Variation in Primary Care Prescribing in Wales due to COVID-19

First Author[1] & Second Author[1,2]

[1] Centre for Health Economics & Medicines Evaluation

Bangor University

LL55 2PZ

[2] Institute for the Psychology of Elite Performance

Bangor University

LL55 2PZ

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to First Author, Postal address. E-mail: my@email.com

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words "**here we show**" or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

Keywords: keywords

Word count: X

Variation in Primary Care Prescribing in Wales due to COVID-19

Methods

Data

- Sources of data
  - GP
  - Prescribing
  - QOF
  - WIMD
- Preparing

Identifying changes in prescribing due to COVID-19

Identifying the effects of COVID-19 is not simple, as its effects are unprecedented and its global impact prevents us from adopting a traditional treatment and control group type design. Instead, we have used time series analysis methods to create counterfactual forecasts

To understand the changes in levels of prescribing due to COVID-19, it was first necessary for us to forecast the levels of prescribing had COVID-19 not affected Wales as a baseline. To do so, we used historic data (January 2015 to December 2020) to identify from a range of time series models, one that was good fit to the data. This was done using cross-validation to reduce the likelihood of choosing a model that was overfitted to the data.

- Given the presence of seasonal variation and existing trends in levels of prescribing, we used various time series forecasting methods to account for existing trends and seasonal variation when creating the forecasts.

- For each analysis, we created a forecast that aimed to answer the question, "What would the levels of prescribing have been if COVID-19 had not affected Wales?" We then used these forecasts to estimate the changes in levels of prescribing due to COVID-19.
  - We viewed this, somewhat complex, approach as superior to more simple approaches (e.g., carrying last years levels forward, or even using the mean of the last three years).
- We did not assume that the existing processes would be the same for all drugs or all GP practices, therefore, we investigated the fit of several different time series models to the pre-COVID data.
  - Using Jan 2015 to Feb 2020 data, we fitted several different models and assessed their accuracy using a cross-validated process to reduce the likelihood of overfitting models to the data.
    - Started with 36 months of data, used 6-month horizon as this was the horizon we would be using for the forecasts, 3-month step (to reduce computation time) **Do we want to re-run this with 1-month?**
    - Only interested in one forecast horizon: six-months
    -
  - Having chosen the "best model" for each practice based on the pre-COVID data, we then used this to forecast the level of prescribing for each practice
- We used time series modelling to forecast Welsh primary care prescribing levels had .
  - We used the `fable` (O'Hara-Wild et al., 2021a) and the `fable.prophet` (O'Hara-Wild, 2020) packages to conduct the time series analyses.
  - Time series linear model

- Decomposition model
- Seasonal naïve (with and without drift)
- Autoregressive integrated moving average (ARIMA; Box et al., 2015)
- Holt-Winters Additive Model (Chatfield, 1978)
- Prophet (Taylor & Letham, 2018)
- Combination models (cf. Thomson et al., 2019)
- Prescribing quantities log transformed and forecasts use median values to reduce bias that back transformation would introduce when using the mean
- Model selection
  - RMSE then MAE

Identifying different prescribing behaviours - LPA

- What is LPA
  - Differences from clustering methods (e.g., k-means and hierarchical)

When conducting LPA, several models are specified and then evaluated. Model selection, including class enumeration, should not be based solely on statistical criteria, but also the statistical adequacy and substantive meaning of the solutions (**???**; Marsh et al., 2009). Indeed, relying solely on statistical criteria in large samples may lead to the inability to identify an "optimal solution" as model fit increases as the number of classes increase (**???**).

In this study we inspected the Bayesian Information Criterion (BIC; Schwartz, 1978) and the bootstrapped likelihood ratio test (BLRT; McLachlan & Peel, 2000) during the class enumeration process as the results of Monte-Carlo simulation study (Nylund et al., 2007) showed them outperform other information criteria and likelihood-based tests.

