# Updating the SICBL Prescribing Model

## Overview

* The current SICBL prescribing model relies on using linear regression to forecast the prescription spend for the remainder of the year.
* Given the clear relationship between the month of the year and the number of dispensing days to the total prescription spend, it is likely a much more simple and transparent model could be used to forecast the spend which would be easy to understand for all users.
* This document first considers 4 simple averaging approaches to modelling to estimate the spend in 2018/19 using data from 2013/14 to 2017/18.
* It then compares the best of the simple models with the results obtained using the linear regression model to estimate the spend in 2018/19, 2019/20 and 2020/21 using data from the previous 5 years for each forecast.
* We find that forecasting by averaging the prescription spend per dispensing day each month and multiplying this by the number of dispensing days in the month provides a good estimate which is comparable to the linear regression model.
* The linear regression model was found to be quite inaccurate at estimating the overall spend in some periods due to issues with the time variable whereby the spend is not necessarily changing linearly with time.
* When quantifying the difference between the models and the true values, the results vary depending on which year is being modelled, but generally speaking there is not a large discrepancy between the simple model and the linear regression model.
* We suggest the simple methodology is adopted going forward as it is much more transparent and results in estimates that are comparable to the linear regression model.

## Current Model

1. Currently the SICBL Prescription Forecasting Model relies on linear regression to forecast the spend on prescriptions for the remainder of the year broken down by each month. From this, we generate a cumulative profile allowing SICBLs to consider how their spend looks so far and the potential spend on prescriptions for the remainder of the year.
2. However, the model is potentially over-complicated for its purpose and often generates a relatively low R-squared value as there may no longer be a linear relationship between time and prescribing spend. Hence, the model may benefit from a simpler, stripped back methodology. This will ensure the model is more transparent to its users and can be easily understood by anyone hoping to use the data.

## Historic Data

1. The prescription spend per month divided by the number of dispensing days that month is shown in Figure 1 from 2013/14 to 2020/21.

Figure : The prescription spend per month divided by the number of dispensing days in that month

1. There appear to be some similarities each month between years – most noticeably a large peak in the prescription spend in December each year (which is further highlighted by the average across all years shown as a dashed red line). This is most likely a result of patients picking up extra supplies before the Christmas public holidays.
2. We can see a clear impact of COVID-19 right at the beginning of the pandemic with a sudden increase in prescription spend in March 2019/20 and April 2020/21. This is likely to be a result of many patients collecting prescriptions due to the uncertainty around when they may be able to collect them again (as well as a larger December peak possibly for similar reasons). However, overall, the data has remained relatively consistent with previous years.
3. Figure 2 shows the timeseries of the prescription spend (blue) and the number of dispensing days (orange) running from 2013/14 until 2020/21. These are shown on different scales as the raw values are not directly comparable. However, we can consider the patterns between the two datasets whereby we see that peaks in the number of dispensing days often coincide with peaks in the prescription spend.

Figure : A timeseries showing the prescription spend and the number of dispensing days each month

1. The previous figures highlight that there appears to be a decent overlap between the prescription spend and the number of dispensing days as well as the month of the year meaning some simpler forecasting may be possible based on these parameters.

## Simple Modelling

1. A much more simple model would involve averaging data from previous years based on a combination of the month and the number of dispensing days. In order to do this, we will consider data from 2013/14 until 2017/18 in order to estimate data from 2018/19. This is done to avoid using any data from the pandemic.
2. The average prescription spend based on the number of dispensing days is shown in Figure 3. As expected, we see the prescription spend increases with the number of dispensing days.

Figure : Average prescription spend from 2013/14 to 2017/18 based on the number of dispensing days.

1. We have considered 4 different simple ways to forecast the spend in 2018/19 using data from 2013/14 to 2017/18:
	1. Month – predict the prescription spend each month based on the average spend of that month
	2. Dispensing Days – predict the prescription spend for each month based on the average from all months with the same number of dispensing days
	3. Average month and dispensing days – predict the prescription spend by averaging the previous 2 values
	4. Average per dispensing day – predict the prescription spend each month by multiplying the average prescription spend per dispensing day of the specified month by the number of dispensing days
2. For each model we have also found the sum of the squared difference between the true value and the predicted value each month and taken the square root of the sum as proxy for estimating the total difference between the true values and the predicted values. A comparison between the different methods of forecasting is seen in Table 1 and Figure 4.
3. Relying only on dispensing days alone (shown in yellow) does not account for any seasonality changes (for example, the forecast for December is much too low as it does not consider the large peak observed as patients collect extra prescriptions ahead of the Christmas public holidays which results in a large spend despite the reduced number of dispensing days).
4. It is important to capture both the supply aspect (dispensing days) and the demand aspect (calendar days) in order to best forecast the overall spend on prescriptions.

