Epidemiology and Infection

**Seasonality of urinary tract infections in the United Kingdom in different age groups: longitudinal analysis of The Health Improvement Network (THIN)**

A. ROSELLO\*1,2, K. B. POUWELS1,3, M. DOMENECH DE CELLÈS4, E. VAN KLEEF5,6, A. C. HAYWARD2,7, S. HOPKINS8,9, J. V. ROBOTHAM1, T. SMIESZEK1,3, L. OPATOWSKI4 and S. R. DEENY10.

# Supplementary material

# Appendix 1- Tables and Figures

## Tables

Table S1. Studies that analysed the seasonality of UTI

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Author** | **Year** | **Country** | **Community/ hospital** | **Organism** | **Sex** | **Age** | **Seasonality** | **Methods** |
| Stansfeld[1] | 1966 | England | Hospital | All | All | 0-12 | In cases >1 year age, 96 in winter and 58 in summer (significant at 1% level) | Unknown |
| Anderson[2] | 1983 | Canada | Community | All | Females | 15 or older | August peak | Edward's test for cyclic variation |
| Pead et al.[3] | 1985 | England | Community | All | Females | 15-25 | *S. saprophyticus* UTI peak in mid-September. Coliform (all Gram-negative bacilli other than Proteus spp. and Pseudomonas spp.) UTI peak in mid-March | Chi-squared test |
| Vorland et al.[4] | 1985 | Norway | Community | *E. coli* | All | All | Higher incidence from September to December (10.2 per 1,000 inhabitants) than from January to April (8.6 per 1000 inhabitants) or May to August (6.2 per 1,000 inhabitants), but non-significant. | Chi-squared test |
| Ferry et al.[5] | 1987 | Sweden | Community | All | All | All | No seasonality in *E. coli* UTI but August peak in *S. saprophyticus* UTIs | Comparing incidence |
| Stamm et al.[6] | 1991 | USA | Outpatient recurrence clinic | All | Females | All | Decrease in incidence November to February | Wilcoxon's signed-rank test |
| Kwok et al.[7] | 2006 | Netherlands | Community | All | All | 0-18 | Decrease in the summer months mainly in children 0-12 | Comparing incidence rates |
| Falagas et al.[8] | 2009 | Greece | Community (house call visits) | All | All | All | UTIs correlate with higher temperatures and decreased relative humidity | Spearman's rank correlation |
| Eriksson et al.[9] | 2013 | Sweden | Community and hospital | *E. coli*, *K. pneumoniae* and *P. mirabilis* aggregated and *S. saprophyticus* | Females | 15-29 | In GP samples, both peak in September, in hospital samples, both peak in August. Stronger seasonality in *S. saprophyticus*. | Chi-squared test |
| Rossignol et al.[10] | 2013 | France, Germany, USA, China Italy, Brazil and Australia | Community (online) | All | All | All | Increases of 8-19% in search trends for UTI-related terms in summer in France, Germany, USA, China and Italy, and peaks in the southern hemisphere austral summer in Brazil and Australia | Google trends analysis, Mann-Whitney test |
| Yolbas et al.[11] | 2013 | Turkey | Community | All | All | 1 month- 15 years | More UTIs in summer (53/150) than overall in winter (46/150), spring (35/150) or autumn (16/150) but difference in seasonality by sex | Comparing incidence |
| Melamed et al.[12] | 2014 | USA | Hospital | All | All | All | Summer peak | Lomb-Scargle periodograms in de-trended data |

Table S2. Akaike information criteria (AIC) for the models of UTI consultations in the UK including a seasonal component with no trend term, a linear trend term and a quadratic trend term, by sex and age group. \*In order to calculate the AIC, the dispersion parameter (theta) was fixed at the estimate derived for the most complex model (the seasonal model with a quadratic trend term). The trend term was given by and the seasonality term by , where was the month of the study (1 to 96).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sex | Age group | AIC of the regression model with no trend term\* | AIC of the regression model with a linear trend term ()\* | AIC of the regression model with a quadratic trend term |
| All UTIs | 14-17 | 1131.302 | 936.8906 | 925.5975 |
| All UTIs | 18-24 | 1375.33 | 1119.663 | 1083.569 |
| All UTIs | 25-45 | 1330.703 | 1249.725 | 1234.055 |
| All UTIs | 46-69 | 1281.198 | 1244.596 | 1231.499 |
| All UTIs | 70-84 | 1193.373 | 1181.468 | 1168.383 |
| All UTIs | 85+ | 1298.537 | 1054.343 | 1036.552 |
| Female UTIs | 14-17 | 1170.251 | 930.8574 | 917.3008 |
| Female UTIs | 18-24 | 1360.837 | 1117.974 | 1078.937 |
| Female UTIs | 25-45 | 1327.862 | 1237.177 | 1221.477 |
| Female UTIs | 46-69 | 1262.241 | 1218.761 | 1205.818 |
| Female UTIs | 70-84 | 1143.959 | 1134.796 | 1123.924 |
| Female UTIs | 85+ | 1241.139 | 1015.206 | 1001.247 |
| Male UTIs | 14-17 | 582.4998 | 568.1808 | 570.5024 |
| Male UTIs | 18-24 | 678.9267 | 624.9811 | 627.5591 |
| Male UTIs | 25-45 | 918.8692 | 802.7613 | 797.7937 |
| Male UTIs | 46-69 | 930.2015 | 922.5998 | 920.3444 |
| Male UTIs | 70-84 | 929.6153 | 931.4508 | 920.3366 |
| Male UTIs | 85+ | 861.6667 | 830.0555 | 819.1208 |