- Entropy
- Posterior probabilities

Reproducability and code

We used R (Version 4.0.2; R Core Team, 2020) and the R-packages *dplyr* (Version 1.0.4; Wickham, François, et al., 2021), *fable* (Version 0.3.0; O'Hara-Wild et al., 2021a, 2021b; O'Hara-Wild, 2020), *fable.prophet* (Version 0.1.0; O'Hara-Wild, 2020), *fabletools* (Version 0.3.0; O'Hara-Wild et al., 2021b), *feasts* (Version 0.1.7; O'Hara-Wild et al., 2021c), *forcats* (Version 0.5.1; Wickham, 2021), *furrr* (Version 0.2.2; Vaughan & Dancho, 2021), *future* (Version 1.21.0; Bengtsson, 2020), *ggplot2* (Version 3.3.3; Wickham, 2016), *lubridate* (Version 1.7.10; Grolemund & Wickham, 2011), *MplusAutomation* (Version 0.8; Hallquist & Wiley, 2018), *papaja* (Version 0.1.0.9942; Aust & Barth, 2020), *prophet* (Version 0.6.1; O'Hara-Wild, 2020; Taylor & Letham, 2020), *purrr* (Version 0.3.4; Henry & Wickham, 2020a), *Rcpp* (Version 1.0.6; Eddelbuettel & François, 2011; Eddelbuettel & Balamuta, 2018), *readr* (Version 1.4.0; Wickham & Hester, 2020), *rebus* (Version 0.1.3; Cotton, 2017), *rlang* (Version 0.4.10; Henry & Wickham, 2020b), *rstan* (Version 2.21.2; Stan Development Team, 2020a), *serCymruTools* (Version 0.1.3; Will, 2021), *StanHeaders* (Version 2.21.0.7; Stan Development Team, 2020b), *stringr* (Version 1.4.0; Wickham, 2019), *tibble* (Version 3.1.0; Müller & Wickham, 2021), *tidyLPA* (Version 1.0.8; Rosenberg et al., 2018), *tidyr* (Version 1.1.3; Wickham, 2020), *tidyverse* (Version 1.3.0; Wickham, Averick, et al., 2019), and *tsibble* (Version 1.0.0; Wang et al., 2020) for all our analyses, functions are available in the `serCymruTools` package (Will, 2021), and all other code is available at https://github.com/w-hardy/sercymru.

Results

Counterfactual forecasts

All drugs

```
## Warning: Removed 2 rows containing missing values (position_stack).
```

