Working paper #2

District characteristics predicting vulnerability to heat

Hanns Moshammer[1]

[1] Department for Environmental Health, ZPH, Medical University of Vienna

March 2024

District characteristics predicting vulnerability to heat

Temporal variation of temperature is associated with (small) changes in mortality rates. Both high and low temperatures cause higher numbers of (daily) deaths in the affected population. Previous studies have not only demonstrated this non-linear (U-shaped) association between temperature and mortality, but also that high temperatures tend to have a more immediate effect than cold temperatures. Whereby at high temperatures the same day temperatures have the strongest effect, at cold temperatures the last weeks' averages are more relevant. This non-linear and differential temporal distribution poses some challenge for quantitative effect estimates. Besides, effects of temperature are small relative to the day-to-day noise in mortality numbers. The smaller the population group (e.g. district population) under study, the more difficult is the detection of the temperature effect in spite of the statistical noise of the data. This problem is enhanced when specific causes of death are to be examined. Besides, while the fact of death is usually undisputed, the cause of death is often less straightforward. Extreme temperatures cause stress and especially old people and people with chronic diseases might more easily succumb to this stress. In that case, the underlying chronic disease is usually entered as the cause of death. Therefore, cause of death is not very specific for temperature related mortality.

Method

Since the cause of death for temperature-related mortality is not very specific, it was decided to analyze daily all-cause mortality per district in Austria. To overcome statistical uncertainties, we aimed for a rather long observation period. Indeed, we obtained daily deaths (per district of the home address) for the years 1970 until 2020. A daily time-series of 51 years is exceptionally long and allows for rather precise estimates of temperature effects. In any epidemiological time-series, temporally varying confounders must be controlled for. Above all, these are long-term trends in average number of deaths (due to changes in population number and structure), seasonal variation, and day of the week. Long-term and seasonal variation can be controlled for through a natural spline. In accordance with the APHEA (respectively APHENA) protocol, we optimized the number of knots for that natural spline by minimizing the sum of the partial autocorrelation [1] and including day of the week as a dummy variable. We selected the optimal number of knots based on daily deaths in all of Austria. In this data-set with large daily numbers, distribution of daily deaths approximated a Poisson distribution well enough to allow for a Poisson regression. The number of knots (320 in total or about 6 knots per year) was then also tested on the cities of Vienna, Graz, Linz and Klagenfurt and on a smaller rural district (Wolfsberg). While the number of knots worked well also with these districts, the assumption of a Poisson distribution no longer held for the smaller districts. Therefore, for the per-district analysis, we decided to do a negative binomial regression analysis (General Additive Model – GAM – with family negative binominal). Because the negative binomial regression needs repeated iterations, the computational time increase substantially with a full model (including meteorological parameters) per district taking between 30 and 60 minutes each.

Thus, in a first step, this base model was calculated for every district (using R version 4.0.3):

Formula 1: gam (Deaths~s(date, bs="cr", k=320)+as.factor(Day_of_Week),data=1970_2020, family=nb)

The residuals from each district were stored for further use.

Step 1: Division of Districts

Some districts have been changed over the course of the years 1970 till 2020. Meteorological data were provided by the Wegener Center for Climate and Global Change (University of Graz) for the current districts. In Lower Austria, the district "Wien Umgebung" was terminated in 2016 and parts of the district were allocated to the neighboring districts Tulln, Korneuburg, Bruck an der Leitha, and St. Pölten Land. Therefore, also for these four districts, a continuous time-series could not be established. In Burgenland, the small town of Rust is for historical reasons a district of its own. But both its area and its population are too small to allow for meaningful statistical analysis. As it lies embedded in the larger district Eisenstadt Umgebung, the daily deaths in Rust were added to those in the larger district and the meteorological data for Eisenstadt Umgebung were used in the models for the combined districts. In Styria, in or around 2012 four new districts were created out of two old districts each. Thus, the new districts could be treated as if they existed already since 1970 by simply adding the daily deaths of the 2 old districts that made up the respective new district. In total, this allowed for the analysis of 111 districts.

Step 2: Selection of temperature

The Wegener Center had modelled temperature data on a 1-km grid over the whole area of Austria. For each day, they provided the daily maximum, minimum, and average temperature. In addition, they also calculated various heat indices, namely Tropical Nights, Summer Days, Hot Days, Heat Episode, and Kysely Days. In addition, they calculated various indices based on temperature and humidity, namely Heat Index, Humidex, and Wet Bulb Global Temperature. These were calculated for the mean and the maximum temperature of the day.