Table S3. Coefficients and 95% confidence intervals (CI) of the models of UTI consultations in the UK by sex and age group. The trend term was given by and the seasonality term by , where was the month of the study (1 to 96). The confidence intervals were calculated using the confint function in R.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sex | ages | (95% CI) | (95% CI) | (95% CI) | (95% CI) | (95% CI) |
| All UTIs | 14-17 | -6.08 (-6.15, -0.00268) | 0.000415 (-6.15, -0.00268) | -6.2e-05 (-6.15, -0.00268) | 0.0994 (-6.15, -0.00268) | -0.154 (-6.15, -0.00268) |
| All UTIs | 18-24 | -5.46 (-5.49, 0.00013) | 0.00193 (-5.49, 0.00013) | -5.85e-05 (-5.49, 0.00013) | 0.0241 (-5.49, 0.00013) | -0.1 (-5.49, 0.00013) |
| All UTIs | 25-45 | -5.93 (-5.97, 4.64e-05) | 0.00173 (-5.97, 4.64e-05) | -3.83e-05 (-5.97, 4.64e-05) | 0.0205 (-5.97, 4.64e-05) | -0.0818 (-5.97, 4.64e-05) |
| All UTIs | 46-69 | -5.76 (-5.79, 0.000349) | 0.00172 (-5.79, 0.000349) | -2.9e-05 (-5.79, 0.000349) | 0.0102 (-5.79, 0.000349) | -0.0763 (-5.79, 0.000349) |
| All UTIs | 70-84 | -5.1 (-5.13, 0.000806) | 0.00222 (-5.13, 0.000806) | -2.98e-05 (-5.13, 0.000806) | -0.00302 (-5.13, 0.000806) | -0.0524 (-5.13, 0.000806) |
| All UTIs | 85+ | -5.1 (-5.13, 0.000806) | 0.00222 (-5.13, 0.000806) | -2.98e-05 (-5.13, 0.000806) | -0.00302 (-5.13, 0.000806) | -0.0524 (-5.13, 0.000806) |
| Female UTIs | 14-17 | -5.4 (-5.47, -0.0029) | 0.000229 (-5.47, -0.0029) | -6.72e-05 (-5.47, -0.0029) | 0.101 (-5.47, -0.0029) | -0.167 (-5.47, -0.0029) |
| Female UTIs | 18-24 | -4.84 (-4.87, 0.000431) | 0.00225 (-4.87, 0.000431) | -6.12e-05 (-4.87, 0.000431) | 0.0236 (-4.87, 0.000431) | -0.101 (-4.87, 0.000431) |
| Female UTIs | 25-45 | -5.36 (-5.4, -6.22e-05) | 0.00164 (-5.4, -6.22e-05) | -3.87e-05 (-5.4, -6.22e-05) | 0.0201 (-5.4, -6.22e-05) | -0.0852 (-5.4, -6.22e-05) |
| Female UTIs | 46-69 | -5.27 (-5.3, 0.000253) | 0.00168 (-5.3, 0.000253) | -2.98e-05 (-5.3, 0.000253) | 0.0114 (-5.3, 0.000253) | -0.0812 (-5.3, 0.000253) |
| Female UTIs | 70-84 | -4.81 (-4.84, 0.000701) | 0.00216 (-4.84, 0.000701) | -2.85e-05 (-4.84, 0.000701) | -0.00206 (-4.84, 0.000701) | -0.0564 (-4.84, 0.000701) |
| Female UTIs | 85+ | -4.81 (-4.84, 0.000701) | 0.00216 (-4.84, 0.000701) | -2.85e-05 (-4.84, 0.000701) | -0.00206 (-4.84, 0.000701) | -0.0564 (-4.84, 0.000701) |
| Male UTIs | 14-17 | -8.43 (-8.61, -0.00787) | 0.00113 (-8.61, -0.00787) | -6.03e-05 (-8.61, -0.00787) | 0.0924 (-8.61, -0.00787) | 0.0412 (-8.61, -0.00787) |
| Male UTIs | 18-24 | -8.08 (-8.19, -0.00684) | -0.00185 (-8.19, -0.00684) | -3.08e-05 (-8.19, -0.00684) | 0.0383 (-8.19, -0.00684) | -0.0673 (-8.19, -0.00684) |
| Male UTIs | 25-45 | -7.8 (-7.85, -0.00216) | 0.000213 (-7.85, -0.00216) | -3.67e-05 (-7.85, -0.00216) | 0.028 (-7.85, -0.00216) | -0.0499 (-7.85, -0.00216) |
| Male UTIs | 46-69 | -6.97 (-7.01, -0.000274) | 0.0015 (-7.01, -0.000274) | -2.24e-05 (-7.01, -0.000274) | 0.00462 (-7.01, -0.000274) | -0.0508 (-7.01, -0.000274) |
| Male UTIs | 70-84 | -5.84 (-5.88, 0.00151) | 0.00323 (-5.88, 0.00151) | -3.39e-05 (-5.88, 0.00151) | -0.0071 (-5.88, 0.00151) | -0.036 (-5.88, 0.00151) |
| Male UTIs | 85+ | -5.84 (-5.88, 0.00151) | 0.00323 (-5.88, 0.00151) | -3.39e-05 (-5.88, 0.00151) | -0.0071 (-5.88, 0.00151) | -0.036 (-5.88, 0.00151) |

Table S4. Month of the year with the highest number of UTI consultations or trimethoprim and nitrofurantoin prescriptions by age group. In brackets, the rate of UTI consultations or trimethoprim and nitrofurantoin prescriptions per 100,000 person years for that month.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Date | Trimethoprim and nitrofurantoin prescriptions in the UK <85 | Trimethoprim and nitrofurantoin prescriptions in the UK 85+ | UTI consultations in the UK <85 | UTI consultations in the UK 85+ | UTI consultations in England <85 | UTI consultations in England 85+ |
| 2008 | Oct (515.08) | Jan (1886.24) | Sep (352.52) | Jan (922.37) | Sep (355.56) | Jan (826.19) |
| 2009 | Sep (535.13) | Oct (1622.4) | Sep (350.65) | Jan (881.78) | Sep (352.6) | Jan (817.95) |
| 2010 | Sep (554.32) | Mar (1612.09) | Sep (344.01) | Jul (838.79) | Sep (339.61) | Jul (764.2) |
| 2011 | Nov (587.15) | Mar (1653.98) | Sep (342.48) | Aug (814) | Sep (345.05) | Nov (742.15) |
| 2012 | Oct (621.67) | Jan (1799.42) | Oct (349.52) | Jan (817.35) | Oct (345.29) | Jan (740.61) |
| 2013 | Oct (622.2) | Jan (1768.77) | Oct (349.18) | Jan (809.73) | Oct (340.93) | Oct (741.06) |
| 2014 | Oct (597.12) | Jan (1687.91) | Jan (325.01) | Jan (756.34) | Oct (317.27) | Jan (645.86) |
| 2015 | Sep (556.59) | Jan (1517.94) | Sep (293.28) | Jan (692.45) | Sep (282.98) | Jan (611.35) |

Table S5. Akaike information criteria (AIC) for models of the scaled UTI consultations in the UK which included a seasonal component and models that did not by age group. \*In order to calculate the AIC for the non-seasonal model, the dispersion parameter (theta) was fixed at the estimate derived for the seasonal model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age group | AIC seasonal model | AIC non-seasonal model\* | % deviance explained by the seasonal model | % deviance explained by the non-seasonal model |
| 14-17 | 960.78 | 1041.58 | 72.21 | 46.84 |
| 18-24 | 1096.01 | 1196.45 | 77.92 | 53.5 |
| 25-45 | 1242.65 | 1293.55 | 54.06 | 26.85 |
| 46-69 | 1254.36 | 1309.21 | 48.88 | 16.55 |
| 70-84 | 1196.56 | 1217.86 | 28.74 | 8.51 |
| 85+ | 1108.25 | 1103.92 | 44.43 | 43.47 |

