. Year of birth ( $B_{Year}$ ) and the demographic information given by the municipality of residence ( $Muni$ ) were the independent variables.

### Methods

A typical regression analysis has the advantage of being very well known to the users, although it is very likely that in many scenarios other properties of the response distribution (e.g. the variance) may also depend on covariates.



**Fig. 1** Age at menarche. The variable has a positive skewed distribution (black line), so one must consider a distribution allowing for positive skewness

**Table 1** Summary of the AaM percentiles by decade of birth since 1920 until 1973. Last column presents the total number of women screened per decade of birth

AaM (Percentiles)								
Decade	5	10	25	50	75	90	95	women
20's	11	12	12	14	15	17	18	19844
30's	11	11	12	14	15	17	17	72206
40's	11	11	12	13	14	16	17	109335
50's	11	11	12	13	14	15	16	122045
60's	10	11	12	13	14	15	16	109266
70's	10	11	12	12	14	15	15	19652

Besides this, one may want to comprehend not only the behaviour of the average person, but the behaviour of those belonging to the population's extremes. To account for these features, we will consider a statistical framework developed within the context of Generalized Additive Models for Location, Scale and Shape (GAMLSS) [29], also known as distributional regression models [30]. These models have several advantages, but for our study the principal characteristic is that it assumes a known parametric family of distributions for the response variable which allows an easy calculation of the conditional percentile curves of the AaM given birth year and municipality of residence. Furthermore, the covariate effects of interest can be of flexible forms (e.g. smoothing functions) and not restricted to the traditional, and perhaps unrealistic, linear effect. At the same time, it ensures that the adjusted percentile curves do not cross. A competing method is quantile regression [31], where the adjusted curves might cross, because it does not assume a

distribution for the response variable and, therefore, can be considered in the realm of non-parametric methods [25]. For their part, other authors have been considering other approaches. Most use simple statistical methods like ANOVA [32] or linear regression [12]. Chumlea et al. [33] considered a probit analysis.

#### Statistical models

Based on data for  $i = 1, \dots, N = 452\,348$  women, we will assume conditional independence of the individual ages at menarche,  $A_{aM}$ , given the covariates  $B_{year}$  and  $M_{uni}$ .

An initial exploratory analysis showed that the distribution of the  $A_{aM}$  is right skewed (*vide* Fig. 1), and a non-linear assumption for the relation between the  $A_{aM}$  and  $B_{year}$  is better supported by the data. Taking this into consideration, we opted to analyse the data with a model within the aforementioned class – GAMLSS. This method offers a highly flexible approach in that constraints to the traditional distributional assumptions,





such as normality, are removed. The method enables the use of skewed distributions without having to transform the data, allowing us to work on the scale of the data, which is a very important feature. Additionally, as already said, all the parameters of the response probability distribution can be modelled by explanatory variables and not only the location. For instance, it is possible to model the variance and the skewness (for some distributions it is also possible to model the kurtosis). All the explanatory variables are introduced into the model parameters through predictors, which can be linear functions of the explanatory variables or can take the form of structured additive predictors with non-linear or smoothing functions of explanatory variables.

The generalized Akaike Information Criterion (GAIC), a model selection measure, was considered to determine the best fitting distribution of the data. It was found that the *Box-Cox Cole and Green* distribution,  $BCCG(\mu, \sigma, \nu)$ , provides the lower GAIC value (i.e. the best fit) when compared to other alternative distributions with continuous support. Specifically, we also considered the Normal, the *t* and the *Box-Cox Power Exponential* distributions, which are all available in GAMLSS. The main statistical model for analysing the data within the GAMLSS framework that we will be dealing with is:

$$\begin{aligned} \text{AaM} &\sim BCCG(\mu, \sigma, \nu), \\ \mu &= \beta_{10} + f_{11}(\text{Byear}) + f_{12}(\text{Muni}), \\ \log(\sigma) &= \beta_{20} + f_{21}(\text{Byear}) + f_{22}(\text{Muni}), \\ \nu &= \beta_{30}. \end{aligned} \quad (1)$$

We defined an additive model for the location parameter,  $\mu$ , which for this distribution represents its median, a very relevant parameter in our work, as we are interested in estimating the percentiles. The linear predictor defined as  $\mu = \beta_{10} + f_{11}(\text{Byear}) + f_{12}(\text{Muni})$  allows for directly modelling the median. For the scale of the response variable,  $\sigma$ , we let the parameter depend on the explanatory variables *Byear* and *Muni*. It is a multiplicative model resulting from the log-link, which in turn ensures positive values for the parameter. For this distribution the scale parameter is approximately the coefficient of variation. The skewness parameter,  $\nu$ , is modelled only with an intercept term. It was found that adding covariates (e.g. age) to its linear predictor did not improve the adjustment because those models produced greater GAIC values.