Table : Different methods of forecasting the monthly spend in 2018/19

|  |  |
| --- | --- |
|  | Forecast of Monthly Spend (£) |
| Actual Data | Forecast (Month) | Forecast (Dispensing Days) | Forecast (average of month and dispensing day) | Forecast (Multiply average per dispensing day by number of dispensing days) |
| April | 643,729,503 | 629,291,786 | 632,149,869 | 630,720,828 | 618,699,298 |
| May | 676,106,275 | 644,114,291 | 636,885,057 | 640,499,674 | 655,480,406 |
| June | 667,185,784 | 650,808,141 | 672,856,378 | 661,832,259 | 661,009,916 |
| July | 660,278,394 | 670,239,566 | 672,856,378 | 671,547,972 | 654,584,857 |
| August | 680,588,999 | 650,853,386 | 687,856,378 | 669,354,882 | 661,250,550 |
| September | 655,658,645 | 677,532,417 | 651,885,057 | 664,708,737 | 655,336,939 |
| October | 720,034,583 | 712,028,407 | 695,412,426 | 703,720,416 | 723,598,460 |
| November | 694,110,101 | 664,921,408 | 677,856,378 | 671,388,893 | 675,760,228 |
| December | 670,658,487 | 692,561,902 | 637,149,869 | 664,855,886 | 680,749,664 |
| January | 688,881,110 | 658,558,822 | 677,856,378 | 668,207,600 | 669,654,779 |
| February | 616,765,638 | 610,258,842 | 637,149,869 | 623,704,355 | 604,807,774 |
| March | 672,818,416 | 670,663,237 | 677,856,378 | 674,259,807 | 670,807,936 |
|  |  |  |  |  |  |
| $\sqrt{\sum\_{}^{}(y\_{i}-\hat{y}\_{i})^{2}}$($×10^{6})$ |  | **73** | **67** | **56** | **50** |

Figure : A comparison of the true data from 2018/19 (black) with all forecasting methods

1. We can see that forecasting by averaging the prescription spend per dispensing day each month and multiplying this by the number of dispensing days that month produces the best results out of the simpler methods. This methodology is further highlighted in Table 2 where we forecast the prescription spend in 2021/22 from the beginning of the year (i.e. we average from 2016/17 to 2020/21 without using any data from 2021/22). It should be noted that the average prescription spend corrects for any policy impacts[[1]](#footnote-2) and hence this is added/removed again to give the final forecast.

Table : The methodology used in the simple model to forecast the spend in 2021/22

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Month | Average prescription spend per dispensing day |  | Number of dispensing days |  | Policy/local adjustment |  | Forecast prescription spend (adjusted) |
|  | 2016/17 to 2020/21 |  | 2021/22 |  | 2021/22 |  | 2021/22 |
|   | £ |  | days |  | £ |  | £ |
| April | 28,901,069 | x | 24 |  | + 7,766,667  |  =  |  701,392,328  |
| May | 29,017,299 | x | 24 |  | + 7,766,667  |  =  |  704,181,837  |
| June | 27,901,760 | x | 26 |  | + 7,766,667  |  =  |  733,212,426  |
| July | 27,333,786 | x | 27 |  | - 8,533,333  |  =  |  729,478,885  |
| August | 27,609,775 | x | 25 |  | - 8,533,333  |  =  |  681,711,043  |
| September | 28,406,993 | x | 26 |  | - 8,533,333  |  =  |  730,048,492  |
| October | 28,659,849 | x | 26 |  | - 17,466,667  |  =  |  727,689,398  |
| November | 28,550,630 | x | 26 |  | - 17,466,667  |  =  |  724,849,701  |
| December | 30,459,408 | x | 24 |  | - 17,466,667  |  =  |  713,559,130  |
| January | 28,246,846 | x | 24 |  | - 33,900,000  |  =  |  644,024,307  |
| February | 27,387,037 | x | 24 |  | - 33,900,000  |  =  |  623,388,890  |
| March | 28,319,183 | x | 27 |  | - 33,900,000  |  =  |  730,717,933  |
|   |   |   |   |   |   |   |   |
|  |  |  |  |  | **Total** |  | **8,444,254,369** |