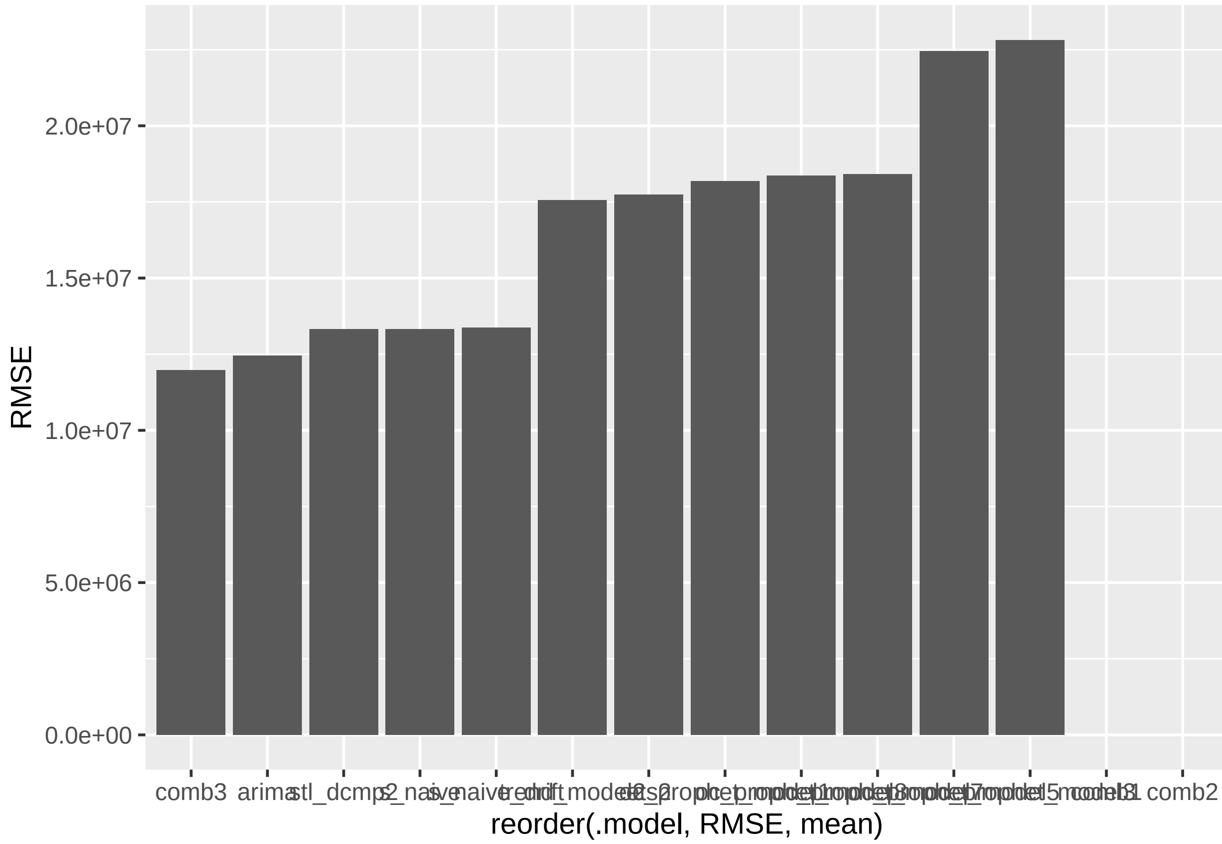

Figure 1

All prescribing, model accuracy

```
## Warning: Removed 108 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 12 row(s) containing missing values (geom_path).
```