The functions  $f_{11}$  and  $f_{21}$ , modelled as cubic splines, represent the temporal effects of the year of birth, and  $f_{12}$  and  $f_{22}$  are the spatially correlated effects of the residence municipalities modelled as an intrinsic autoregressive process (IAR), a limiting case of the conditional autoregressive models (CAR) of Besag et al. [34]. Thus, we are

considering that the spatial effect is a Markov random field (MRF), i.e. assuming that the spatial random effects (our spatial variables) have a joint distribution which is specified by considering conditional independence locally. The IAR models are a typical choice when dealing with a dependent variable observed in geographical areas sharing borders, because we expect neighbouring areas to have more similar observed values for the AaM than areas farther apart. The consequence of this approach is that the parameter estimates for neighbouring locations are shrunk towards its mean.

## Results

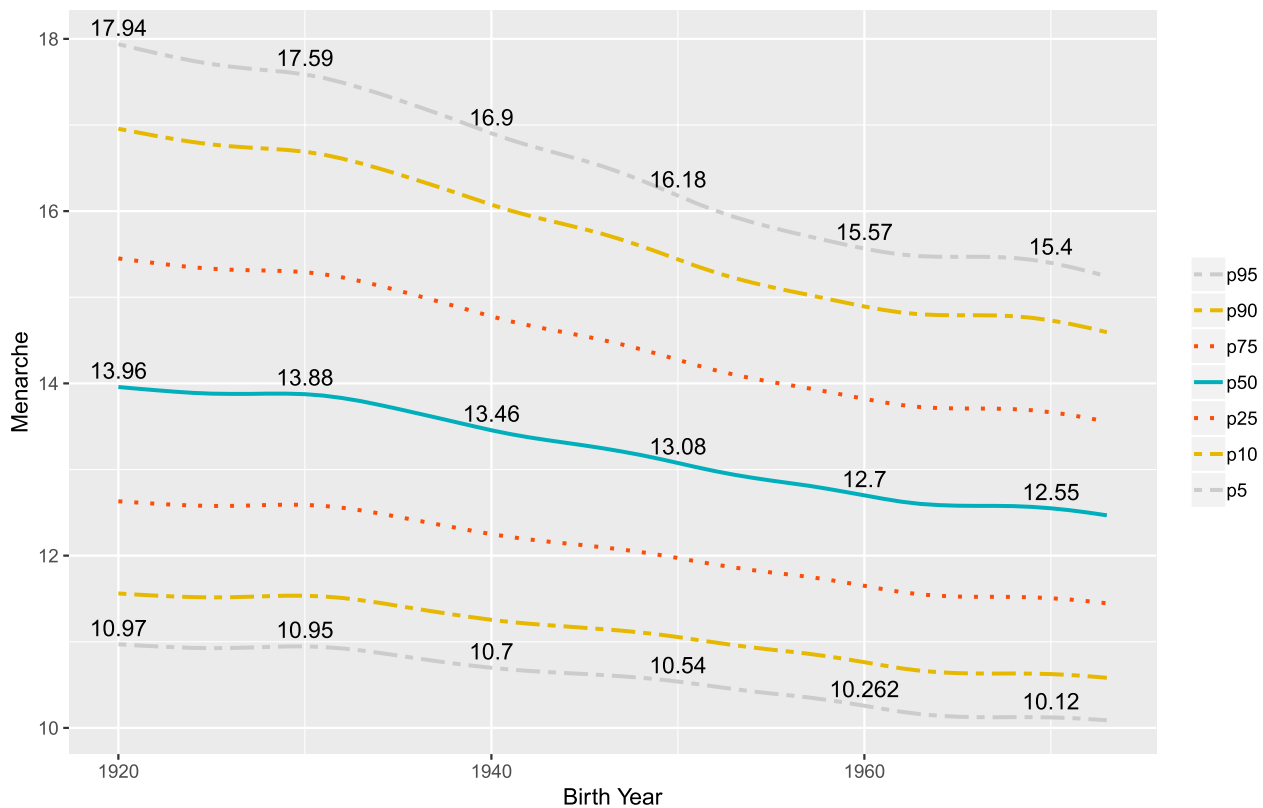
The conditional percentile values were thus obtained via the R software (v 4.1.0), namely using the main package *gamlss* (version 5.1-4) and two additional packages that provide a set of functions to fit models with spatial variables (*gamlss.spatial* and *gamlss.add*).