## Figures

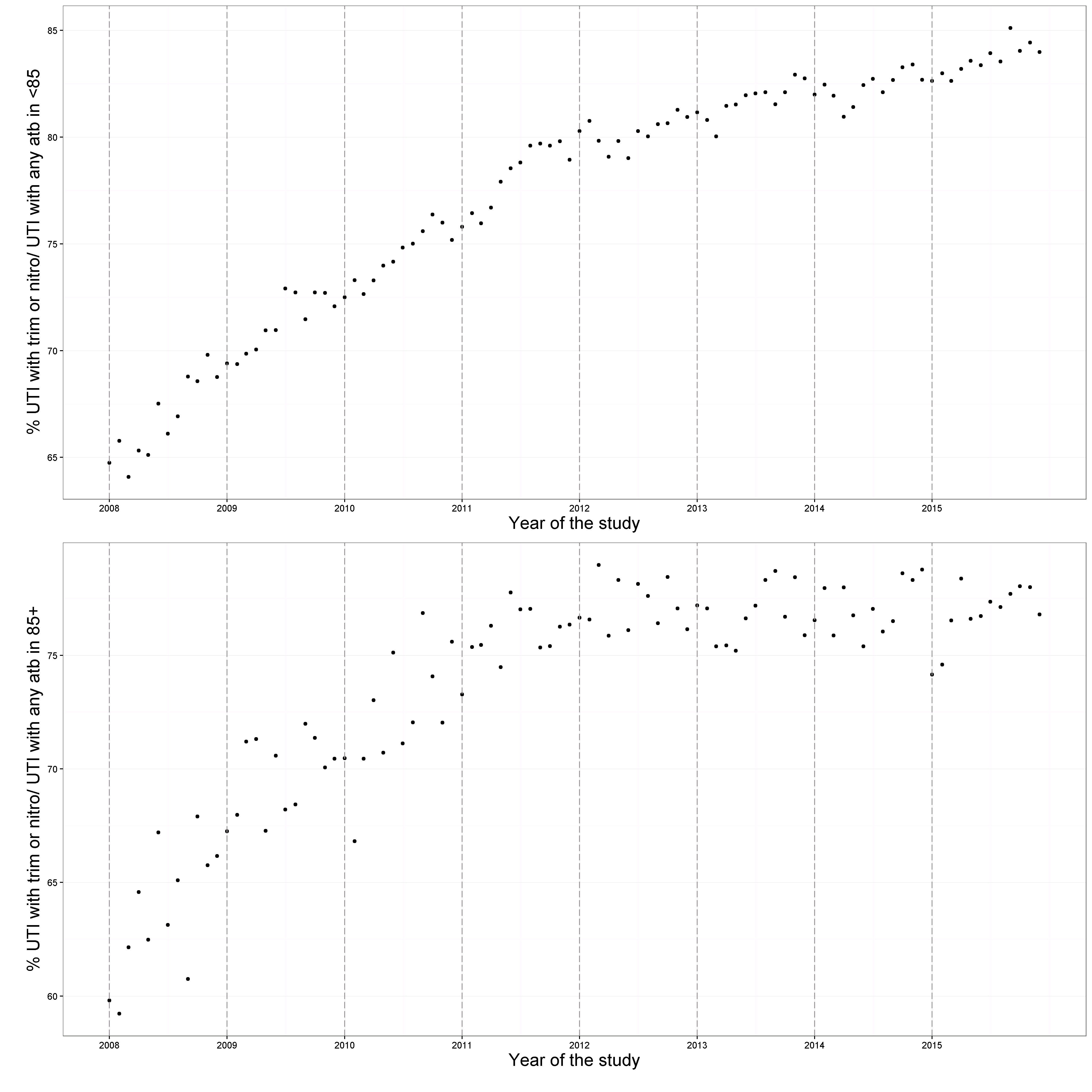


Figure S1. Percentage of monthly UTI consultation coded with any antibiotic prescription on the same day for which that antibiotic was trimethoprim or nitrofurantoin, by age group.

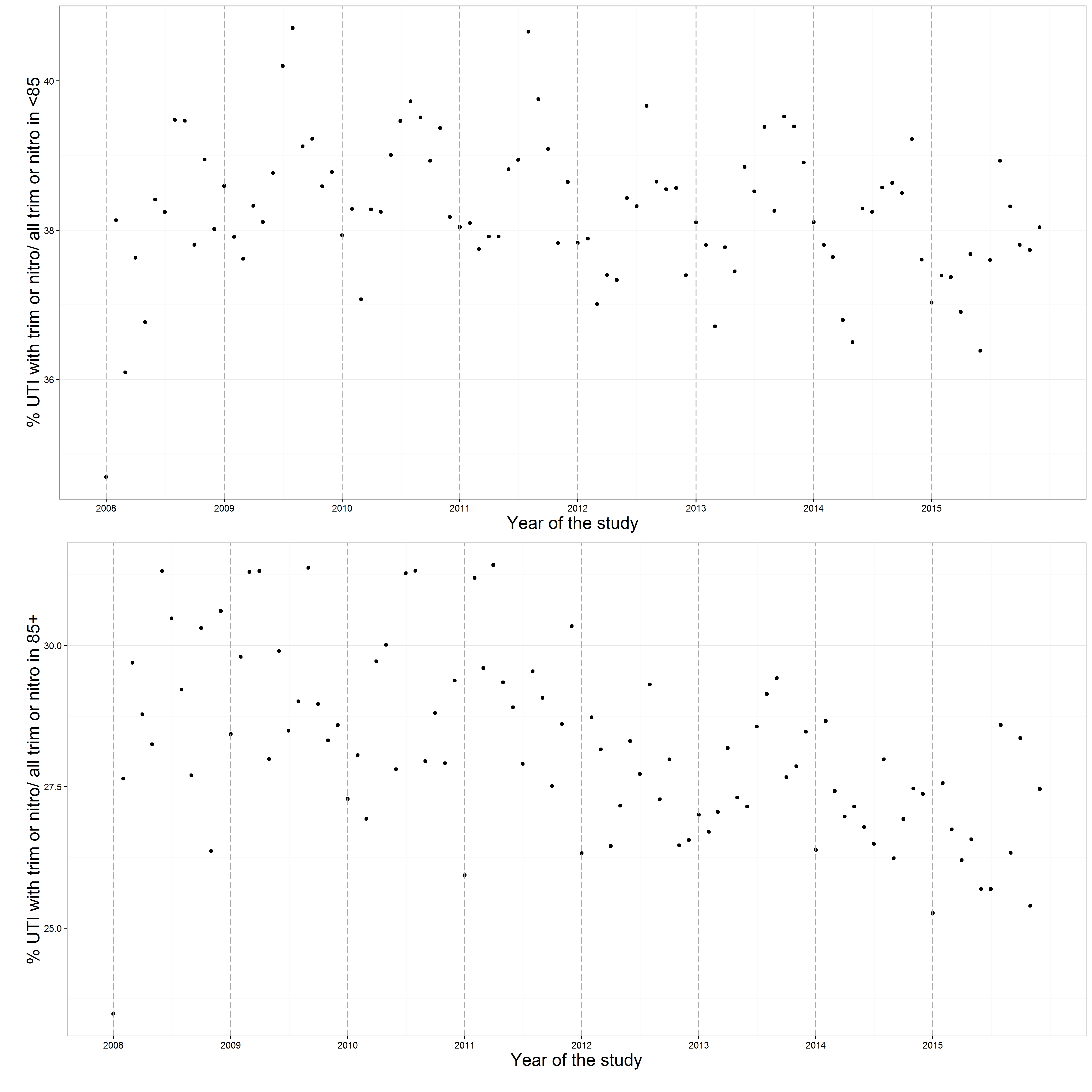


Figure S2. Percentage of monthly trimethoprim and nitrofurantoin prescriptions that had a UTI consultation coded on the same day for those aged under 85 and 85 or over. Nitrofurantoin and trimethoprim are almost exclusively prescribed for UTI, therefore this can be interpreted as a proxy for coding of UTI consultation.