Since all these measures of temperature were provided on a 1-km grid and not per district, there remained the question how to arrive at temperature measures representative of each district. For that, three different approaches were applied: The first approach just calculated the unweighted average over all grid points within the respective district. The second approach weighted each grid point according to the population density in the respective grid square so that grid point temperatures in more populated areas of the district received a stronger weight when calculating the district average. The third approach reasoned that in an alpine district, most of the population would be clustered in the low-lying parts of the district, where the temperature would also be the highest. Therefore, this approach simply assumed the hottest grit point to be representative of the district.

To select the best measure of temperature, a step-wise approach was conducted for selected districts, namely again the cities of Vienna, Graz, Linz and Klagenfurt and on a smaller rural district (Wolfsberg). First, the mean unweighted temperature measures were compared in a simple model, that just added same day temperature measure, the average of the past 14 days' temperature, and the square of the same day temperature to the base model. In all the tested districts, mean temperature performed best according to the Akaike Information Criterion. It performed better than the other measures of temperature (minimum and maximum temperature) and also better than the combined indices also accounting for humidity. Upon adding binary indicators of hot or very hot days, this did not improve model performance. Among the indicators of heat episodes, only the definition according to Kysely provided a slightly better model fit. Indeed, adding a binary variable of "Kysely Day" provided a slightly better fit than only examining the continuous effect of temperature. But this gain was not considered worth the effort, because the gain was so small. Therefore, it was decided to investigate only the effect of "mean temperature" further.

Next it was tested, which approach to represent district-wide temperature based on grid data was the best. In that case, the population weighted mean performed best in nearly all tested districts. Only in Linz, the unweighted mean performed slightly better. Therefore, it was decided to use the population weighted mean of the daily mean temperature.

Step 3: Testing and comparing different models

Already in the selection process of the best temperature measure, a "simple" model including same-day temperature, same-day temperature squared, and arithmetic mean of temperature over the past

14 days was applied. This decision was based on previous studies on temperature effects in Vienna [2]. In that study we had demonstrated for Vienna, that this parametric model represented a fair approximation of the more complex temperature-effect association as modelled with a non-linear distributed lag model. Indeed, as now using the current base model and the population-weighted means of the daily mean temperatures, this simple model was not inferior to the non-linear distributed lag model according to the Akaike Information Criterion, this time tested in the 3 largest cities (Vienna, Graz, Linz). As demonstrated in Figure 1 for Vienna, the distributed lag effects are rather complex, with very strong acute effects of hot days and not so strong, but longer lasting effects of cold days. But that complex model comes with the disadvantage of many more parameters used respectively degrees of freedom lost. In addition, for the final step of a meta-regression, easy to interpret parametric effect estimates were needed.

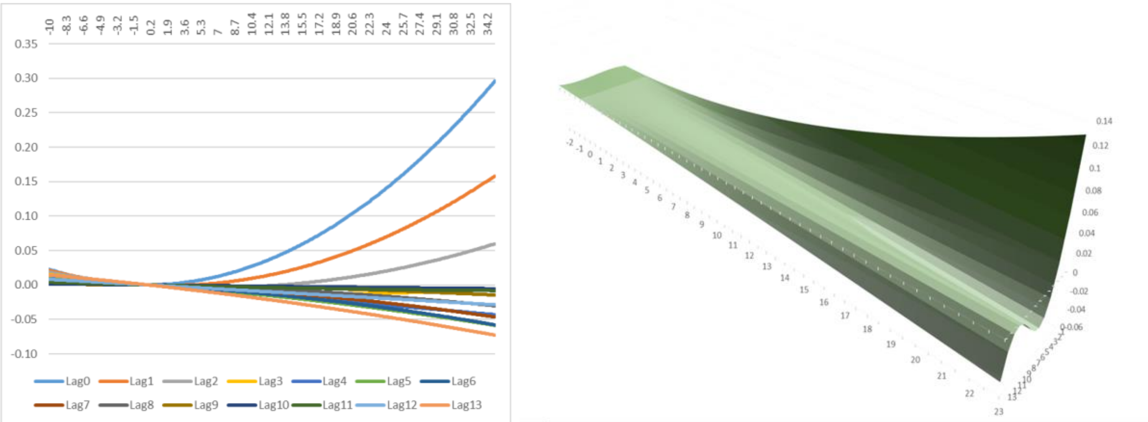

Figure 1: Distributed lag model of temperature effects in Vienna (daily number of deaths, 1970-2020). Same-day temperature (lag 0) has a very strong effect on hot days. With increasing latency (up to 14 days, lag 13), a nearly linear effect of cold temperature develops.