The chosen model is written below in terms of the R syntax based on the *gamlss* package:

```
m1<-gamlss(AaM ~ cs(Byear) +
  ga(~s(Muni, bs = "mrf", xt=xt1),
    method="REML"),
  sigma.fo = ~ cs(Byear) +
  ga(~s(Muni, bs = "mrf", xt=xt1),
    method="REML"),
  nu.fo = ~1,
  tau.fo = ~1,
  family = BCCG, data=MenaAge)
```

## Age at menarche – temporal trends

A first distributional regression model without covariates was considered in order to estimate the median AaM for the overall population since 1920, resulting in a point estimate of 13.05 years old. Then we limited the model to one explanatory variable *Byear* for exploring the temporal trends in the central region of Portugal. It should be noted that the percentiles curves displayed in Fig. 2 are not linear. This behaviour could not be captured within a typical linear regression analysis and is facilitated by the non-linear approach permitted by the generalized additive models. Additionally, the percentiles 25%, 50%, 75%, 90% and 95% are steadily decreasing at a greater rate since 1920 than the percentiles 5% and 10%, which could mean that there is not much more space for reductions in the menarche age for the girls with the earliest menarches. The stagnation trend reported by other studies also appears to be emerging, mainly in the 1960s. From 1970 onwards the downward trend reappears, which is in line with the results published in [18].



**Fig. 2** Estimated percentile curves for the AaM by year of birth since 1920. The values printed are for the percentiles 5%, 50% and 95% for the years 1920, 1930, ..., 1970

### Age at menarche – spatio-temporal trends

Each municipality is likely to have the median and the variability of the AaM close to those of its neighbours. Bearing this in mind, spatial effect,  $f_{12}(\text{Muni})$  and  $f_{22}(\text{Muni})$ , were added to the model in (1) aiming at capturing the influence of neighbouring locations not available by the observed covariates. Figures 3 and 4 present the spatio-temporal trends in the percentiles for the AaM, which are already accounting for the spatial dependency of the response variable (AaM). It is clear that the patterns through the decades have remained unchanged, i.e. regions with the larger values in 1920 maintain their position in 1970, although all have been decreasing, with only a few exceptions. The northern municipalities tend to present larger values for all the considered percentiles.

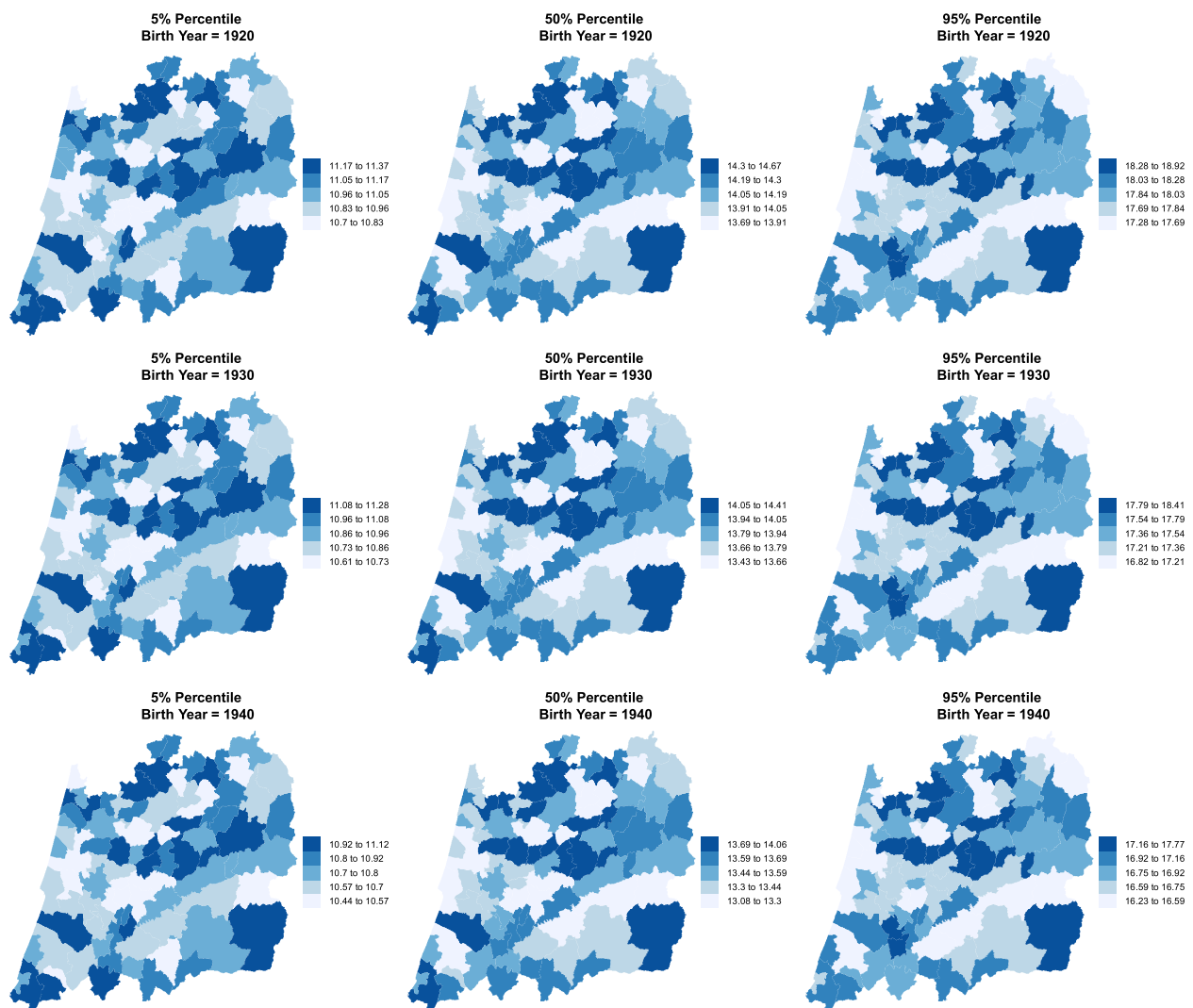
Figure 5 depicts the effects of each of the covariates (Byear and Muni) on the AaM location parameter, i.e. its median. The interpretation is relatively straightforward. A positive time effect was found until approximately 1950 and thereafter a negative effect. This means that a woman born before 1950 had an AaM above the median for the overall population and accounting all the years, i.e. 13.05, and a woman born after 1950 will tend to have an AaM below 13.05. For the spatial effects, a