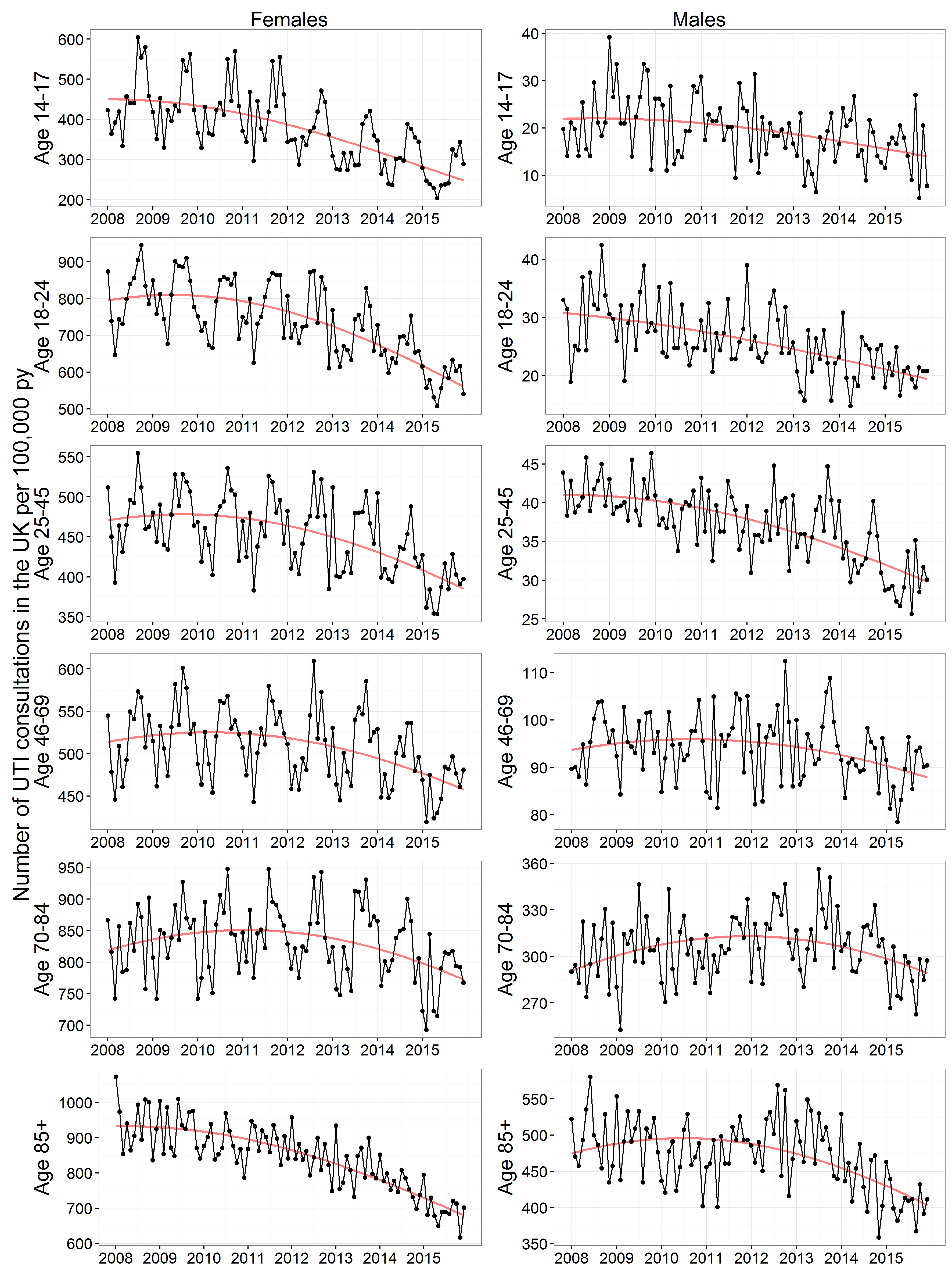


Figure S3. Monthly UTI consultations coded by GPs per 100,000 person years in the UK by age group and sex. The central red lines represent the fitted trend predictions from the seasonal regression model. This was a negative binomial polynomial regression model of degree two with the number of patients registered at each of the GP practices on the 1st of July (mid-year) each year of the study as offset. The UTI consultations were de-duplicated to one per 30-day period. The y axes differ between panels.