The non-linear effect of temperature can also be examined using natural splines. This is again exemplified for Vienna in Figure 2. It demonstrates that the chronic effect can be represented by a linear association, while the acute effect can either be approximated by a linear threshold model or by a parabolic function (temperature and temperature squared). For the sake of simplicity, the latter version was chosen. In a further step the residuals from the base model were also used to run a linear threshold model.

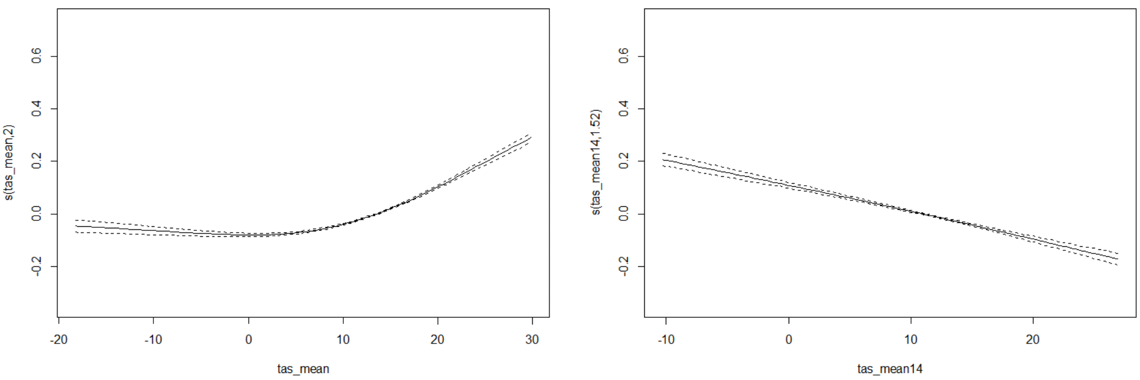

Figure 2: Non-linear effect of same-day temperature (left panel) and of 14-day average of temperature (right panel, always non-weighted average of the mean temperatures of the grid-points).

As expected, the 14-day average always had a negative and usually highly significant coefficient. The squared temperature was positive in nearly all districts and usually significantly different from zero. The same-day temperature had values below or above zero and only reached significance in some districts. The 10-day average thus displays a clear fairly linear effect with decreasing temperatures. The squared temperature is mostly driven by the extreme hot temperatures. With a coefficient of zero for the same-day temperature, the parabola defined by temperature and temperature squared would have its minimum at zero degree Celsius. Of course, this temperature-effect curve is shifted because the adverse effect of low temperatures is already mostly covered by the sub-chronic exposure effects. Without the 14-day average, the parabola would automatically shift to the right (to warmer temperatures). The parabola describing the effects of same-day temperature would shift to the right (to warmer temperatures), if the coefficient of "temperature" was negative, and to the left (to cooler temperatures), if it was positive.

Thus, on the one hand, a full negative binomial models in R (the base model, Formula 1) was run plus same-day temperature, same-day temperature squared, and 14-day running average temperature. On the other hand, also the residuals of each district from the base model were used. These residuals were first meant to be used for another project, where the effect of air pollutants and possible interaction between pollutant and temperature should have been examined. But because data on air pollution were not available for the whole range from 1970 till 2020 and because the period with available data varied between districts and the same base model should have been used for all districts, the residuals had to be calculated first in order to run separate linear models on the effect of temperature and pollutants on the residuals.

Because a threshold model is easily implemented in STATA (formula 2), the residuals were used to run a threshold model:

Formula 2: nl(residual={coeff}*AVA+(temp>{threshold=10})*(temp-{threshold})*{coeff2}+{const}),
where AVA is the 14-day running average temperature, and temp is the same-day temperature.