Specific drugs

Discussion

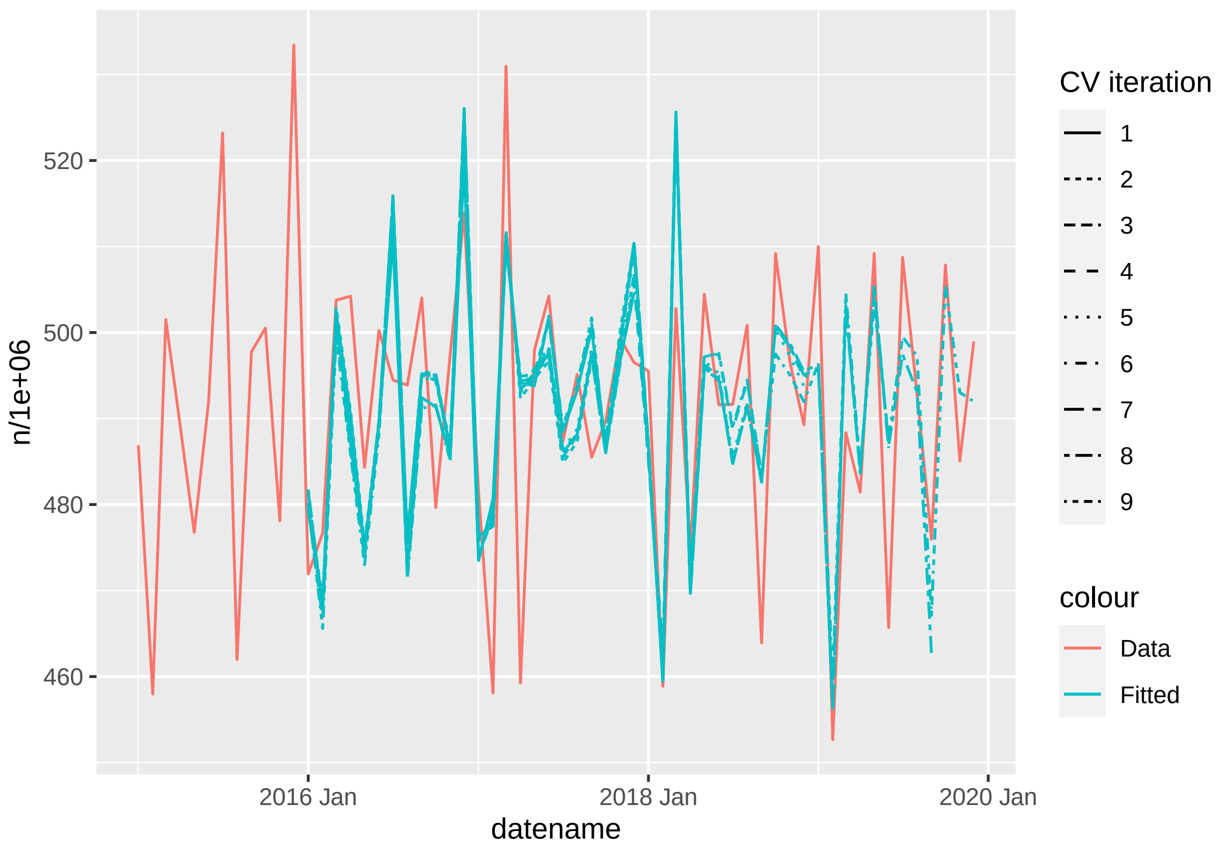

Figure 2

All prescribing, best model CV fit

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.* https://github.com/crsh/papaja

Bengtsson, H. (2020). *A unifying framework for parallel and distributed processing in r using futures.* https://arxiv.org/abs/2008.00553

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. .., & Ljung, G. M. (2015). *Time series analysis: forecasting and control* (5th ed.). John Wiley & Sons.

Chatfield, C. (1978). The Holt-Winters Forecasting Procedure. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *27*(3), 264–279.