municipality with a negative effect means that women residing there have ages at menarche below the median when comparing to the overall population. On the other side, areas with a positive spatial effect will tend to have ages at menarche above the median of the overall population. Looking to Figs. 3 and 4 and comparing them to Fig. 5, it is clear that larger values of the median AaM are associated with a positive spatial effect and *vice-versa*.

Another interesting feature of the temporal effect for the year of birth is obtained considering the `getPEF` function (Partial Effect function) which allows us to calculate the slope of the curve at each year (*vide* Table 3). For example, for the year 1960, in `gamlss` syntax this can be done by

```
slope <- getPEF(m1, "Byear", plot=TRUE)
slope(1960, deriv=1)
```

and the result is the slope of the curve's tangent for the year 1960, which is not very different from the other years, at least until 1960. The result is approximately  $-0.04$ , meaning that for that year the median AaM was reduced at a rate of  $0.04 \times 365 \approx 14.6$  days, or 4.87 months per decade. This value is in accordance with



**Fig. 3** Estimated AaM by region and birth year. Percentiles 5%, 50% and 95% for the central region of Portugal for the years 1920, 1930 and 1940

the findings of other studies, e.g. [35] which examines the evolution of the menarche in the period 1830–1960 in Europe and reported a decreasing of about 3 months per decade. For the year 1970, we have seen from Fig. 2 a curve with a horizontal aspect, meaning a stabilization. Calculating the slope at this time, the value of  $-0.013$  appears, meaning that the median AaM is reducing in this year at a rate of  $0.013 \times 365 = 4.745$  days, or 1.58 months per decade, pointing to a slowing down in the reduction rate.