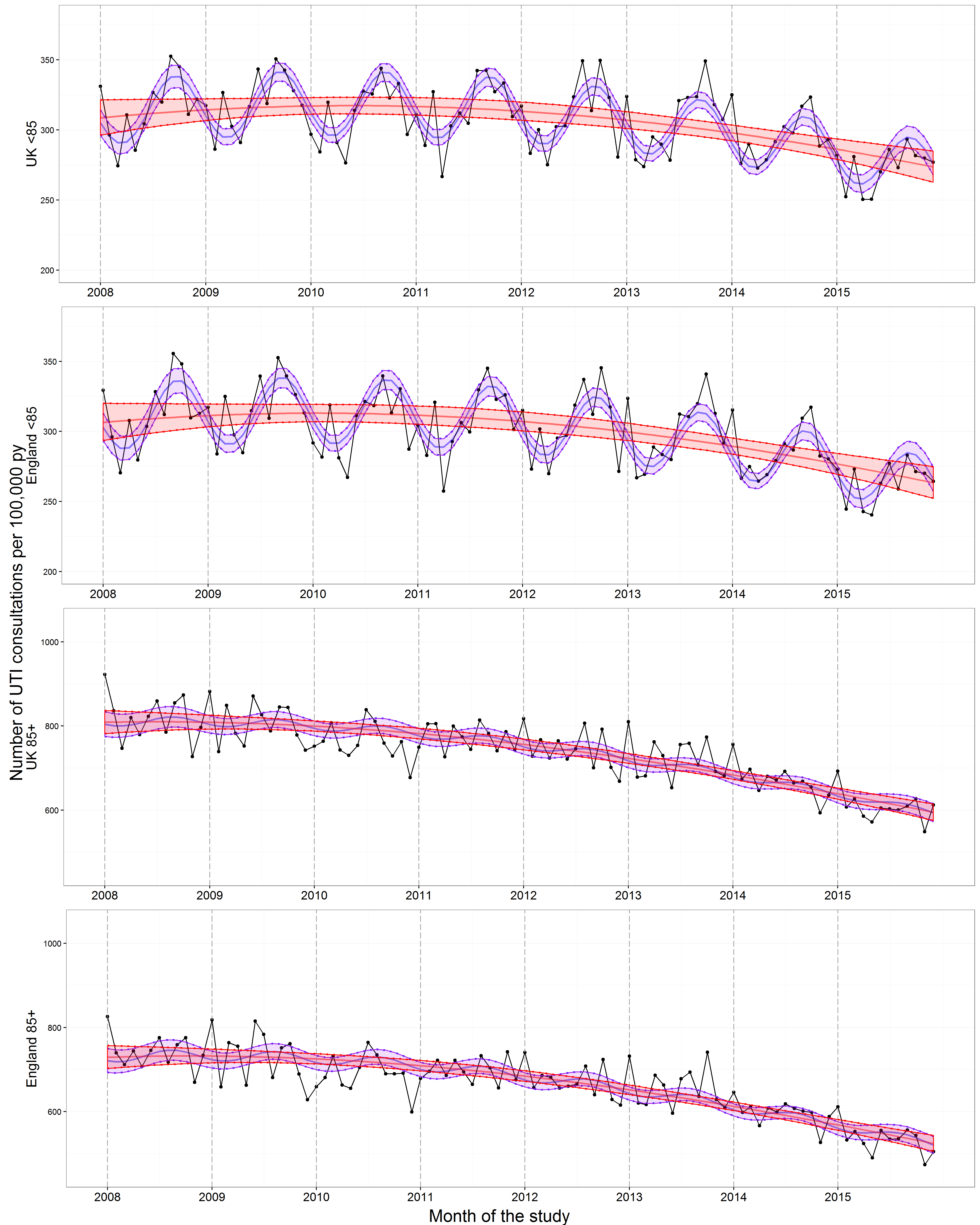


Figure S4. Monthly UTI consultations coded by GPs per 100,000 person years in England and in the UK by age group. The central red lines represent the fitted predictions of the negative binomial polynomial regression model of degree two with the number of patients registered at each of the GP practices on the 1st of July (mid-year) each year of the study as offset. The central blue lines represent the fitted predictions of the same model but with a seasonal component included. The shaded areas represent the 95% confidence intervals for their respective models. These were calculated using the standard errors from the predict function, which calculates the confidence intervals around the mean. The UTI consultations were de-duplicated to one per 30-day period. The y axes differ between panels.

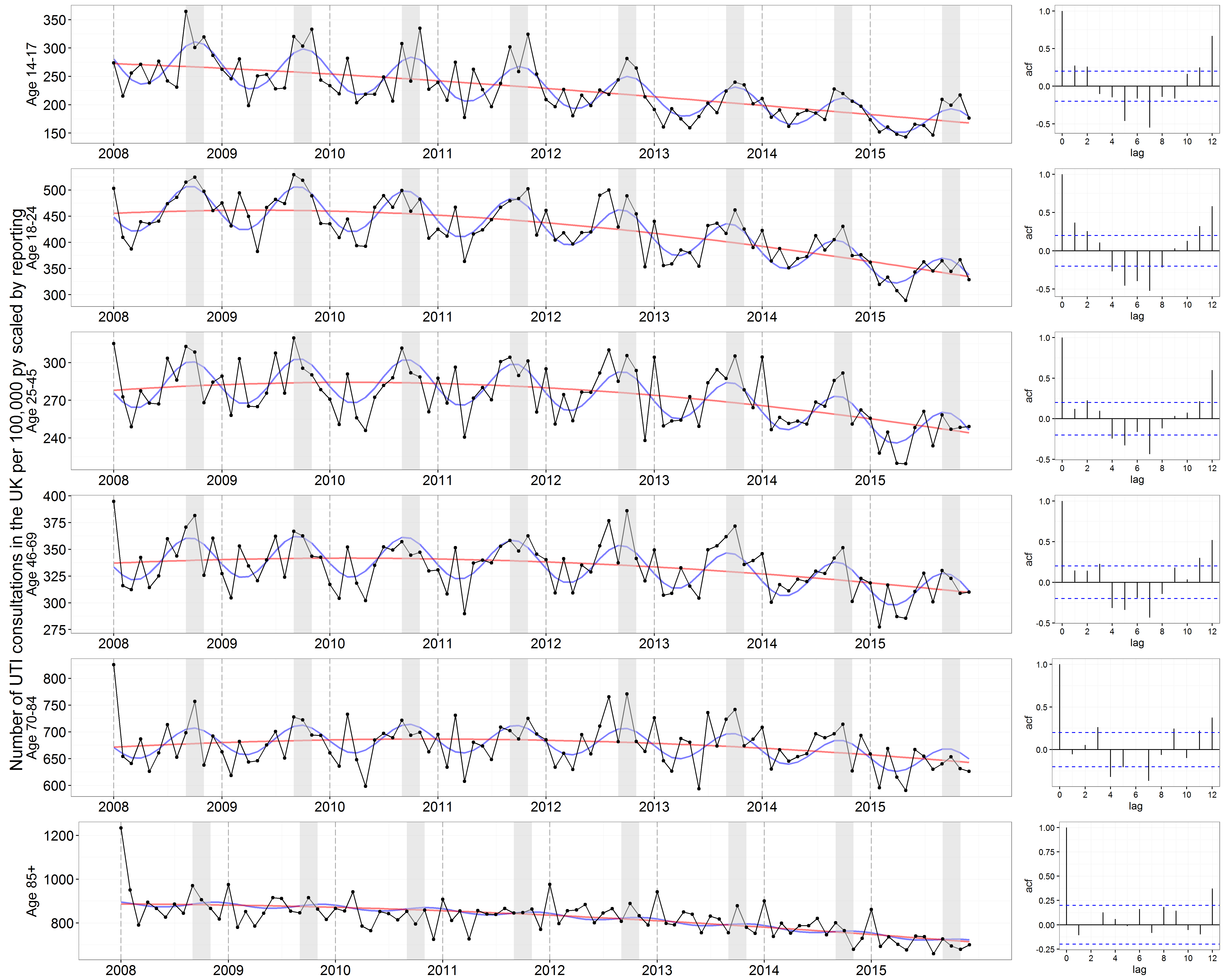


Figure S5. Scaled monthly UTI consultations coded per 100,000 person years in the UK by age group. The UTI consultations were de-duplicated to one per 30-day period. The red lines represent the fitted predictions of the negative binomial polynomial regression model of degree two with the number of patients registered at each of the GP practices on the 1st of July (mid-year) each year of the study as offset. The blue lines represent the fitted predictions of the same model but with a seasonal component included. No confidence intervals are presented as these were scaled predictions. The monthly UTI consultations were scaled for each age group by dividing by a scaling factor. This scaling factor was the percentage of UTIs coded in each month (the percentage of monthly trimethoprim and nitrofurantoin prescriptions that had a UTI consultation coded on the same day) divided by the maximum percentage coded over the study period for that age group. The right panels show the correlograms for the residuals of the regression models without seasonality at lags of 0-12 months for each age group. The y axes differ between panels.