Cotton, R. (2017). *Rebus: Build regular expressions in a human readable way.*

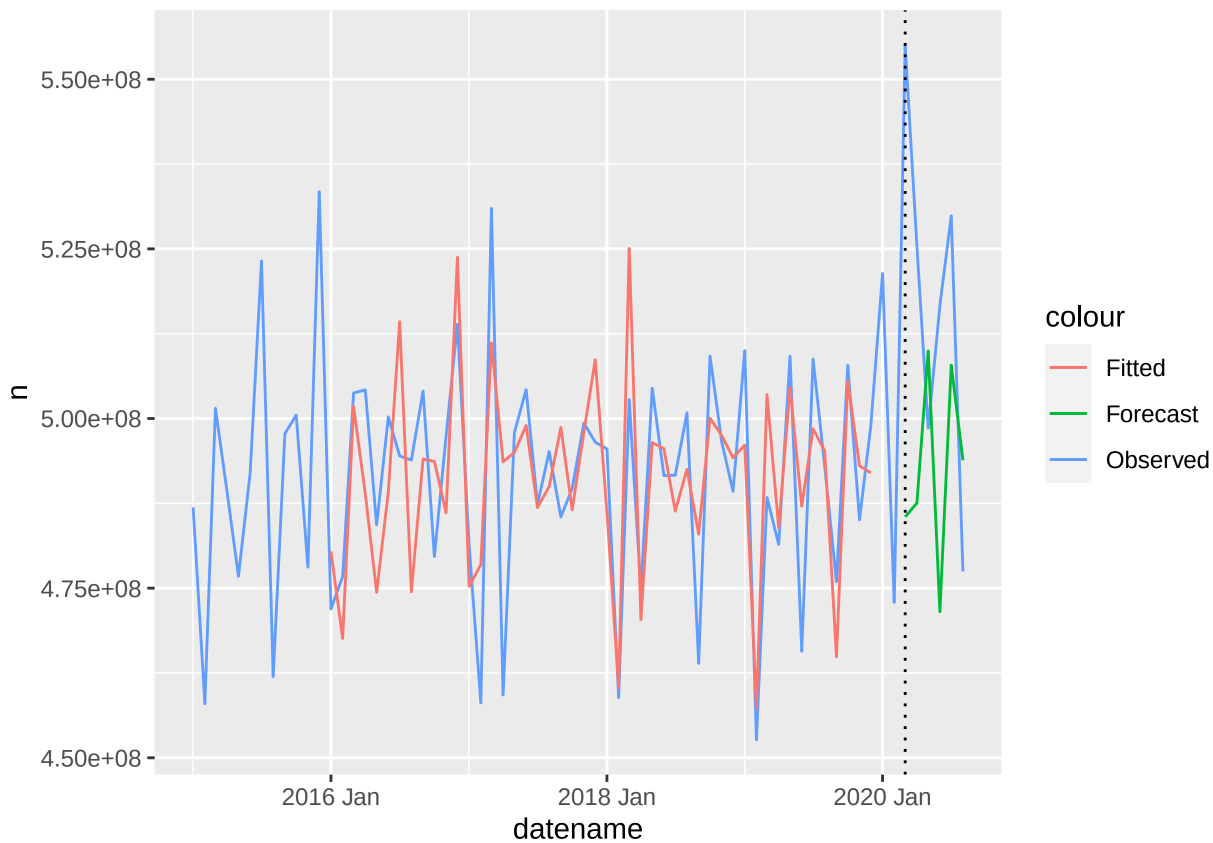

Figure 3

All prescribing, best model fit and forecast

https://CRAN.R-project.org/package=rebus

Eddelbuettel, D., & Balamuta, J. J. (2018). Extending extitR with extitC++: A Brief Introduction to extitRcpp. *The American Statistician*, *72*(1), 28–36. https://doi.org/10.1080/00031305.2017.1375990

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. https://doi.org/10.18637/jss.v040.i08

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. https://www.jstatsoft.org/v40/i03/

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling*, 1–18.

