Appendix I

Model Choice

We employed two generalized linear models to study the relationship between polypharmacy and socio-economic and health factors. We focused our investigation on medication count (the number of medications a respondent reported to be taking) and polypharmacy status.

When considering the medication counts, the outcome variable was over-dispersed and so a standard Poisson regression model was not suitable. We explored negative binomial and quasi-Poisson modelling options to account for the overdispersion. The models gave similar prediction accuracy, yet for the negative binomial model 3.5% of (absolute) residuals were greater than 2, with largest residual 4.7, compared to 8.7% and 6.8 for the quasi-Poisson. The negative binomial model predicts around 41.1% of responses to be zero, while the observed proportion of zero counts is 49.2%. However, exploring a zero-inflated negative binomial model predicted even fewer zero counts (39.3%), while also producing larger residuals.

In order to look more specifically at factors affecting the risk of polypharmacy, we considered two binary (logistic) regression models categorizing respondents as either no polypharmacy (1-4 medications) and polypharmacy. In the first model we defined polypharmacy as being on 5-9 medications, while in the second we defined it as being on more than 5 medications. The latter compares no polypharmacy with the combined polypharmacy and extreme polypharmacy groups. Respondents not taking any medications are excluded from this analysis so that the control group is more homogeneous with regards to their overall health.

Dataset Characteristics

The Household Health Survey was conducted in two waves, the first in 2014 and the second in 2017. We decided to not do a repeated measures analysis as the majority of households were only surveyed in one wave, and the outcomes we were interested in were not assumed to change significantly over time. The 867 individuals that were surveyed in both waves were removed entirely from our analysis to avoid introducing bias caused by correlation in their responses.

We also excluded individuals that reported to be taking more than 20 prescribed medications. These higher counts are less reliable and often the result of a misunderstanding or misinterpreting the survey question. For example, the respondent might report the number of pills taken that week instead of the number of distinct medications.

There were 27 polypharmic (5 or more medications) respondents who claimed to have no medical conditions. These were removed from the negative binomial model analysis as they were unlikely to be reliable. However, they were included for the logistic regression analysis since the effect of extreme counts is reduced by grouping into polypharmacy categories.

Variables

For the negative binomial model, the outcome variable was defined as the number of prescribed medications respondents reported taking within the last week. The variables age, sex, and ethnicity were included as markers of personal demographic. Measures of socioeconomic status included education, employment, income deprivation, financial hardship, debt, neighbourhood type, and tenure of housing. The variables live alone and sense of belonging explored environmental factors, while the variables smoking and alcohol described lifestyle choices. Physical health status was

assessed with the variables morbidity, cardiovascular medication, and self-reported health. Mental health status was assessed with the depression variable, which was measured using the nine-item Patient Health Questionnaire (PHQ-9). Finally, use of NHS services was represented by variables for number of visits and the distance to each of three types of services: primary care (including GP/nurse visits in or out of home), A&E, and walk-in units. See Table I for full details of the variables used in the analysis.

For the logistic regression models, the outcome variable is an indicator variable for polypharmacy. The variable selection procedure selected fewer variables for the logistic model compared to the negative binomial model, in particular, education, financial hardship, tenure of housing, live alone, sense of belonging, and NHS services distances were not included. However, a variable for side effects was added to this model. Since the question regarding side effects only applies to individuals taking prescribed medications, it was not possible to include the variable in the negative binomial model due to the presence of respondents with zero medication count.

Table I defines each of the variables used in the analysis and details the corresponding questions in the Household Health Survey, along with any changes we made to the scoring of the variable for the purpose of analysis. A description is given for variables derived from sources other than the Household Health Survey. Numerous questions in the survey had a free-text response ("Other, please specify"), which we classified into existing groups where possible.

The following significant analysis choices were made:

- Primary care, A&E and walk-in usage were reported as counts in the survey responses. However, due to the assumed log linear relationship with covariates in a negative binomial model, a large count in any of NHS service usage variables results in an unduly large fitted value. An ordinal form of these variables decreased the magnitude of the residuals and gave smaller root mean square errors when cross validating the model. The grouping for the ordinal variables was determined by the quantiles of the corresponding count variable.
- The Household Health Survey provides two possible measures for morbidity: a binary yes/no response or a categorised health condition list. Reading out the categories in the second question elicited positive responses in many respondents who said "no" to the first. Therefore, a combination of the two answers was required, and morbidity was defined as taking a positive response to either of the two questions. The number of health conditions was then coded as an ordinal variable. Multimorbidity is defined as two or more long-term health conditions, however, we split the ordinal variable into more categories to better capture the relationship between health conditions and polypharmacy.
- Depression was coded as a binary variable, where a person is classified as depressed if they scored 10 or higher on the PHQ-9 test. A PHQ-9 score ≥10 is the advised screening cut-off point for major depression, with a test sensitivity of 88% and a specificity of 88% [1]. Since the PHQ-9 test is a screening tool and not a clinical diagnosis, we were conservative with our choice of cut-off point and opted to classify individuals who scored between 10 or higher as depressed.