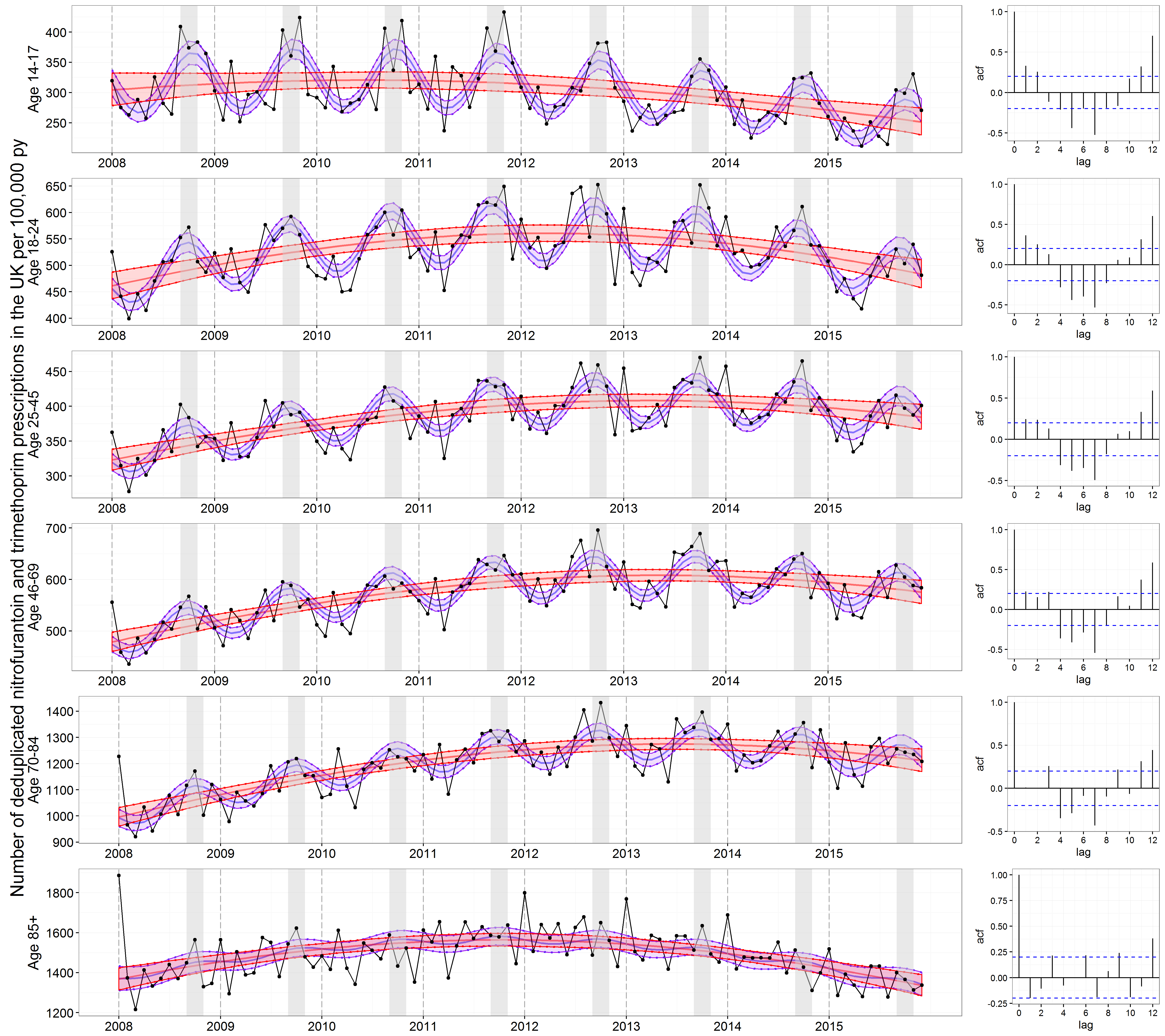


Figure S6. Monthly nitrofurantoin and trimethoprim prescriptions administered by GPs per 100,000 person years in the UK by age group. The nitrofurantoin and trimethoprim prescriptions were de-duplicated to one per 30-day period. The central red lines represent the fitted predictions of the negative binomial polynomial regression model of degree two with the number of patients registered at each of the GP practices on the 1st of July (mid-year) each year of the study as offset. The central blue lines represent the fitted predictions of the same model but with a seasonal component included. The shaded areas represent the 95% confidence intervals for their respective models. These were calculated using the standard errors from the predict function, which calculates the confidence intervals around the mean. The right panels show the correlograms for the residuals of the regression models without seasonality at lags of 0-12 months for each age group. The y axes differ between panels.