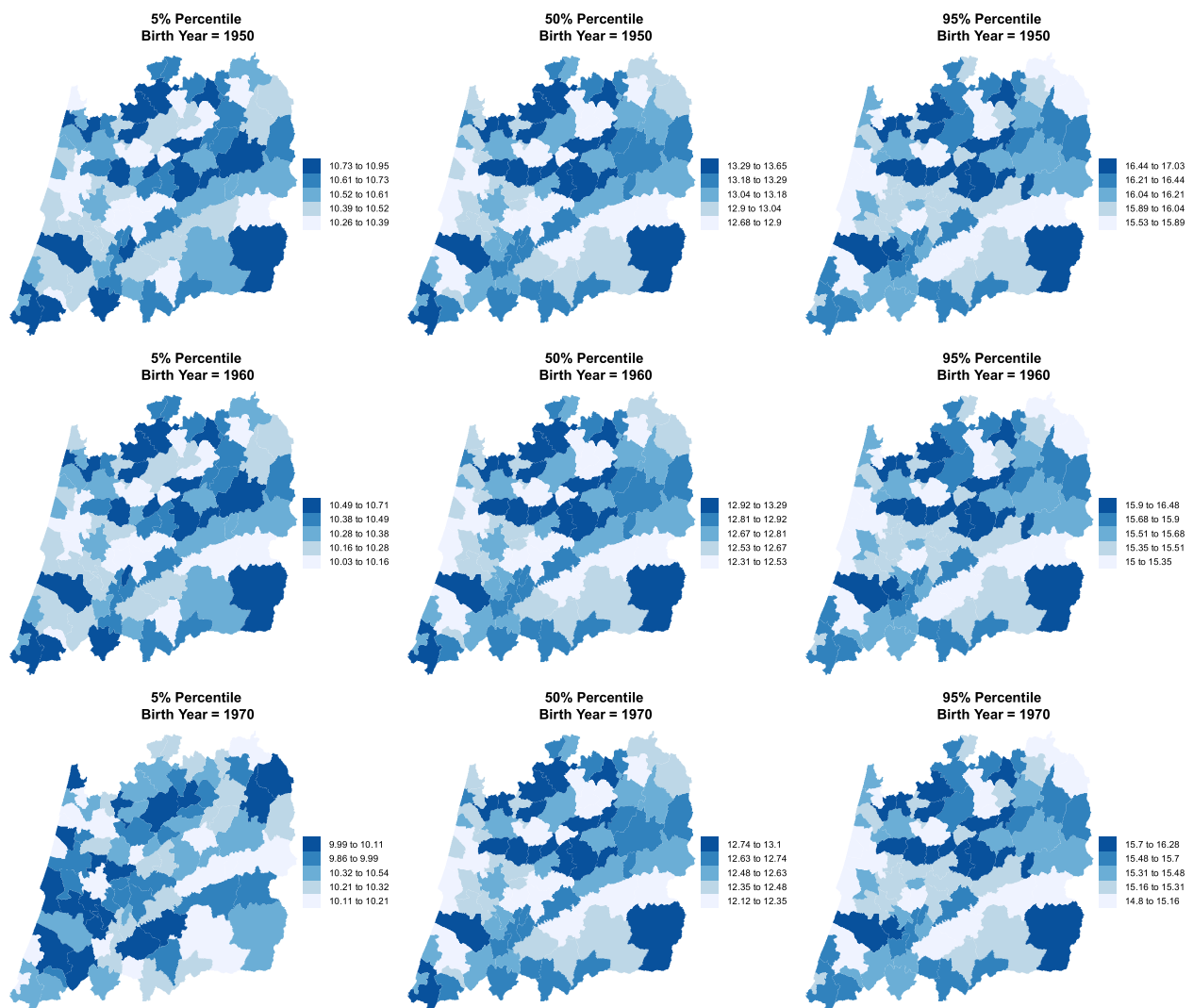
## Discussion

Estimation of the percentile curves for a response variable is widely used in medicine for checking whether an individual has an abnormally low or high value of the response variable (given the covariates of interest), and hence whether she/he is potentially at risk. For

example, let us think of a region where breast cancer risk is high and that girls living there tend to have an early menarche when compared to the rest of the country. And suppose that all other characteristics generally known for being linked to breast cancer are equal across that country. That specific region may be the target of differentiated public policies.

The dataset here analysed is based on the largest sample size reported to date in Portugal, although it does not include women born in the last three decades, meaning that the percentiles of these youngest women cannot be ascertained. Nevertheless, and to the best of our knowledge, this study is the first to analyse and produce percentile curves for the women residing in the Central Portugal since 1920 and to use a distributional regression approach.