[1] Kroenke K, Spitzer RL, Williams JB; The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001 Sep 16(9):606-13.

Table I: Measures used in the analysis

Variable	Source	Description / Survey question	Original scoring	Recording for analysis
Wave	N/A	Whether the respondent was surveyed in wave 1 or wave 2.		0 = Wave 1 1 = Wave 2
Age band	Office of National Statistics	N/A	1 = Under 16 years 2 = 16-17 years 3 = 18-24 years 4 = 25-34 years 5 = 35-44 years 6 = 45-54 years 7 = 55-64 years 8 = 65-74 years 9 = 75 years and over	1 = 18-34 years 2 = 35-44 years 3 = 45-54 years 4 = 55-64 years 5 = 65+ years
Sex	Office of National Statistics	N/A	1 = Male 2 = Female 3 = Other	0 = Male 1 = Female
Ethnicity	Office of National Statistics	N/A	1 = English / Welsh / Scottish / Northern Irish / British 2 = Irish 3 = Gypsy or Irish Traveller 4 = Any other White background, please specify 5 = White and Black Caribbean 6 = White and Black African 7 = White and Asian 8 = Any other Mixed / Multiple ethnic background, please specify 9 = Indian 10 = Pakistani 11 = Bangladeshi 12 = Chinese	0 = White 1 = BME

			13 = Any other Asian background, please specify 14 = African 15 = Caribbean 16 = Any other Black / African / Caribbean background, please specify 17 = Arab 95 = Any other ethnic group, please specify	
Education	Office of National Statistics	Do you have any educational qualifications for which you received a certificate? Do you have any professional, vocational or other work-related qualifications for which you received a certificate? What is your highest qualification?	1 = Yes 2 = No 1 = Yes 2 = No 1 = At degree level or above 2 = Another kind of qualification	1 = High *(degree or above)* 2 = Medium *(other qualification)* 3 = Low *(none)*
Employment	Office of National Statistics	N/A	1 = Going to school or college full time (including on vacation) 2 = In paid employment or self-employed (or temporarily away) 3 = On a Government scheme for employment training 4 = Doing unpaid work for a business that you own, or that a relative owns 5 = Waiting to take up paid work already obtained 6 = Looking for paid work or a Government training scheme 7 = Intending to look for work but prevented by temporary sickness or injury	0 = Unemployed 1 = Employed

			8 = Permanently unable to work because of long-term sickness or disability 9 = Retired from paid work Looking after the home or family 10 = Doing something else, specify	
Income deprivation	Office of National Statistics	The income domain of the Index of Multiple Deprivation (IMD). The domain measures at Lower Super Output Area (LSOA) level the proportion of the population experiencing deprivation relating to low income.		
Managing financially / financial hardship	Wealth and Assets Survey	How well would you say your household is managing financially these days?	1 = Doing well 2 = Getting by 3 = Struggling	1 = Doing well 2 = Getting by 3 = Struggling
Debt	(Adapted from) Understanding Society	I would now like to ask you about any debts, credit or loans you may have, apart from mortgages. Do you currently owe any money in any of the following ways? Please don't include debts on your credit card that you pay off in full every month.	1 = Credit Card 2 = Hire Purchase (i.e. Brighthouse) 3 = Payday lender 4 = Pawn Shop (i.e. Cash Converter) 5 = Local companies, including Moneyshop 6 = Bank Overdraft 7 = Fixed term loan from a Bank or Building 8 = Society (EXCLUDING a mortgage) 9 = Loan from a Credit Union 10 = Loan from a finance company 11 = Loan from an unlicensed money lender 12 = Loan from a friend or relative 13 = Loan or advance on wages from your employer 14 = Social Fund loan 15 = Student Loans Company 95 = Other (please specify)	0 = No 1 = Yes *(of any kind)*

			96 = None of these	
Neighbourhood type	Defined with Local Authority input		1 = Neighbourhood for Learning 2 = Deprived comparator 3 = Less deprived comparator	0 = Less deprived 1 = Deprived
Tenure of housing	Health Survey for England	In which of these ways does your household occupy this accommodation?	1 = Own it outright 2 = Buying it with the help of a mortgage or loan 3 = Part rent and part mortgage (shared ownership) 4 = Rent it 5 = Live here rent-free (incl. rent-free in relative's/friend's property excluding squatting) 6 = Squatting 7 = Other	1= Own 2 = Mortgaged 3 = Rent or other
Live alone	Office of National Statistics	How many people live here including you?	*Numeric*	*Live alone* 0 = No 1 = Yes
Sense of belonging	Community life / Citizenship survey	How strongly you feel you belong to your immediate neighbourhood?	1 = Very strongly 2 = Fairly strongly 3 = Not very strongly 4 = Not at all strongly 5 = Don't know	0 = Positive 1 = Negative
Smoking	(Adapted from) Merseyside Lifestyle Survey	Which best describes you? If asked, smoking refers to any kind of tobacco, including cigarettes, roll ups, pipe tobacco, cigars, or shisha.	1 = I have never smoked 2 = I used to smoke occasionally but do not smoke at all now 3 = I used to smoke daily but do not smoke at all now 4 = I smoke occasionally but not every day 5 = I smoke daily	1 = Never 2 = Ex-smoker 3 = Current smoker
Alcohol	Merseyside Lifestyle Survey	Do you ever drink alcohol?	1 = Yes 2 = No	1 = Never 2 = Irregular *(fewer than one a week)*