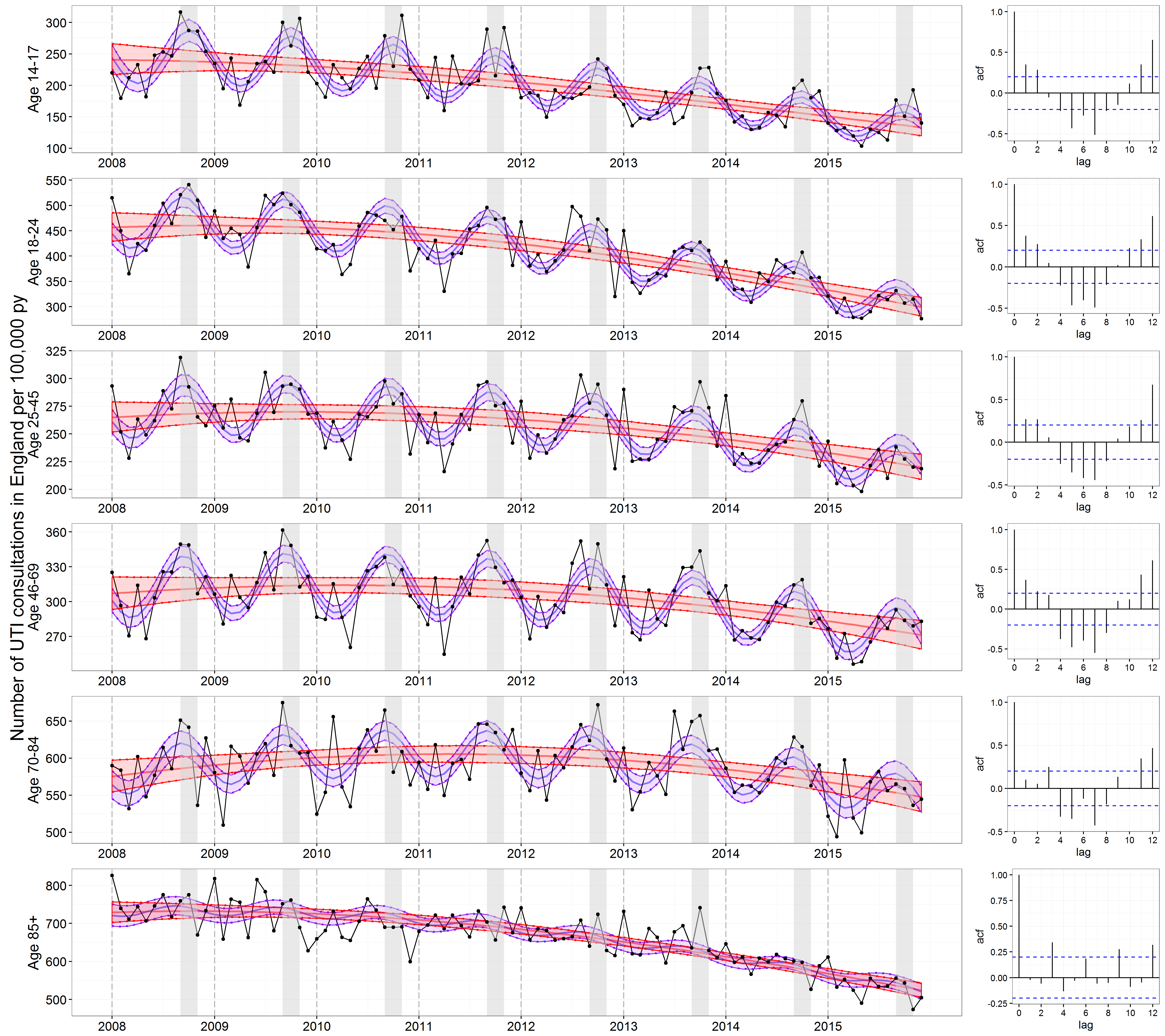


Figure S7. Monthly UTI consultations coded per 100,000 person years in England by age group. The UTI consultations were de-duplicated to one per 30-day period. The central red lines represent the fitted predictions of the negative binomial polynomial regression model of degree two with the number of patients registered at each of the GP practices on the 1st of July (mid-year) each year of the study as offset. The central blue lines represent the fitted predictions of the same model but with a seasonal component included. The shaded areas represent the 95% confidence intervals for their respective models. These were calculated using the standard errors from the predict function, which calculates the confidence intervals around the mean. The right panels show the correlograms for the residuals of the regression models without seasonality at lags of 0-12 months for each age group. The y axes differ between panels.

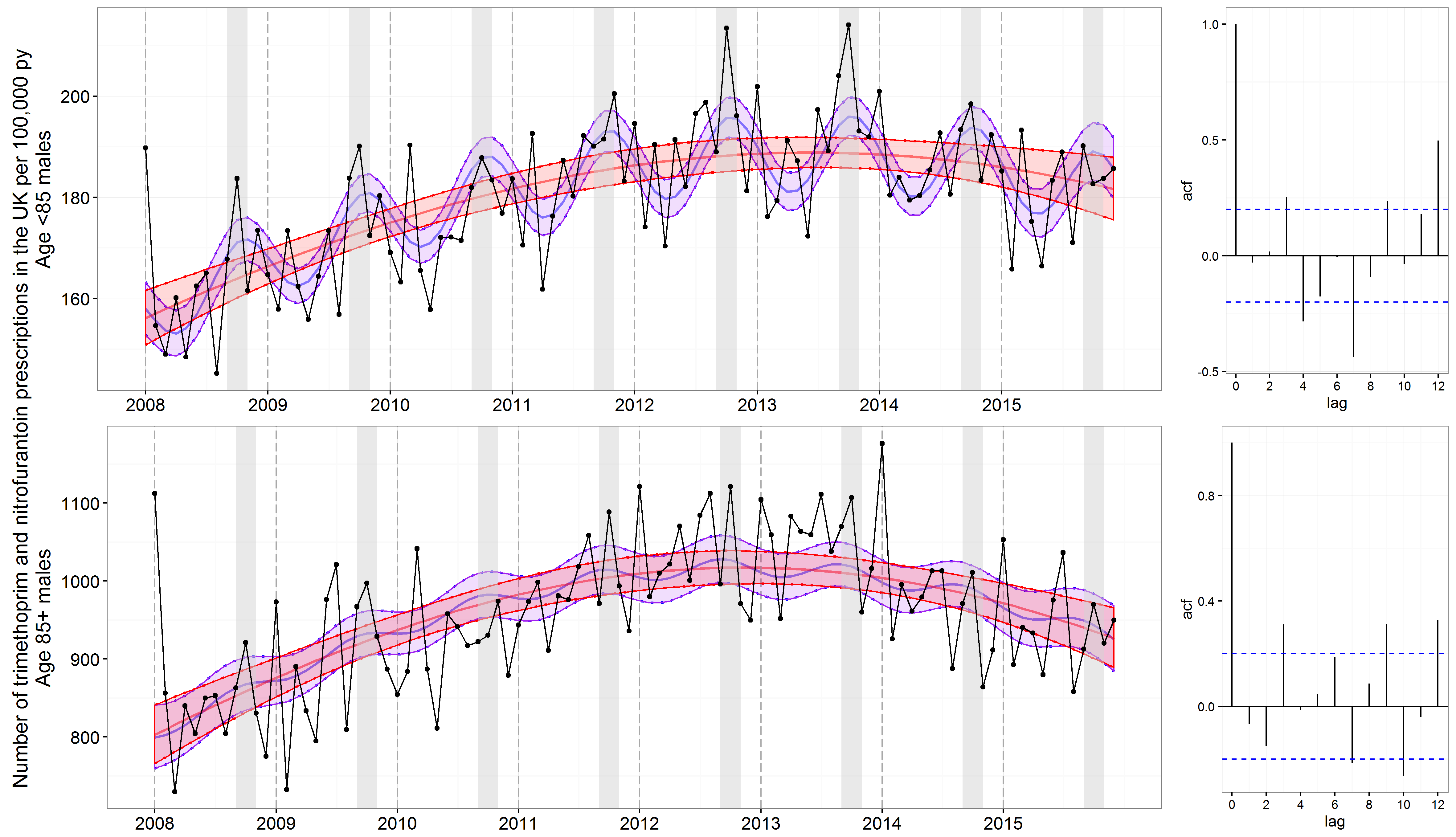


Figure S8. Monthly nitrofurantoin and trimethoprim prescriptions administered by GPs to males per 100,000 person years in the UK by age group. The nitrofurantoin and trimethoprim prescriptions were de-duplicated to one per 30-day period. The central red lines represent the fitted predictions of the negative binomial polynomial regression model of degree two with the number of patients registered at each of the GP practices on the 1st of July (mid-year) each year of the study as offset. The central blue lines represent the fitted predictions of the same model but with a seasonal component included. The shaded areas represent the 95% confidence intervals for their respective models. These were calculated using the standard errors from the predict function, which calculates the confidence intervals around the mean. The right panels show the correlograms for the residuals of the regression models without seasonality at lags of 0-12 months for each age group. The AIC of the model in those aged under 85 decreases (from 1262.5 to 1237.4) by including seasonality in the elderly, but remains similar in those aged 85+ (928.0 in the model without seasonality and 932.9 in the model with seasonality). The y axes differ between panels.

# Appendix 2

*Dataset*

THIN is a validated database of primary care consultation data covering over 3.7 million active patients which are demographically representative of the UK.[13–15] The dataset contains individual pseudonomysed patient ID, prescription details, consultation date and time, reason for consultation (recorded through diagnostic code), patient registration details and patient clinical and demographic information.