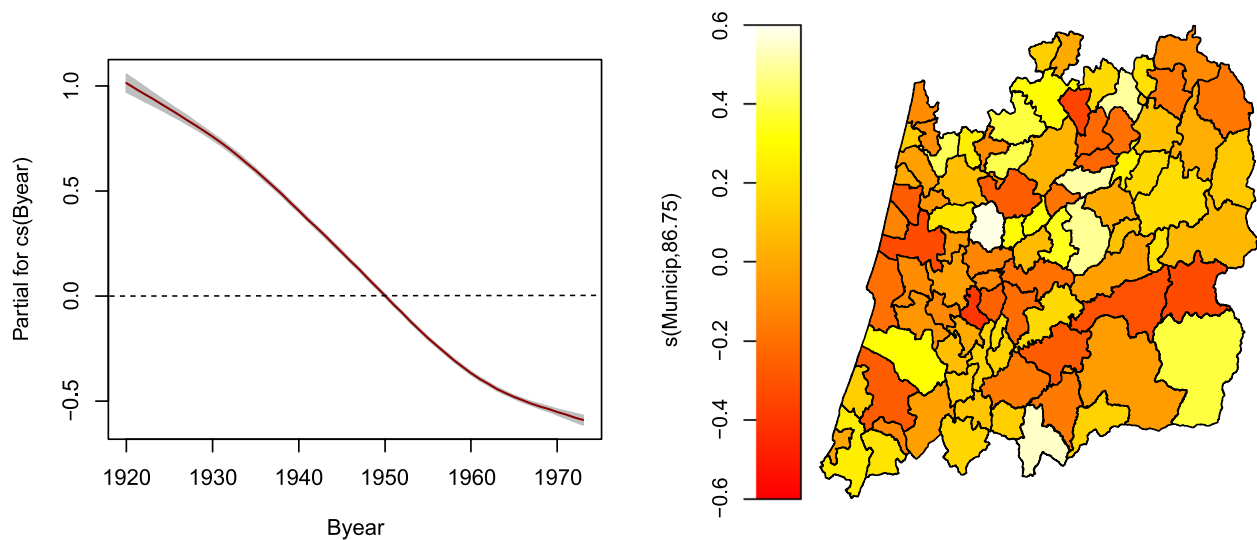
**Fig. 4** Estimated AaM by region and birth year. Percentiles 5%, 50% and 95% for the central region of Portugal for the years 1950, 1960 and 1970

Probably not all spatial effects are being correctly captured because the spatial information in the data set regards the place of residence at the time of the participation in the screening program, i.e. when the women are adults, as already mentioned in the data description. However, we assume that these women have always been living there. This is more likely to be true for the interior areas than for the areas closest to the coast given that municipalities in Central and Eastern Portugal have lower levels of wealth. Despite the above, we do believe that the spatial results are quite robust because of the dimension of the sample here analysed. That said, some works, e.g. [36] show that the place of residence, and even the altitude of the residence area [37] during childhood, are likely to be more effective in estimating regional differences for the AaM.

The probability distribution that we chose for describing the AaM,  $BCCG(\mu, \sigma, \nu)$ , has several advantages for have being considered here. First its location parameter,  $\mu$ , is interpreted as the median of the distribution, which when describing the percentiles is a benefit. Additionally, we are free to control the shape parameter,  $\nu$ , representing the amount of skewness of the distribution, and from Fig. 1 we know that our distribution is positive skewed.

## Conclusion

We resorted to a distributional regression model to estimate several spatio-temporal percentiles for the AaM in the central region of Portugal considering a long period of time and using a representative sample with a large number of observations. The results show that within the period 1920 – 1973, all the percentiles obtained for the AaM depict a decreasing trend. For a girl born in 1920,



**Fig. 5** Estimated effects for the AaM model showing the temporal variation of the birth year effects (left) together with the clustered spatial effects (right)

we would expect a median AaM of around 14 years. For a girl born in 1973, this value is about 1.5 years lower (Fig. 2). Interestingly, a pattern of AaM distribution by municipality emerges. Although all its percentiles are decreasing, the regions with the highest percentile values for AaM in 1920 are the same for 1970 (Figs. 3 and 4), with only a few exceptions.

In the middle of the 20th century the decreasing rate achieved its maximum of around 15.23 days per year (Table 3). This trend after 1973 seems to be slowing down or even stopping. We should note that the median's rate of decrease is not replicated by the other percentiles. For instance, percentiles 90% and 95% decreased about 2.6 years in the span of 50 years. On the other hand, percentiles 5% and 10% decreased only

about 0.8 years in the same range of 50 years. Early menarches (below percentile 5%) occur in the coastal (western) area of Portugal's central region for all the decades considered and later menarches (above percentile 95%) occur in the central-north area.

#### Acknowledgements

The authors would like to thank the Portuguese Cancer League for providing the data.

#### Authors' contributions

VR, RM and BS performed the data analysis and wrote the manuscript. All authors then revised and approved the manuscript prior to submission.

#### Funding

This work was partially funded by Fundação para a Ciência e a Tecnologia (FCT) through the projects UIDB/00006/2020 and POCI/01/0145/FEDER/029443 – SHSADReM – Addressing Social and Health Challenges through new developments in Structured Additive Distributional Regression Models.