		On average, how often do you drink alcoholic drinks?	1 = Every day of the week 2 = Four to six times a week 3 = One to three times a week 4 = A couple of times a month 5 = Less than once a month 6 = Don't know/never	3 = Regular *(one or more times a week)*
Morbidity	Office of National Statistics / Health Survey for England	Do you have any physical or mental health conditions or illnesses lasting or expected to last for 12 months or more?	1 = Yes 2 = No	*Number of health conditions* 1 = No conditions 2 = One 3 = Two 4 = Three or four 5 = Five or more
	Psychiatric Morbidity Survey	Have you ever had any of [these health conditions] over the past 12 months?	1 = Cancer 2 = Diabetes 3 = Epilepsy/fits 4 = Migraine or other frequent headaches 5 = Dementia or Alzheimer's disease 6 = Any mental health issue 7 = Cataracts / eyesight problems (even if corrected with glasses or contacts) 8 = Ear/hearing problems (even if corrected with a hearing aid) 9 = Stroke 10 = Heart attack/angina 11 = High blood pressure 12 = Bronchitis/emphysema 13 = Asthma 14 = Allergies 15 = Stomach ulcer or other digestive problems 16 = Liver problems 17 = Bowel/colon problems 18 = Bladder problems/incontinences	

			19 = Arthritis 20 = Bone, back joint or muscle problems 21 = Gout 22 = Skin problems 95 = Other, please specify 96 = None of these	
Cardiovascular	Health Survey for England	Have you taken any of these classes of medication in the last week?	1 = Cardiovascular medicine 2 = Anti-hypertensive medicines 3 = Lipid-lowering medicines 4 = Antiplatelet medicines 5 = Proton pump inhibitors 6 = Analgesics and/or NSAIDs 7 = Antidepressant medicines 8 = Medicines for asthma or COPD 9 = Antidiabetic medicines 10 = Antibacterial medicines 11 = Antipsychotic medicines 12 = Contraceptive pill 95 = Other (please specify) 96 = None of these	*Takes cardiovascular medicine* 0 = No 1 = Yes
Self-reported health	EQ-5D-3L	To help people say how good or bad their health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0. We would like you to indicate on this scale how good or bad your own health is today, in your opinion.		0 = Good *(50 or greater)* 1 = Poor *(less than 50)*
Side effects	N/A	Do any of your medications cause side effects or bother you in any way?	1 = Yes 2 = No	0 = No 1 = Yes

Depressed	Patient Health Questionnaire (PHQ-9)	Assesses how often participants had been bothered by pro problems such as "Feeling down, depressed, or hopeless" over the past two weeks. Depression severity (calculated by summing the scores across the statements)	0 = Not at all 1 = Several days 2 = More than half the days 3 = Nearly every day 0-4 none, 5-9 mild, 10-14 moderate, 15-19 moderately severe, 20-27 severe.	0 = No *(score 0-9)* 1 = Yes *(score 10-27)*
Primary care usage	SANAD2 trial	Have you, over the past 12 months because of any condition you have or other health reasons: Been seen by a practice nurse at the GP's surgery? Been seen by the family doctor or another GP at the surgery? Been seen by a nurse at home? Been seen by the family doctor or another GP at home?	1 = Yes 2 = No If yes, please specify how many times in the past 12 months	1 = None 2 = 1-3 visits 3 = 4-6 visits 4 = 7-18 visits 5 = >18 visits
A&E usage	SANAD2 trial	Have you been to a hospital casualty/A&E/urgent care department over the past 12 months because of any condition you have or other health reasons?	1 = Yes 2 = No If yes, please specify how many times in the past 12 months	1 = None 2 = 1 visit 3 = 2-3 visits 4 = >3 visits
Walk-in unit usage	SANAD2 trial	Have you been to a walk-in centre or minor injury unit over the past 12 months because of any condition you have or other health reasons?	1 = Yes 2 = No If yes, please specify how many times in the past 12 months	1 = None 2 = 1-2 visits 3 = >2 visits
Polypharmacy	Health Survey for England	How many different types of prescribed medication have you taken this week?		*Medication count* Numeric *Polypharmacy*

				1 = no polypharmacyy *(1-4 medications)* 2 = Polypharmacy *(5-9 medications)* 3 = Extreme polypharmacy *(10+ medications)*

Appendix II

Additional logistic regression analysis

Table A shows the results of a second logistic regression analysis carried out to explore the risk factors for polypharmacy and extreme polypharmacy.