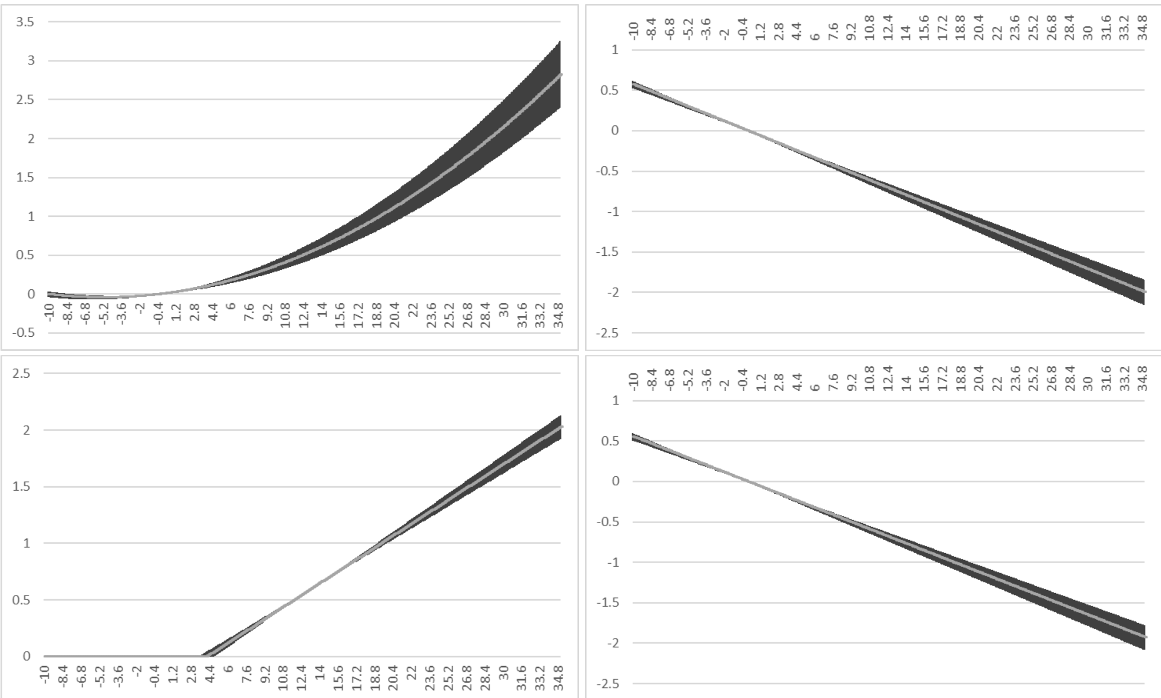

Figure 3: Non-linear effect of same-day temperature (left panel) and of 14-day average of temperature (right panel, always weighted average of the mean temperatures of the grid-points). Example: total of Vienna (23

districts). Top panel: quadratic model, bottom panel: linear threshold model for same-day temperature. Effects on residuals from the base model.

This threshold model accounted for a linear association between residuals (i.e. daily deaths after controlling for the base model) and the 14-day average temperature (expecting a negative association), and for a threshold model with no effect below the threshold and a linear increase of same-day temperature above that threshold.

Interestingly, when the two models (14-day average for both, either linear threshold model or quadratic model for same-day temperature) were applied on the residuals for the total of Vienna (Figure 3), the results differed slightly from the full model run in R (Figure 2). Nevertheless, the two models that were both applied on the residuals, gave comparable results. Also, the main facts (clear negative temperature-effect line for the 14-day running average, clearly increasing case numbers with increasing same-day temperature above a certain threshold, also compared well to the full model results in R, if not the exact absolute values. So, for example, the absolute value of the (negative) coefficient for the 14-day average was much stronger in the STATA models. The same was true for the effects of same-day high temperatures, while the thresholds were quite similar in both models.

Characteristics of districts

District-wide characteristics were primarily calculated from 2020 and 2022 data from Statistik Austria: population density (both based on total area and of dwelling area), percentage of persons aged less than 15 years, percentage of persons aged 75 and older, percentage of persons with Austrian citizenship, percentage of persons born in Austria, percentage of persons without even primary education, percentage of persons with university degree, percentage of homeless people, and average income. In previous studies [3,4], other district characteristics were already examined. While in the COVID Study [3], the 23 districts of Vienna were not included, all districts of Austria (but for the four new Styrian districts) were included in the melanoma study [4].