1. We will consider this simple model going forward and compare with the results produced by the linear regression model.
2. Figure 5 shows a comparison of the simple model with the old linear regression model at predicting the prescription spend in 2018/19, 2019/20 and 2020/21 using data from the previous 5 years.
3. Both models do a reasonable job of estimating the spend in all 3 years (though large discrepancies are seen for March 2019/20 and April and December 2020/21 which were the months with the largest differences during the pandemic compared with previous years). The linear regression model appears to have overestimated the spend by quite a large amount in 2018/19. As mentioned, this is likely to be caused by issues with the time variable whereby there has not been a linear relationship between time and spend. However, the pattern of the spend still remains a reasonable estimate.
4. In order to consider the precision of the model, for each forecast year, we have compared the monthly spend with the minimum/maximum spend of that month over the previous 5 years’ worth of data. The results are shown in Table 3 where we see a range of -3.7% below the minimum to 2.5% above the maximum for the simple model, and 4.0% above the maximum and no months below the minimum for the linear regression model.



Figure : A comparison of the simple model with the old linear regression model at predicting the prescription spend in 2018/19, 2019/20 and 2020/21 using data from the previous 5 years.

Table 3: The greatest percentage difference between the forecast spend in each financial year and the minimum/maximum spend for any given month over the previous 5 years for the simple model (top) and linear regression model (bottom)

|  |
| --- |
| **Simple Model** |
|  | 2018/19 | 2019/20 | 2020/21 |
| Largest % below minimum for any given month | -3.7% | None below minimum | None below minimum |
| Largest % above maximum for any given month | None above maximum | 2.2% | 2.5% |
|  |
|  | **Largest % below minimum** | **-3.7%** |
| **Largest % above maximum** | **2.5%** |
| **Linear Regression model** |
|  | 2018/19 | 2019/20 | 2020/21 |
| Largest % below minimum for any given month | None below minimum | None below minimum | None below minimum |
| Largest % above maximum for any given month | 4.0% | 1.6% | 2.8% |
|  |
|  | **Largest % below minimum** | **None below minimum** |
| **Largest % above maximum** | **4.0%** |

1. Generally speaking, the model is used to estimate the proportion of spend in each month. Table 4 and Table 5 show the proportion of spend and cumulative proportion of spend each month respectively for the true values compared with the simple and linear regression models for 2018/19, 2019/20 and 2020/21.
2. In order to determine how well the models are estimating the proportions of spend, we have calculated the sum of the absolute difference between each model and the true value for each year (a lower value represents better agreement between the model and the true value).
3. The sum of the absolute difference varies year-on year. The linear regression model better estimates the proportions in 2018/19, the simple model better estimates the proportions in 2019/20, and there is very little difference in 2020/21.

Table : A comparison of the percentage spend each month (as a proportion of total yearly spend) between the true values to the simple model and the linear regression model

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2018/19** | **2019/20** | **2020/21** |
| True Value | Simple Model | Linear Regression Model | True Value | Simple Model | Linear Regression Model | True Value | Simple Model | Linear Regression Model |
| April | 8.00% | 7.80% | 8.07% | 7.71% | 7.78% | 7.98% | 8.34% | 7.82% | 7.91% |
| May | 8.40% | 8.26% | 8.40% | 8.28% | 8.20% | 8.37% | 7.82% | 7.93% | 8.08% |
| June | 8.29% | 8.33% | 8.24% | 7.79% | 7.96% | 8.09% | 8.23% | 8.49% | 8.43% |
| July | 8.21% | 8.25% | 8.34% | 8.08% | 8.50% | 8.49% | 8.50% | 8.64% | 8.55% |
| August | 8.46% | 8.34% | 8.30% | 8.55% | 8.29% | 8.33% | 7.69% | 8.07% | 8.09% |
| September | 8.15% | 8.26% | 8.35% | 8.29% | 8.17% | 8.33% | 8.64% | 8.58% | 8.54% |
| October | 8.95% | 9.12% | 9.01% | 8.96% | 8.98% | 9.06% | 8.91% | 8.89% | 8.90% |
| November | 8.63% | 8.52% | 8.57% | 8.36% | 8.60% | 8.55% | 8.33% | 8.21% | 8.39% |
| December | 8.33% | 8.58% | 8.23% | 8.40% | 8.56% | 8.21% | 8.80% | 8.47% | 8.22% |
| January | 8.56% | 8.44% | 8.35% | 8.49% | 8.47% | 8.32% | 8.29% | 8.21% | 8.22% |
| February | 7.66% | 7.63% | 7.69% | 7.77% | 8.00% | 7.88% | 7.76% | 7.68% | 7.77% |
| March | 8.36% | 8.46% | 8.46% | 9.34% | 8.49% | 8.38% | 8.69% | 9.02% | 8.90% |
|  |  |  |   |   |  |   |  |  |   |
| **Sum of absolute difference each month** |   | **1.4%** | **1.2%** |  | **2.6%** | **3.1%** |  | **2.4%** | **2.4%** |