Table A: Logistic regression model comparing the probability of polypharmacy and extreme polypharmacy (5-20 medications) to no polypharmacy (1-4 medications) (n=2848).

Parameter	Parameter level	Coefficient OR	Confidence interval (95%)	p-value
Intercept		**0.021**	**(0.010, 0.047)**	**<0.001**
Wave	**2**	**1.267**	**(1.026, 1.565)**	**0.028**
Neighbourhood type	*Reference: Less deprived*			
	Deprived	**1.543**	**(1.055, 2.256)**	**0.026**
Age band	Reference: 18-34			
	35-44	**2.388**	**(1.464, 3.897)**	**<0.001**
	45-54	**3.236**	**(2.045, 5.121)**	**<0.001**
	55-64	**4.917**	**(3.137, 7.706)**	**<0.001**
	65+	**4.955**	**(3.180, 7.722)**	**<0.001**
Sex	Female	0.957	(0.778, 1.177)	0.676
Ethnicity	BME	0.688	(0.426, 1.114)	0.128
Working	**Yes**	**0.624**	**(0.463, 0.840)**	**0.002**
Income (IMD score)		0.912	(0.375, 2.216)	0.839
Debt	**Yes**	**0.697**	**(0.532, 0.912)**	**0.009**
Smoking	*Reference: never*			
	Ex-smoker	**1.314**	**(1.035, 1.668)**	**0.025**
	Current smoker	0.911	(0.699, 1.187)	0.489
Alcohol	*Reference: Never*			
	Irregular	0.867	(0.662, 1.135)	0.298
	Regular	**0.655**	**(0.514, 0.835)**	**0.001**
Self-reported health	**Poor**	**2.024**	**(1.588, 2.579)**	**<0.001**
Side effects	**Yes**	**2.119**	**(1.605, 2.797)**	**<0.001**
Depressed	Yes	1.128	(0.867, 1.469)	0.370
Health condition	*Reference: no conditions*			
	One	1.564	(0.923, 2.651)	0.097
	Two	**2.589**	**(1.523, 4.400)**	**<0.001**
	Three or four	**4.328**	**(2.560, 7.319)**	**<0.001**
	Five or more	**11.958**	**(6.718, 21.288)**	**<0.001**
Cardiovascular	**Takes medication**	**3.003**	**(2.324, 3.878)**	**<0.001**
Primary care (no. visits in last year)	*Reference: none*			
	1-3	**0.574**	**(0.377, 0.874)**	**0.010**
	4-6	0.966	(0.639, 1.462)	0.871
	7-18	1.304	(0.861, 1.975)	0.211
	>18	1.284	(0.764, 2.159)	0.345
A&E (no. visits in last year)	*Reference: none*			

	1	1.054	(0.792, 1.403)	0.718
	2-3	0.941	(0.684, 1.296)	0.711
	>3	**2.389**	**(1.550, 3.683)**	**<0.001**
Walk-in unit (no. visits in last year)	*Reference: none*			
	1-2	0.967	(0.707, 1.322)	0.834
	>2	1.251	(0.743, 2.107)	0.399

We note that all variables that were significant in the first regression analysis (Table 1, main article) are also significant in this analysis. In addition, being surveyed is wave 2 is significantly associated with a 27% increase in odds of polypharmacy compared to wave 1. Living in a more deprived neighbourhood is significantly associated with a 54% increase in odds of polypharmacy. Finally, having 1-3 primary care appointments decreases the odds of polypharmacy by 42% compared to no appointments.

Appendix III

Diagnostics

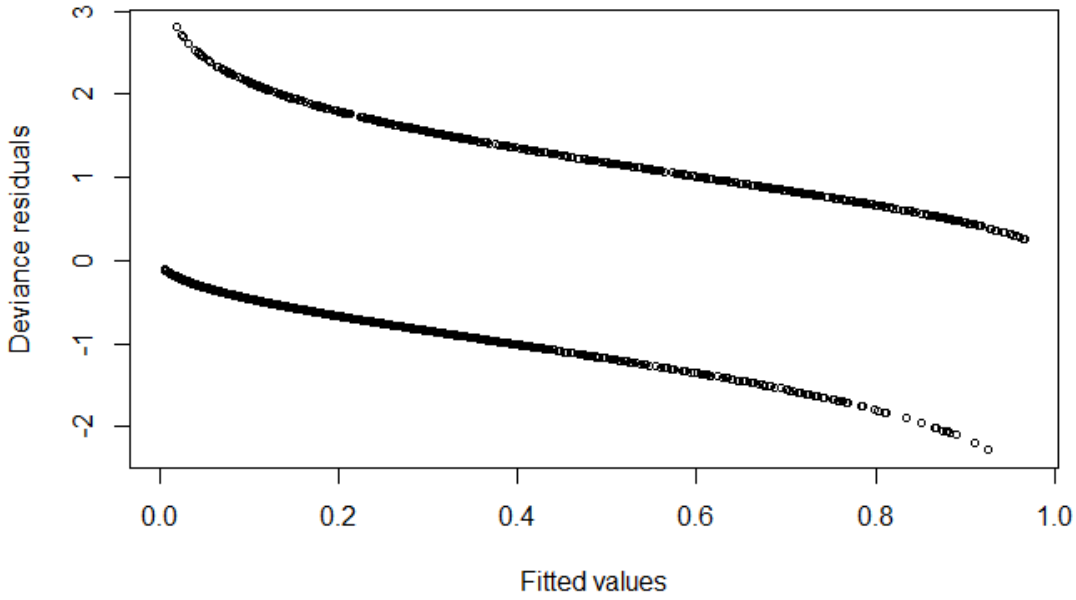
First, we present the results of diagnostic testing for the logistic regression model comparing the probability of polypharmacy (5-9 medications) to no polypharmacy (1-4 medications) (corresponding model output displayed in Table 1 of the main paper).