From the melanoma study [4], the district characteristics "urban" district were used, where most districts were considered rural. With districts, where the rural part is an independent district and the large central town is a separate district, the central town district was considered "urban". The city of Vienna is also a separate federal country and consists of 23 districts. These 23 districts were considered "Vienna". Altitude above sea level was taken from the altitude of the main town of each district. Percentage of smokers (daily smokers and never smokers) was taken from Austrian Health Interview Surveys (ATHIS). We used 2014 data for the melanoma study and 2019 data for the COVID study. Because the ATHIS lacked power to estimate smoking prevalence per district sufficiently, only smoking prevalence per NUTS3 region was reported. It was assumed that every district within the same NUTS3 region had the same smoking prevalence.

From the COVID study [3], unemployment rate, percent of commuters (working outside of the district), percentage of voters of the main parties (for the parliamentary elections in 2019), percentage of persons working in agriculture, and number of tourist nights divided by population were used.

Because of the small number of data points (111 districts, or 88 districts without the districts of Vienna or 107 districts without the new Styrian districts), only univariate linear regressions were run between single characteristics of the districts and the coefficients of the temperature models. Only in rare occasions where an association was very likely confounded by another variable (tourism confounded by altitude), a model with two independent variables was run.

Results

As with the full model in R, also in the threshold model on the residuals, the 14-day average always had a negative and usually highly significant coefficient. The threshold was in most districts clearly above zero and the increase in residual death numbers above the threshold was nearly always significant. Only for 8 of the 111 districts, usually small districts, the coefficient did not reach significance (p<0.05).

The effects of the district characteristics on the coefficients are displayed in Table 1. For the sake of clearness, only those effects that at least displayed a tendency (p<0.1) are included. These are the result of univariate linear meta-regressions weighted by 2022 population number per district.

Characteristic	Average	Same-day	Temp²	Average	Threshold	Coeff. above
Density 1		**-1.29e-07**	**4.31e-09**		*0.0001321*	**6.33e-07**
Density 2		**-2.01e-07**	**6.72e-09**		*0.0000849*	**3.91e-07**
Age>75		**0.0003342**				**-0.0021118**
AT citizen	**-0.0000573**	**0.0001222**	**-2.58e-06**		*-0.0684247*	**-0.0004073**
AT born	**-0.0000483**	**0.0001114**	**-2.56e-06**		**-0.0667124**	**-0.0003726**
basic educ.						*0.0006108*
University		**-0.0001209**	**4.05e-06**			
Homeless		**-0.0064602**	*0.0001433*		**8.002941**	**0.0257093**
Net income			**9.24e-09**			
Vienna*		**-0.0025282**	**0.0000848**		**1.798318**	**0.0088554**
Altitude	*2.54e-06*	**5.98e-06**	**-3.69e-07**	*4.03e-06*	**-0.011658**	**-0.0000403**
Daily smoke	**0.0215685**	*-0.0113768*				
Working	**-0.0001799**	**0.00019**				
Unemployed	**0.000323**	**-0.0002982**				
Commuters	*-0.000034*	**0.0000311**				
Vote V						*-0.0295035*
Vote S		**-0.0089014**	**0.0004176**		**19.12868**	
Vote F	**-0.0236632**	**0.0151456**			**22.61899**	*0.0619391*
Vote N	*0.0235424*		**-0.0009042**		*-29.69967*	
Vote valid	**-0.0138134**		**0.0004456**		**21.15035**	
Agriculture	**-0.0200135**	**0.0143876**				
Tourism	**0.0000319**		**-8.93e-07**	**0.0000359**	**-0.0406086**	**-0.0001016**
Smoke daily						**0.180741**
Smoke never						**-0.1362675**

Table 1: Coefficients from the meta-regression on two models: (left) quadratic model (added in R to the base model), (right) Threshold model (in STATA on the residuals of the base model). **Bold**: p < 0.05, *italic*: p < 0.1
* Vienna compared to rural districts. Other urban districts did not differ significantly from rural districts.

Discussion

This analysis aimed for rather simple models returning a small set of coefficients that can be interpreted with ease. Therefore, distributed lag effects were replaced by an acute effect of same-day temperature and a sub-acute effect of the running 14-day average. Also, non-linear associations were approximated either by a quadratic polynomial or by a linear threshold model. These approximations worked pretty well, but still were approximations only. In addition, only daily temperature was included in the model. More complex indicators of perceived temperature that also included humidity did not provide a better model fit. Nevertheless, other meteorological variables would likely also have some impact on health and wellbeing.