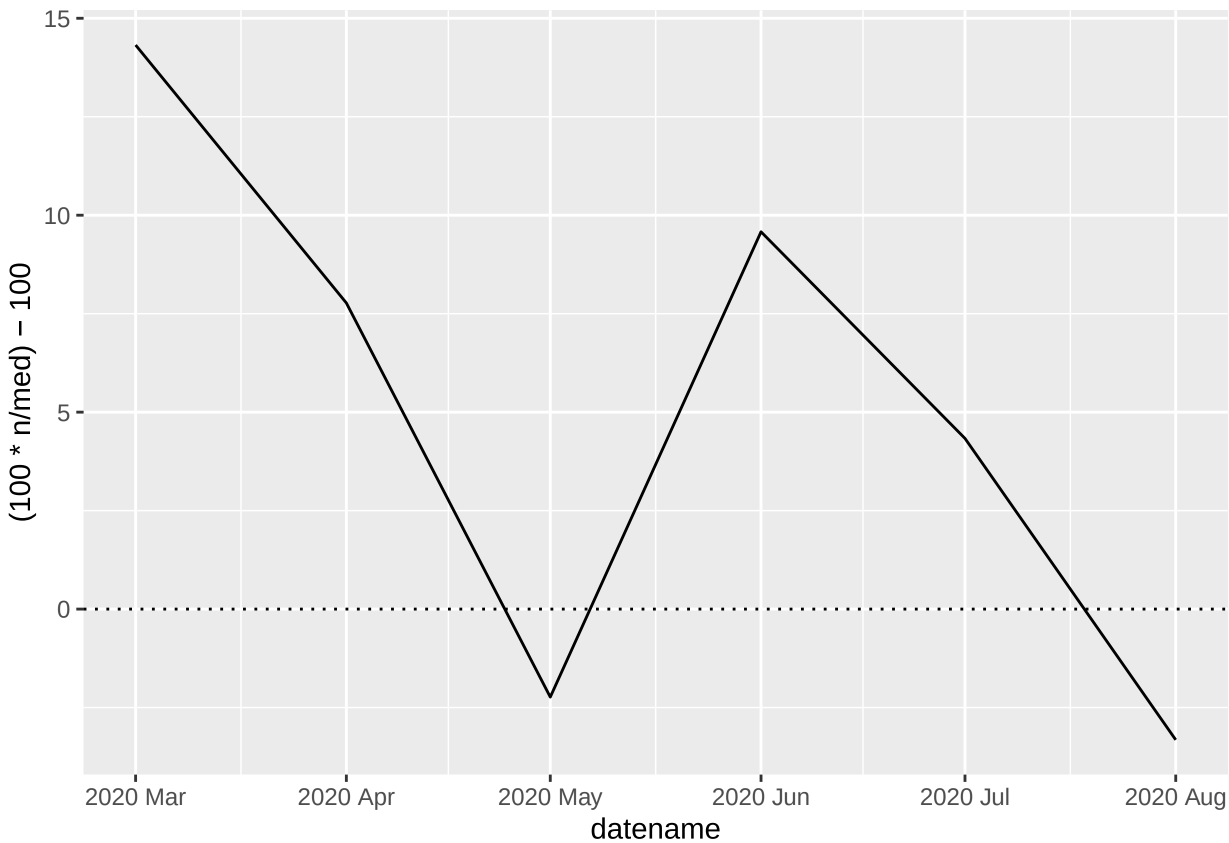

Figure 4

Percentage difference in all prescribing

https://doi.org/10.1080/10705511.2017.1402334

Henry, L., & Wickham, H. (2020a). *Purrr: Functional programming tools.* https://CRAN.R-project.org/package=purrr

Henry, L., & Wickham, H. (2020b). *Rlang: Functions for base types and core r and 'tidyverse' features.* https://CRAN.R-project.org/package=rlang

Marsh, H. W., Muthén, B. O., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). *Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching* (Vol. 16, pp. 439–476). https://doi.org/10.1080/10705510903008220

McLachlan, G., & Peel, D. (2000). *Finite mixture models.* Wiley.

Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. https://CRAN.R-project.org/package=tibble

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

O'Hara-Wild, M. (2020). *Fable.prophet: Prophet modelling interface for 'fable'*. https://CRAN.R-project.org/package=fable.prophet

O'Hara-Wild, M., Hyndman, R., & Wang, E. (2021a). *Fable: Forecasting models for tidy time series*. https://CRAN.R-project.org/package=fable

O'Hara-Wild, M., Hyndman, R., & Wang, E. (2021b). *Fabletools: Core tools for packages in the 'fable' framework*. https://CRAN.R-project.org/package=fabletools

O'Hara-Wild, M., Hyndman, R., & Wang, E. (2021c). *Feasts: Feature extraction and statistics for time series*. https://CRAN.R-project.org/package=feasts

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C. J., & Schmidt, J. A. (2018). TidyLPA: An r package to easily carry out latent profile analysis (lpa) using open-source or commercial software. *Journal of Open Source Software*, *3*(30), 978. https://doi.org/10.21105/joss.00978

Schwartz, G. (1978). Estimating the dimesion of a model. *The Annals of Statistics1*, *6*(2), 461–464.

Stan Development Team. (2020a). *RStan: The R interface to Stan*. http://mc-stan.org/

Stan Development Team. (2020b). *StanHeaders: Headers for the R interface to Stan*. https://mc-stan.org/

Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *American Statistician*, *72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Taylor, S., & Letham, B. (2020). *Prophet: Automatic forecasting procedure.* https://CRAN.R-project.org/package=prophet

Thomson, M. E., Pollock, A. C., Önkal, D., & Gönül, M. S. (2019). Combining forecasts: Performance and coherence. *International Journal of Forecasting*, *35*(2), 474–484. https://doi.org/10.1016/j.ijforecast.2018.10.006

Vaughan, D., & Dancho, M. (2021). *Furrr: Apply mapping functions in parallel using futures.* https://CRAN.R-project.org/package=furrr

Wang, E., Cook, D., & Hyndman, R. J. (2020). A new tidy data structure to support exploration and modeling of temporal data. *Journal of Computational and Graphical Statistics*, *29*(3), 466–478. https://doi.org/10.1080/10618600.2019.1695624

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations.* https://CRAN.R-project.org/package=stringr

Wickham, H. (2020). *Tidyr: Tidy messy data.* https://CRAN.R-project.org/package=tidyr

Wickham, H. (2021). *Forcats: Tools for working with categorical variables (factors).* https://CRAN.R-project.org/package=forcats

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation.* https://CRAN.R-project.org/package=dplyr

Wickham, H., & Hester, J. (2020). *Readr: Read rectangular text data.* https://CRAN.R-project.org/package=readr

Will, H. (2021). *SerCymruTools: Tools for analysing gp prescribing data.*