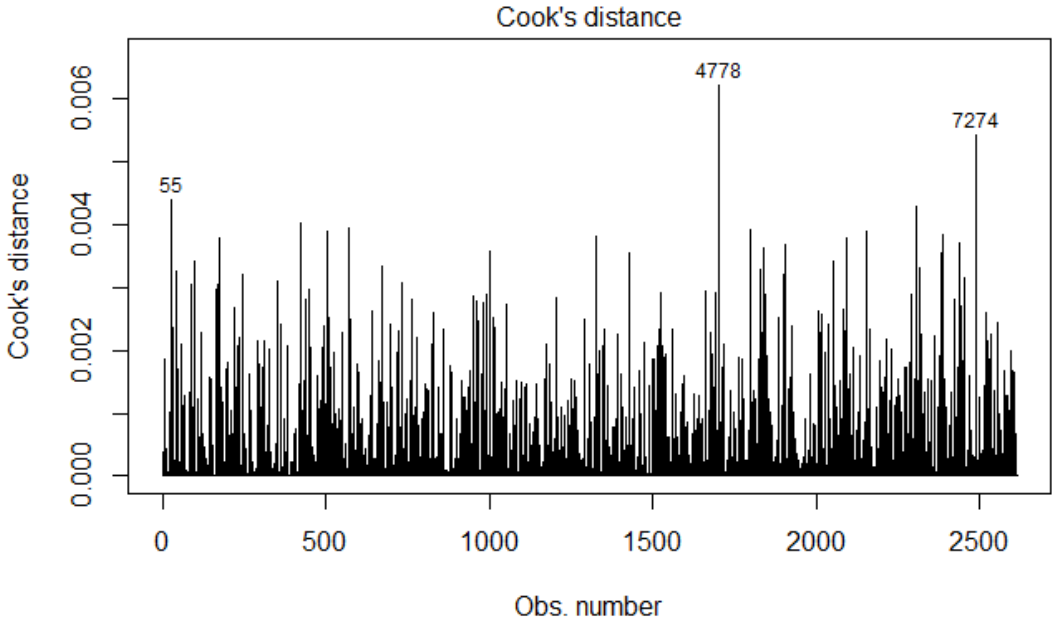
The first two plots show the residuals follow a pattern as to be expected from a logistic regression analysis. The maximum (absolute) deviance residual is 2.81, and 97.1% of the (absolute) deviance residuals are less than 2, suggesting a good model fit.

In addition, an analysis of Cook's distance does not suggest any highly influential points (given the large sample size of this study, we use the simple guideline that a Cook's distance larger than 1 is considered highly influential).

Very similar diagnostic results are observed for the additional logistic regression analysis in Appendix II. We do not present the results here.

No


Deviance residuals
3
2
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0
Fitted values