In the quadratic model, the coefficient of temperature squared should be higher if the effect of (same-day) temperature is stronger at high temperatures. Therefore, the sign of the coefficient from the meta-regression should be the same as for the “Coefficient above the threshold” in the threshold model. A negative coefficient of the temperature in the quadratic model should shift the parabola to the right and therefore the sign of the coefficient should be the opposite to that of the threshold in the threshold model. On the whole, these expectations were met (see Table 1). The coefficients for the average temperature (14 days) should be similar in both models. Surprisingly, with the threshold model, the coefficients of the meta-regression only rarely reached significance. But when they did, they were in the same direction and in the same order of magnitude as for the quadratic model.

Certainly, not all characteristics of districts that displayed a significant effect on any of the coefficients of the temperature-mortality association had a causal impact. For example, tourism, where significant, lost its significance when combined with altitude, which remained significant in the more complex model. A more sophisticated meta-regression controlling for possible confounders was outside the scope and the power of this study. Although usually significant, effect estimates per district already came with substantial uncertainty and hence, a linear regression with only 111 data points at maximum lacked the necessary power. It was surprising that despite the poor power, many of the coefficients were highly significant. Even when significant, some factors would only serve as proxies for more complex (socio-economic) conditions. This is likely especially true for the district-wide voting behavior. Still, some factors seem plausible as determinants of vulnerability to temperature (extremes). This is likely true for population density, percentage of (very) old people, percentage of immigrants, percentage of persons with university degree, percentage of homeless people, percentage of unemployed, and altitude of the district. Why Vienna differed from the rest of the districts, is not completely clear. Why smoking prevalence in 2014 was significantly predictive in the districts minus the new Styrian districts, while smoking prevalence in 2019 in the districts minus those in Vienna, is not clear either. Nor is it clear how or if smoking prevalence is linked to temperature related vulnerability. All these issues would require a more complex multivariate analysis and even then, it is not clear if all these issues could be solved.

Extreme temperatures act as stressors to the human body thereby reducing physical and mental performance. In the case of pre-existing diseases or reduced stress-tolerance and adaptability, extreme temperatures can even lead to death. Daily deaths therefore serve as a proxy (or as the "tip of the iceberg") for more general effects on health and wellbeing. But not only extreme temperatures act as stressors. Also fast changes in temperature have such an effect. A more sophisticated model would therefore also include measures of temperature change, either between consecutive days, or within a day as the difference between maximum and minimum temperature.

Therefore, the actual effect estimates for temperatures might not be exactly accurate. Still, this would not invalidate the relative differences in the effect estimates between districts. These differences were the main objective of this study. This is also why the same models were run on every district, even though for some districts, mainly depending on population size, different and maybe even simpler models, would have worked quite as good or in some instances even slightly better.

Conclusion

This district-wise analysis and meta-regression does provide some valuable insight as to vulnerability factors (on the district level) that affect the impact of temperature (extremes) on mortality risk. Besides the altitude of the district (which likely is a proxy for the average temperature and thus for temperature adaptation), several socio-economic indicators have turned out to be relevant in that field.

References

1. Samoli, E., Peng, R., Ramsay, T., Pipikou, M., Touloumi, G., Dominici, F., Burnett, R., Cohen, A., Krewski, D., Samet, J., & Katsouyanni, K. (2008). Acute effects of ambient particulate matter on mortality in Europe and North America: results from the APHENA study. Environmental Health Perspectives, 116(11), 1480–1486. https://doi.org/10.1289/ehp.11345
2. Weitensfelder L, Moshammer H. Evidence of Adaptation to Increasing Temperatures Int. J. Environ. Res. Public Health 2019 17(1), 97 doi: 10.3390/ijerph17010097
3. Moshammer H, Poteser M, Weitensfelder L COVID-19: Regional Differences in Austria. Int. J. Environ. Res. Public Health 2022 19(3):1644. Doi: 10.3390/ijerph19031644
4. Haluza D, Simic S, Moshammer H. Temporal and Spatial Melanoma Trends in Austria: An Ecological Study. Int. J. Environ. Res. Public Health 2014 11(1):734-748.