#### Availability of data and materials

The data sets used and/or analysed during the current study are available from the corresponding author on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

This project has used data from the Breast Cancer Screening data from Portugal, which have ethical approval from the Faculty of Psychology of the University of Coimbra Ethics Committee and the Portuguese Cancer League. Usage of data derived from the records is according to Portuguese and European laws and regulations. All women signed the informed consent prior to the screening procedure. The study was carried out according to the tenets of the Declaration of Helsinki. All methods were carried out in accordance with relevant guidelines and regulations.

##### Consent for publication

Not applicable.

**Table 3** Estimated decreasing rates per year (in days) of the median AaM for girls born since 1920 up to 1970

Birth Year	Decreasing Rate
1920	8.91
1925	9.1
1930	10.47
1935	13.04
1940	14.43
1945	14.71
1950	15.23
1955	13.72
1960	10.48
1965	6.26
1970	4.75

**Competing interests**

The authors declare no competing interests.

Received: 4 October 2022 Accepted: 17 August 2023

Published online: 07 September 2023

**References**

- Papadimitriou A. The evolution of the age at menarche from prehistorical to modern times. *J Pediatr Adol Gynec.* 2016;29(6):527–30.
- Harris MA, Prior JC, Koehoorn M. Age at Menarche in the Canadian Population: Secular Trends and Relationship to Adulthood BMI. *J Adolesc Health.* 2008;43(6):548–54. <https://doi.org/10.1016/j.jadohealth.2008.07.017>.
- Freedman D. Trends in Reproductive, Smoking, and other Chronic Disease Risk Factors by Birth Cohort and Race in a Large Occupational Study Population. *Ann Epidemiol.* 2002;12(6):363–9. [https://doi.org/10.1016/S1047-2797\(01\)00295-2](https://doi.org/10.1016/S1047-2797(01)00295-2).
- Malina RM, Reyes MEP, Tan SK, Little BB. Secular change in age at menarche in rural Oaxaca, southern Mexico: 1968–2000. *Ann Hum Biol.* 2004;31(6):634–46. <https://doi.org/10.1080/03014460400018085>.
- Won YJ, Jung KW, Oh CM, Park EH, Kong HJ, Lee DH, et al. Geographical Variations and Trends in Major Cancer Incidences throughout Korea during 1999–2013. *Cancer Res Treat.* 2018;50(4):1281–93. <https://doi.org/10.4143/crt.2017.411>.
- Gottschalk MS, Eskild A, Hofvind S, Gran JM, Bjelland EK. Temporal trends in age at menarche and age at menopause: a population study of 312,656 women in Norway. *Hum Reprod.* 2020;35(2):464–71. <https://doi.org/10.1093/humrep/dez288>.
- Liu W, Yan X, Li C, Shu Q, Chen M, Cai L, et al. A secular trend in age at menarche in Yunnan Province, China: a multiethnic population study of 1,275,000 women. *BMC Public Health.* 2021;21(1). <https://doi.org/10.1186/s12889-021-11951-x>.
- Nabhan AF, Mburu G, Elshafeey F, Magdi R, Kamel M, Elshebiny M, et al. Women's reproductive span: a systematic scoping review. *Hum Reprod Open.* 2022;2022(2). <https://doi.org/10.1093/hropen/hoac005>.
- Gaudineau A, Ehlinger V, Vayssières C, Jouret B, Arnaud C, Godeau E. Factors associated with early menarche: results from the French Health Behaviour in School-aged Children (HBSC) study. *BMC Public Health.* 2010;10(1):175.
- Kaplowitz PB. Link between body fat and the timing of puberty. *Pediatrics.* 2008;121(Supplement 3):S208–17.
- Lazzeri G, Tosti C, Pammolli A, Troiano G, Vieno A, Canale N, et al. Overweight and lower age at menarche: evidence from the Italian HBSC cross-sectional survey. *BMC Womens Health.* 2018;18(1):168.
- Cabanés A, Ascunce N, Vidal E, Ederra M, Barcos A, Erdozain N, et al. Decline in age at menarche among Spanish women born from 1925 to 1962. *BMC Public Health.* 2009;9(1):449.
- Sohn K. Improvement in the biological standard of living in 20th century Korea: Evidence from age at menarche. *Am J Hum Biol.* 2017;29(1):e22882.