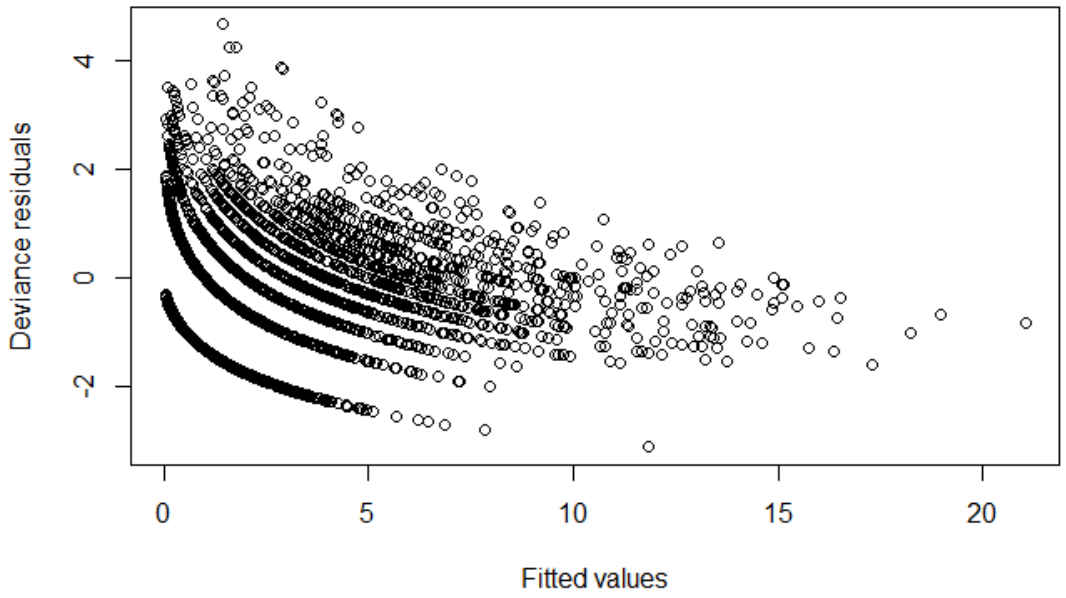

Cook's distance
4778
7274
55
Cook's distance
0.006
0.004
0.002
0.000
0
500
1000
1500
2000
2500
Obs. number

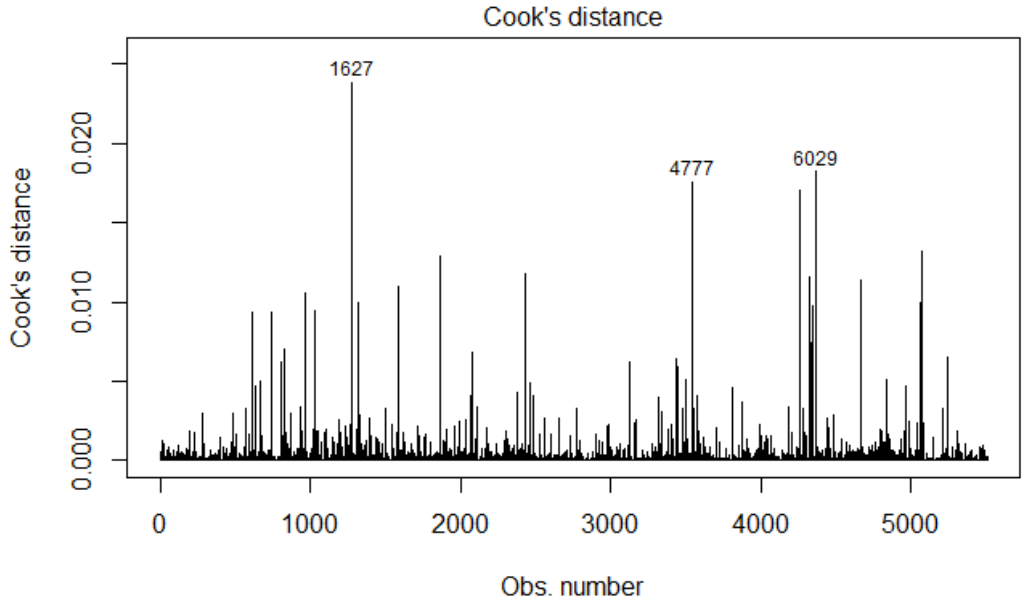
Now, turning our attention to the negative binomial model for number of prescribed medications reported by respondents (corresponding model output displayed in Table 2 of the main paper).

An argument for the use of the negative binomial model over the Poisson model due to over dispersion in the data has already been given in Appendix I. As mentioned previously, 96.5% of the (absolute) deviance residuals are less than 2. There are, however, 33 observations with residuals greater than 3, and 3 residuals with observations greater than 4.

An analysis of Cook's distance shows there are not any observations with a Cook's distance greater than 1 but there are a few points that could be argued as "significantly greater than the rest". However, removing these points from the model does not significantly impact the model output or model fit.

Overall, the count model fits reasonably well, but there is arguably some room for improvement. However, given the large number of covariates considered, the steps taken to mitigate the impact of inaccurate survey responses, and using a model that accounts for over-dispersion, it is unlikely this could be achieved without either additional data or more advanced statistical modelling techniques. This could be an area for further research.




Cook's distance
1627
4777
6029
Cook's distance
0.000
0.010
0.020
0
1000
2000
3000
4000
5000
Obs. number

Appendix IV

The analyses in Appendices IIV-V use the subset of respondents who reported taking at least one prescribed medication.

Question 83 of the questionnaire asked respondents to rate the difficulty they have with certain aspects of taking their medicines. The responses to Q83 are displayed in Table B. Note that where percentages do not sum to 100 across rows, the remaining percentage is NAs (either refused to answer or answered "Don't know/not applicable").