Table : A comparison of the cumulative percentage spend each month (as a proportion of total yearly spend) between the true values to the simple model and the linear regression model

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2018/19** | **2019/20** | **2020/21** |
| True Value | Simple Model | Linear Regression Model | True Value | Simple Model | Linear Regression Model | True Value | Simple Model | Linear Regression Model |
| April | 8.00% | 7.80% | 8.07% | 7.71% | 7.78% | 7.98% | 8.34% | 7.82% | 7.91% |
| May | 16.40% | 16.06% | 16.46% | 15.98% | 15.98% | 16.35% | 16.17% | 15.75% | 15.99% |
| June | 24.69% | 24.40% | 24.70% | 23.77% | 23.94% | 24.45% | 24.39% | 24.24% | 24.42% |
| July | 32.90% | 32.65% | 33.04% | 31.85% | 32.44% | 32.94% | 32.90% | 32.88% | 32.97% |
| August | 41.36% | 40.99% | 41.34% | 40.40% | 40.72% | 41.27% | 40.59% | 40.94% | 41.06% |
| September | 49.50% | 49.25% | 49.70% | 48.69% | 48.90% | 49.60% | 49.23% | 49.52% | 49.60% |
| October | 58.45% | 58.37% | 58.71% | 57.65% | 57.88% | 58.66% | 58.14% | 58.41% | 58.51% |
| November | 67.08% | 66.89% | 67.27% | 66.01% | 66.47% | 67.21% | 66.47% | 66.62% | 66.90% |
| December | 75.41% | 75.47% | 75.51% | 74.41% | 75.03% | 75.42% | 75.27% | 75.08% | 75.12% |
| January | 83.97% | 83.92% | 83.85% | 82.90% | 83.50% | 83.74% | 83.56% | 83.30% | 83.34% |
| February | 91.64% | 91.54% | 91.54% | 90.66% | 91.51% | 91.62% | 91.31% | 90.98% | 91.10% |
| March | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
|  |  |  |   |   |  |   |  |  |   |
| **Sum of absolute difference each month** |   | **2.2%** | **1.2%** |  | **4.1%** | **9.2%** |  | **3.0%** | **2.9%** |

1. One final comparison between the models is shown in Table 6 where we have calculated the ratio of the cumulative spend at each month between the actual spend and the forecast spend for each model. A ratio of 100% would imply the cumulative proportion of spend up until that month was equal between the actual spend and forecast spend. For each month, we have found the deviation from 100% and summed this across all months to determine how well the model determines cumulative spend throughout the year. As with the previous comparisons we see the simple model performing better in 2019/20, the linear regression model performing better in 2018/19, and relatively similar results in 2020/21 (though a slightly better performance by the linear regression model).

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2018/19** | **2019/20** | **2020/21** |
| Simple Model | Linear Regression Model | Simple Model | Linear Regression Model | Simple Model | Linear Regression Model |
| Ratio of cumulative proportions between actual and forecast | Deviation from 100% (i.e. deviation from if cumulative proportions were equal) | Ratio of cumulative proportions between actual and forecast | Deviation from 100% (i.e. deviation from if cumulative proportions were equal) | Ratio of cumulative proportions between actual and forecast | Deviation from 100% (i.e. deviation from if cumulative proportions were equal) | Ratio of cumulative proportions between actual and forecast | Deviation from 100% (i.e. deviation from if cumulative proportions were equal) | Ratio of cumulative proportions between actual and forecast | Deviation from 100% (i.e. deviation from if cumulative proportions were equal) | Ratio of cumulative proportions between actual and forecast | Deviation from 100% (i.e. deviation from if cumulative proportions were equal) |
| April | 102.6% | 2.6% | 99.2% | 0.8% | 99.1% | 0.9% | 96.6% | 3.4% | 106.7% | 6.7% | 105.4% | 5.4% |
| May | 102.1% | 2.1% | 99.6% | 0.4% | 100.0% | 0.0% | 97.7% | 2.3% | 102.6% | 2.6% | 101.1% | 1.1% |
| June | 101.2% | 1.2% | 100.0% | 0.0% | 99.3% | 0.7% | 97.2% | 2.8% | 100.6% | 0.6% | 99.9% | 0.1% |
| July | 100.8% | 0.8% | 99.6% | 0.4% | 98.2% | 1.8% | 96.7% | 3.3% | 100.1% | 0.1% | 99.8% | 0.2% |
| August | 100.9% | 0.9% | 100.0% | 0.0% | 99.2% | 0.8% | 97.9% | 2.1% | 99.1% | 0.9% | 98.9% | 1.1% |
| September | 100.5% | 0.5% | 99.6% | 0.4% | 99.6% | 0.4% | 98.2% | 1.8% | 99.4% | 0.6% | 99.2% | 0.8% |
| October | 100.1% | 0.1% | 99.6% | 0.4% | 99.6% | 0.4% | 98.3% | 1.7% | 99.5% | 0.5% | 99.4% | 0.6% |
| November | 100.3% | 0.3% | 99.7% | 0.3% | 99.3% | 0.7% | 98.2% | 1.8% | 99.8% | 0.2% | 99.4% | 0.6% |
| December | 99.9% | 0.1% | 99.9% | 0.1% | 99.2% | 0.8% | 98.7% | 1.3% | 100.2% | 0.2% | 100.2% | 0.2% |
| January | 100.1% | 0.1% | 100.1% | 0.1% | 99.3% | 0.7% | 99.0% | 1.0% | 100.3% | 0.3% | 100.3% | 0.3% |
| February | 100.1% | 0.1% | 100.1% | 0.1% | 99.1% | 0.9% | 99.0% | 1.0% | 100.4% | 0.4% | 100.2% | 0.2% |
| March | 100.0% | 0.0% | 100.0% | 0.0% | 100.0% | 0.0% | 100.0% | 0.0% | 100.0% | 0.0% | 100.0% | 0.0% |
|  |  |  |   |  |   |  |   |  |  |  |   |  |
| **Sum of deviation from 100%** |  | **8.7%** |  | **3.1%** |  | **8.2%** |  | **22.6%** |  | **13.1%** |  | **10.8%** |