In order to obtain the monthly rate of de-duplicated UTI consultations, nitrofurantoin prescriptions and trimethoprim prescriptions by age and sex, for 2008-2015, we extracted UTI diagnostic codes (listed in Appendix 3), Patient ID, trimethoprim and nitrofurantoin prescriptions (derived from the prescribing information in THIN), country, date of UTI consultation/prescription, date of registration at GP, date of de-registration at GP, patient age, patient sex, and patients registered on the 1st of July (mid-year) each year for 2008-2015 at each of the GP practices present in THIN during the whole duration of the study (for this, practice ID was required). UTI consultations and nitrofurantoin and trimethoprim prescriptions from UK practices meeting acceptable standard for research (as suggested by the THIN Data Guide for Researchers) were de-duplicated to one per patient per 30-day period in order to approximate episodes of infection (one nitrofurantoin or trimethoprim prescription during the 30-day period) and subsequently aggregated by age group, sex and moth of the study. The denominator population was the number of patients (of the corresponding age group and sex) registered at each of the GP practices on the 1st of July (mid-year) each year of the study.

*Reasoning for analysing both GP consultations and antibiotic prescriptions*

Consultation codes in THIN are known to be poorly recorded[16,17]. However, all prescriptions made by GP practices reporting to THIN are automatically included in the database and do not suffer from this reporting bias. Hence, the analysis of UTI consultations was repeated for trimethoprim and nitrofurantoin prescriptions. Both trimethoprim and nitrofurantoin are almost exclusively prescribed for UTIs and account for the majority of antibiotics used for UTIs in primary care.

Only the rate of UTI consultations (and not antibiotic prescriptions) were used to assess the trend in UTIs over time, because nitrofurantoin and trimethoprim prescriptions for UTI as a proportion of all antibiotic prescriptions increased over the study period. (Figure S1, Appendix 1). Although coding for UTI consultations by GPs was poor, it remained stable over the study period (the percentage of trimethoprim and nitrofurantoin prescriptions that had a UTI consultation coded on the same day fluctuated between 35-41% during the study period), enabling the study of trend over time (Figure S2, Appendix 1).

# Appendix 3

Read codes for UTI

'K190300', 'K190400', '1AG..00', 'K190311', 'K190.11', '14D7.00', 'L166z11', 'L166800', 'K190.00', 'K190500', 'K190z00', 'K190000', 'K190100', 'K190200', 'K190600', 'K190X00', 'Q40y100', '1J4..00', '46U3.00', '4617.00', 'K190011', 'L166600',

'K15..00', 'K150.00', 'K15z.00', 'K152000', 'K154.00', 'K154000', 'K154300', 'K154400', 'K154600', 'K154800', 'K154z00', 'K15y.00', 'K15y200', 'K15y300', 'K15yz00', 'A32y300', 'K153.11', 'K151.00', 'K152y00', 'K152.00', 'K152z00', 'K155.00', '14D4.00',

'L166.11', 'L166500', 'K101.00', 'K101000', 'K101100', 'K101200', 'K101300', 'K101400', 'K101500', 'K101z00', 'K106.00', 'K100.00', 'K100000', 'K100100', 'K100200', 'K100300', 'K100400', 'K100500', 'K100600', 'K100z00', 'K10y000', 'A160200', 'K104.00',

'K10..00', 'K102.00', 'K102000', 'K102100', 'K102200', 'K102z00', 'K103.00', 'K105.00', 'K10y.00', 'K10y000', 'K10y100', 'K10y200', 'K10y300', 'K10y400', 'K10yz00','K10z.00', 'K10..11'.

# Appendix 4

The following negative binomial model was fit to the rates of consultations and prescriptions:

Where, was the month of the study; was the number of consultations and prescriptions at month ; was the intercept; was a quadratic term defined as , used to account for the decreasing trend observed in the rates; was a seasonality term defined as ; and was an offset used to model the rates of consultations and prescriptions instead of the counts.

Negative binomial models were best suited to model the rates of consultations and prescriptions due to the overdispersion in the data.

# References

1. **Stansfeld JM**. Clinical observations relating to incidence and aetiology of urinary-tract infections in children. *British medical journal* 1966; **1**: 631–5.

2. **Anderson JE**. Seasonality of symptomatic bacterial urinary infections in women. *Journal of Epidemiology and Community Health* 1983; **37**: 286–290.

3. **Pead L, Maskell R, Morris J**. Staphylococcus saprophyticus as a urinary pathogen: a six year prospective survey. *British medical journal (Clinical research ed.)* 1985; **291**: 1157–9.

4. **Vorland LH, Carlson K, Aalen O**. Antibiotic resistance and small R plasmids among Escherichia coli isolates from outpatient urinary tract infections in northern Norway. *Antimicrobial agents and chemotherapy* 1985; **27**: 107–13.

5. **Ferry S, Burman LG, Mattsson B**. Urinary tract infection in primary health care in northern Sweden. I. Epidemiology. *Scandinavian journal of primary health care* 1987; **5**: 123–8.

6. **Stamm WE, *et al.*** Natural history of recurrent urinary tract infections in women. *Reviews of infectious diseases* 1991; **13**: 77–84.

7. **Kwok W-Y, *et al.*** Incidence rates and management of urinary tract infections among children in Dutch general practice: results from a nation-wide registration study. *BMC pediatrics* 2006; **6**: 10.

8. **Falagas ME, *et al.*** Effect of meteorological variables on the incidence of lower urinary tract infections. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 2009; **28**: 709–12.

9. **Eriksson A, Giske CG, Ternhag A**. The relative importance of *Staphylococcus saprophyticus* as a urinary tract pathogen: distribution of bacteria among urinary samples analysed during 1 year at a major Swedish laboratory. *APMIS* 2013; **121**: 72–78.

10. **Rossignol L, *et al.*** A method to assess seasonality of urinary tract infections based on medication sales and google trends. *PloS one* Public Library of Science, 2013; **8**: e76020.

11. **Yolbaş I, *et al.*** Community-acquired urinary tract infections in children: pathogens, antibiotic susceptibility and seasonal changes. *European review for medical and pharmacological sciences* 2013; **17**: 971–6.

12. **Melamed RD, Khiabanian H, Rabadan R**. Data-driven discovery of seasonally linked diseases from an Electronic Health Records system. *BMC bioinformatics* 2014; **15 Suppl 6**: S3.

13. **Blak BT, *et al.*** Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Informatics in primary care* 2011; **19**: 251–5.

14. *IMS Health*. (http://www.epic-uk.org/our-data/our-data.shtml). Accessed 21 April 2017.

15. **Bourke A, Dattani H, Robinson M**. Feasibility study and methodology to create a quality-evaluated database of primary care data. *Informatics in primary care* 2004; **12**: 171–7.

16. **James GDR, *et al.*** Use of long-term antibiotic treatment in COPD patients in the UK: a retrospective cohort study. *Primary Care Respiratory Journal* Nature Publishing Group, 2013; **22**: 271–277.

17. **Petersen I, *et al.*** Oral antibiotic prescribing during pregnancy in primary care: UK population-based study. *Journal of Antimicrobial Chemotherapy* Oxford University Press, 2010; **65**: 2238–2246.