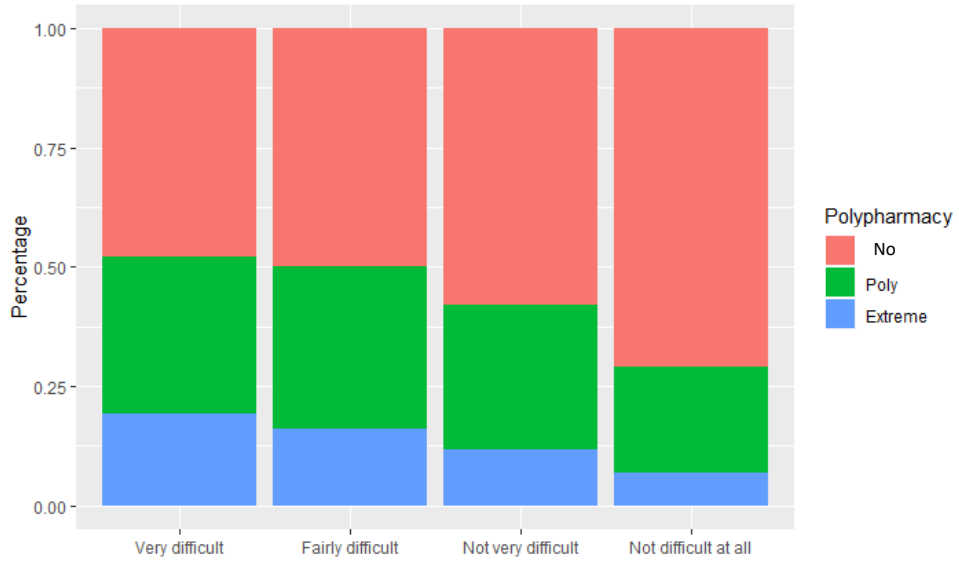
Table B: Percentages (counts) of responses for each of the 5 statements in Q83.

	Not difficult at all	Not very difficult	Fairly difficult	Very difficult
Open or close the medication's packaging	78.9% (2478)	8.54% (268)	7.23% (227)	2.77% (87)
Read the print on the packaging	75.3% (2365)	10.2% (322)	8.54% (268)	3.38% (106)
Remember to take all the pills / dose	76.9% (2415)	10.7% (336)	7.23% (227)	2.48% (78)
Get your refills in time	82.3% (2583)	8.47% (266)	4.81% (151)	1.43% (45)
Take more than one medication at the same time	80.6% (2532)	8.89% (279)	3.06% (96)	0.92% (29)

Appendix V

Polypharmacy and difficulty taking more than one medication at the same times, as reported by Q83, are significantly associated (p-value < 0.001). The below plot illustrates the changes in reported difficulty for each polypharmacy grouping.

Figure A: Percentage of respondents in each polypharmacy category by level of difficulty with taking more than one medication at the same time.



Appendix VI

To investigate the respondents experiencing difficulties with medication further, a binary variable was created using the responses from Q83 as follows:

- 0 = Little or no difficulty ("Not difficult at all" and "Not very difficult" responses)
- 1 = Difficulty ("Fairly difficult" and "Very difficult" responses)
- NA for "Don't know/not applicable" responses or if they refused to answer

The binary variables for each individual were then summed across the 5 statements to create a difficulty score. For example, if someone had answered "Fairly difficult" or "Very difficult" for each statement then they would have a score of 5. If they had answered "Not difficult at all" or "Not very difficult" for each statement they would have a score of zero. Table C shows the counts for each score.

Table C: Percentages (counts) of each score when summing the binary variable created for Q83.

Score	**0**	**1**	**2**	**3**	**4**	**5**
Percentage (count)	77.9% (2445)	10.9% (196)	6.2% (196)	2.9% (92)	0.8% (25)	1.3% (41)

Table D is a cross tabulation of the scores from Table C by whether or not they reported a mental health condition. The percentages are row percentages.

Table D: Percentages (counts) of difficulty scores by mental health condition.

Score	**No mental health condition**	**Reported a mental health condition**
0	81.6% (1996)	18.4% (449)
1	78.3% (267)	21.7% (74)
2	68.9% (135)	31.1% (61)
3	54.3% (50)	45.7% (42)
4	60.0% (15)	40.0% (10)
5	48.8% (20)	51.2% (21)

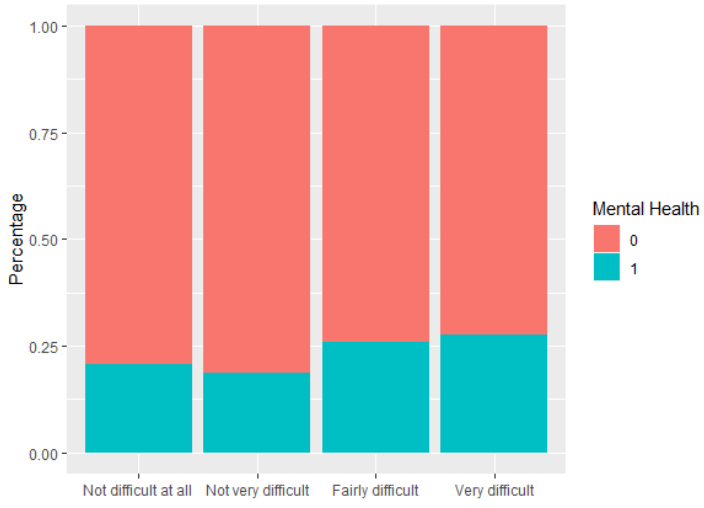
Table D shows that the percentage with a mental health condition increases with difficulty score. Out of those experiencing the greatest amount of difficulty across all statements (i.e. those with a score of 5) over half of them reported to have a mental health condition.