- Onland-Moret N, Peeters P, Van Gils C, Clavel-Chapelon F, Key T, Tjønneland A, et al. Age at menarche in relation to adult height: the EPIC study. *Am J Epidemiol.* 2005;162(7):623–32.
- Lalys L, Pineau JC. Age at menarche in a group of French schoolgirls. *Pediatr Int.* 2014;56(4):601–4.
- Talma H, Schönbeck Y, van Dommelen P, Bakker B, van Buuren S, HiraSing RA. Trends in menarcheal age between 1955 and 2009 in the Netherlands. *PLoS ONE.* 2013;8(4):e60056.
- Finer LB, Philbin JM. Trends in ages at key reproductive transitions in the United States, 1951–2010. *Womens Health Issues.* 2014;24(3):e271–9.
- Padez C. Social background and age at menarche in Portuguese university students: a note on the secular changes in Portugal. *Am J Hum Biol.* 2003;15(3):415–27.
- Parent AS, Rasier G, Gerard A, Heger S, Roth C, Mastrorandi C, et al. Early onset of puberty: tracking genetic and environmental factors. *Horm Res Paediatr.* 2005;64(Suppl. 2):41–7.
- Burgess S, Thompson DJ, Rees JM, Day FR, Perry JR, Ong KK. Dissecting causal pathways using Mendelian randomization with summarized genetic data: application to age at menarche and risk of breast cancer. *Genetics.* 2017;207(2):481–7.
- Mokha R, Kaur Al, Kaur N. Age at menarche in urban-rural Punjabi Jat Sikh girls. *Anthropologist.* 2006;8(3):207–9.
- Duarte E, de Sousa B, Cadarso-Suarez C, Rodrigues V, Kneib T. Structured additive regression modeling of age of menarche and menopause in a breast cancer screening program. *Biom J.* 2014;56(3):416–27.
- Kneib T. Beyond mean regression. *Stat Mod.* 2013;13(4):275–303.
- Stasinopoulos MD, Rigby RA, Bastiani FD. GAMLSS: a distributional regression approach. *Stat Mod.* 2018;18(3–4):248–73.
- Rigby R, Stasinopoulos D, Voudouris V. Discussion: A comparison of GAMLSS with quantile regression. *Stat Mod.* 2013;13(4):335–48.
- Alves JGR, Cruz DB, Rodrigues VL, Gonçalves ML, Fernandes E. Breast cancer screening in the central region of Portugal. *Eur J Cancer Prev.* 1994;3:49–53.
- Bastos J, Rodrigues V, Paap E, Broeders M, Pina M, Cruz D, et al. Breast cancer screening effectiveness in Portugal central Region. *Eur J Cancer Prev.* 2017;26:S204–7.
- Dorn LD, Sontag-Padilla LM, Pabst S, Tissot A, Susman EJ. Longitudinal reliability of self-reported age at menarche in adolescent girls: Variability across time and setting. *Dev Psychol.* 2013;49(6):1187.
- Rigby R, Stasinopoulos D. Generalized additive models for location, scale and shape,(with discussion). *Appl Stat.* 2005;54:507–54.
- Umlauf N, Kneib T. A primer on Bayesian distributional regression. *Stat Mod.* 2018;18(3–4):219–47.
- Koenker R. Quantile Regression. vol. 38. Cambridge University Press; 2005. [https://scholar.google.com/scholar?hl=pt-PT&as\\_sdt=0%2C5&q=Koenker+R.+Quantile+Regression.+vol.+38.+Cambridge+University+Press+2005&btnG=](https://scholar.google.com/scholar?hl=pt-PT&as_sdt=0%2C5&q=Koenker+R.+Quantile+Regression.+vol.+38.+Cambridge+University+Press+2005&btnG=).
- Lee MH, Kim SH, Oh M, Lee KW, Park MJ. Age at menarche in Korean adolescents: trends and influencing factors. *Reprod health.* 2016;13(1):121.
- Chumlea W, Schubert C, Roche A, Kulin H, Lee P, Himes J, et al. Age at menarche and racial comparisons in US girls. *Pediatrics.* 2003;111(1):110–3.
- Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann I Stat Math.* 1991;43(1):1–59.
- Tanner JM, Eveleth PB. Variability between populations in growth and development at puberty. *Puberty, Biologic and Psychosocial Components.* Leiden: HE Stenfert Kroese; 1975. p. 256–73.
- Okasha M, McCarron P, Smith GD, McEwen J. Age at menarche: secular trends and association with adult anthropometric measures. *Ann Hum Biol.* 2001;28(1):68–78.
- Jansen E, Herrán O, Fleischer N, Mondul A, Villamor E. Age at menarche in relation to prenatal rainy season exposure and altitude of residence: results from a nationally representative survey in a tropical country. *J Dev Orig Hlth Dis.* 2017;8(2):188–95.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