Table : The deviation of the cumulative proportions between the actual spend and the forecast spend for each year and each model

1. In Table 7 we see the average deviation from 100% of the ratio between the forecast spend and the actual spend across all 3 years. This highlights the inaccuracy of both models at the beginning of the financial year. Slight improvements are seen from June onwards, with larger improvements from around September.

|  |  |
| --- | --- |
|  | Average deviation from 100% |
| Simple Model | Linear Regression Model |
| April | 3.4% | 3.2% |
| May | 1.6% | 1.2% |
| June | 0.8% | 1.0% |
| July | 0.9% | 1.3% |
| August | 0.8% | 1.1% |
| September | 0.5% | 1.0% |
| October | 0.3% | 0.9% |
| November | 0.4% | 0.9% |
| December | 0.4% | 0.6% |
| January | 0.4% | 0.5% |
| February | 0.5% | 0.5% |
| March | 0.0% | 0.0% |

Table 7: The average deviation from 100% of the ratio between the forecast and actual monthly spend for each month across the 3 years

## Caveats

1. There are several caveats to this model which should be considered when interpreting these results (though most would have also been the case for the linear regression model).
2. Forecasts are done based on the previous 5 years’ worth of data and hence the forecasts themselves will not change within a financial year (i.e. the forecasts are fixed at the beginning of a financial year). However, as more data becomes available, the forecasted data is replaced with real data, meaning the overall proportions of spend will change throughout the year.
3. The proportion of spend is less accurate at the beginning of the financial year (as seen in Table 7) and hence the outputs of the model should be used with caution for the first few months of any financial year.
4. The calculations done in this document rely on using National level data. However, there is likely to be a greater forecasting error when using smaller populations (e.g. when reducing the data down to SICBL level).
5. This is a simple model and does not include any risk monitoring when forecasting the prescription spend. In other words, this model should not be relied upon to account for changes including demographic changes; changes in morbidity; the emergence of new substitute or complementary technology; changes in prices; local or international market or supply chain conditions; wider national or international policy changes; local, national or global events or circumstances that interrupt business as usual; new trends that are not represented in the source data.
6. Data are averaged over the previous 5 years and hence the total spend estimates may be an underestimate if growth in spend has occurred over the last 5 years. This is unlikely to impact the initial proportions of spend (as it would affect all months equally) but may have an effect as estimates are replaced by real data.
7. This model is likely to be impacted by the effects of the COVID-19 pandemic on prescription spend data. However, the spend only shows a noticeable difference in March and April 2020 meaning averages are unlikely to be dramatically changed by the effects of the pandemic.
1. The community pharmacy sector receive money as retained buying margin, i.e. the profit pharmacies can earn on dispensing drugs through cost-effective purchasing. The margin element is a target that DHSC aim to deliver by adjusting the reimbursement price of drugs and the delivery of margin to pharmacy is carefully monitored by DHSC. Averaging results may be very noisy without correcting for policy impacts as they may have occurred in any month of the year. [↑](#footnote-ref-2)