Difficulty score and mental health condition are significantly associated at the 5% level (chi-squared test p-value < 0.001) but the counts for scores 4 and 5 are quite small, so this relationship should be treated with caution. Furthermore, the survey does not detail the exact mental health conditions these responses represent. However, 79% of respondents who reported to have a mental health condition also reported to be taking anti-depressants.

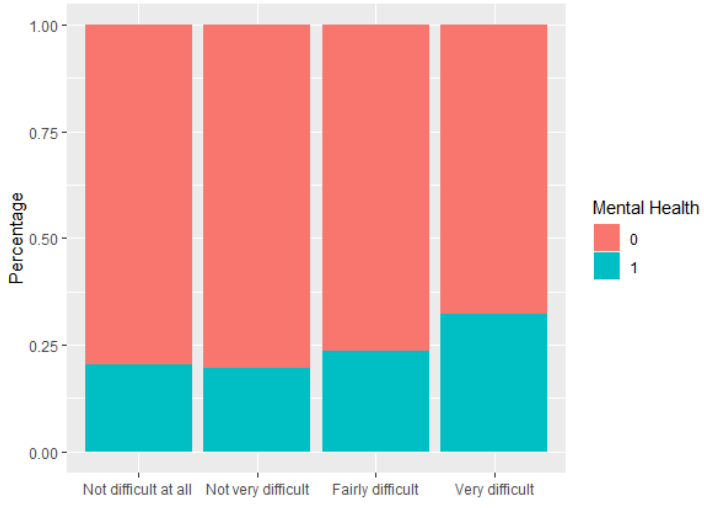
Appendix VII

The below plots illustrate the percentages by mental health condition for the responses to Q83.

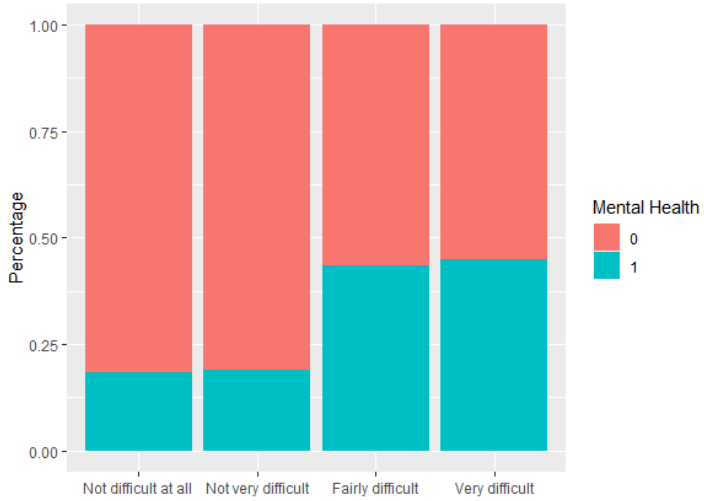
Statement 1 – opening/closing packaging



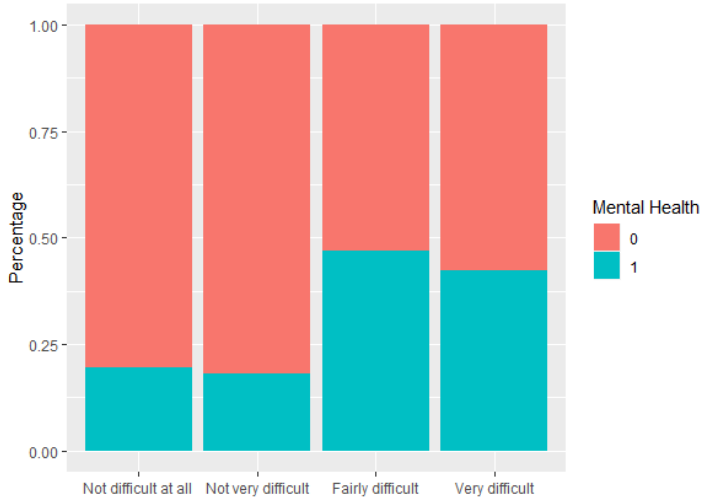
Statement 2 – reading packaging



Statement 3 – remember to take all pills/doses



Statement 4 – getting refills on time



Statement 5 – taking more than one at once