Nudge vs Superbugs: A behavioural economics trial to reduce the overprescribing of antibiotics – pre-analysis plan

We pre-registered this trial on the AEA Social Science Registry on 11 September 2017 and this pre-analysis plan was lodged on the AEA Social Science Registry at the same time. While this occurred after the launch of the trial on 9 June 2017, it was before we had received any data or commenced any analysis. Our trial pre-registration can be found here:

https://www.socialscienceregistry.org/trials/2420

1. Introduction

1.1 Abstract and motivation

The Australian Government Australia's First National Antimicrobial Resistance Strategy 2015-2019 aims to educate the public and healthcare community about the dangers of Antimicrobial Resistance and reduce over-prescription of antibiotics.

This project supports the Strategy's aims by evaluating the effectiveness of measures to reduce unnecessary prescriptions of the eight most common antibiotics by high-prescribing General Practitioners (GPs) in Australia. The measures involve various behaviourally-informed letters; 'high-prescribing GPs' refer to those in the top 30 per cent of prescribers.

1.2 Research questions

1) Will sending high prescribing GPs letters containing a social norm about their prescription rate compared to their peers reduce subsequent prescriptions compared to control or compared to a business as usual education letter?
2) Will sending delayed prescribing stickers to high prescribing GPs reduce prescription rates compared to control and is this strategy superior to a social norm letter?
3) Will social norm and delayed prescribing letters be more effective at reducing prescriptions in relatively high prescribing GPs (top 15 per cent) compared to relatively low prescribing GPs (second 15 per cent).

1.3 Interventions

- **Control** – Doctors will receive no letter.
- **Education** – Doctors will receive a business as usual educational letter aimed to increase their understanding of anti-microbial resistance.
- **Education with social norm text** – Same as above, however, also includes the text social norm "You prescribe more antibiotics than <x%> of prescribers in <region>".

- **Social norm with graph** – Provides a text social norm (as above) along with a graph to make the information clear and salient. Does not contain the educational text from the above treatment groups.
- **Delayed prescribing** - Letter to encourage use of a delayed prescribing sticker (to stick on a script), to encourage the patient to see if they improve before filling the script.

1.4 Limitations

This study uses data on prescriptions filled and not prescriptions dispensed by GPs. In order for a prescription to be counted a patient must take the script to a pharmacy and purchase the medication. Therefore, any prescriptions dispensed by GPs and not subsequently filled will not be captured in this trial.

2.1 Sampling

2.1 Population and Sampling Frame

Our study examines the behaviour of a subclass of general practitioners in Australia who are identified on the basis of having high rates of prescription for antibiotics relative to prescription levels in their statistical region. Specifically, we examine prescription of the following eight antibiotics: Amoxicillin; Cephalexin; Amoxicillin with Clavulanic Acid; Roxithromycin; Clarithromycin; Erythromycin; Cefaclor; and Trimethoprim with Sulfamethoxazole.

To identify this class we began with all general practitioners (GPs) in Australia, and excluded all non-GP prescribers, all prescribers with missing prescription data in the month preceding randomisation, and all prescribers with less than 50 scripts or less than 500 consults. There was also a small number of prescribers with prior prescribing rates that were well above the average prescription rate. To deal with this, we removed the top two per cent of prescribers. [*This paragraph was edited post-trial for clarity and accuracy.*]

This initial filtering leaves us with a pool of 22,310 general practitioners that have reliable prescription data. Our subclass of interest in this analysis consists of general practitioners who are in the top 30 per cent of these prescribers in their region (defined by the ABS ASGS SA4), based on their average prescription rate over the 12 months prior to the trial. This gives a subclass of 6,649 individual general practitioners who are 'high prescribers'.

Our sample was clustered at the clinic/practice level to prevent spill-over effects. Our 6,649 individual 'high prescribing' GPs were grouped into 3198 clusters (with an average of 2.07 individual GPs in each cluster). Clusters were randomised to five arms (the control group and four treatment groups - see section 2.3 Randomisation). The result was approximately 640 clusters in each of the five arms with each arm containing approximately 1,330 individual general practitioners.

2.2 Power

Using historical prescription data we ran a regression of our pre-trial outcome variable on past prescriptions. We found that that past prescriptions explained 25% of the variation in our outcome variable (R^2 = 0.25). Accounting for this explained variation reduced our pre-trial outcome standard deviation from 22.77 to 19.65

As this is a cluster randomised trial, we also accounted for clustering of GPs at the clinic level in our sample size calculations. Using historical prescription data we calculated an intracluster correlation coefficient (ICC) of 0.26, an average cluster size of 2.07, and from these a design effect of 1.28.

Using our adjusted standard deviation and applying our design effect we calculate that a sample size of approximately 1,330 per group provides 90% power at a 5% significance level to detect a 2.75% reduction in the antibiotic prescription rate (from 100.4 to 97.7 scripts per 1,000 consults). This is a small standardised effect size ($d = 0.13$).

2.3 Randomisation

The five trial arms (control group and four treatment groups) were randomised at the clinic/practice level. Practices were matched into quintuplets using an optimal greedy algorithm based on: the number of general practitioners in the cluster (this variable was given primacy for matching), cluster average prescription rate over the previous 12 months, and the cluster SEIFA score (a measure of socioeconomic advantage/disadvantage for the geographic area containing the practice). Practices were randomised into five arms within each quintuplet. Both matching and randomisation were implemented using a custom R script utilising the blockTools package.

3. Trial data

3.2 Data collection and processing

Pre-randomisation data was extracted and post-randomisation data will be extracted from Commonwealth Government administrative systems. Data on prescriptions does not capture all scripts dispensed by GPs, instead it captures scripts filled. Consults data is captured by a different data collection, thus it is necessary to calculate prescription rates per 1,000 consults manually (see Section 4.1: Primary Outcome).

Pre-randomisation baseline data on GPs such as clinic address, name, sex, age and major specialty were extracted from the same administrative system. Regional average prescription rates, used as the comparator in the social norm letters, were calculated by averaging individual (12-month average) prescription rates in a given region (defined by the ABS ASGS SA4). Because there is sometimes ambiguity in clinic addresses, it was necessary to manually check and assign individual doctors to clinics to ensure the integrity of the clustering variable.

All data processing and analysis steps will be performed using custom R script and will involve manual checks at each stage to ensure there are no errors introduced. Data will be collected on a rolling basis. We will receive preliminary data that may not contain all prescriptions prior to receiving finalised data. We will run analyses on the preliminary data that will be used for internal reporting purposes.

3.4 Missing data, attrition and non-compliance

Based on past administrative data we expect there will be missing outcome data. We tried to minimise this possibility by excluding any prescriber with outcome data missing in the month before randomisation. This is detailed above in the filtering protocols in Section 2.1: Sampling.

It will not be possible to determine the cause of missingness in our final outcome, nor will it be possible to verify if those in our treatment groups received and read the letters they were sent. However, because our data collection instrument is blind to assignment (data collection is mandated by regulatory processes and is collected for other reasons) and it is difficult to imagine a mechanism by which treatment could influence data collection, we expect missing data will not be statistically related to random allocation into the treatment groups.

To test for asymmetrical missingness we will run a linear regression of an attrition indicator on treatment, baseline covariates and treatment-covariate interactions and perform an F-test of the hypothesis that all interaction coefficients are zero. The covariates in this regression will be the same as those used in the covariate balance test outlined at Section 4.2: Balance Checks.

4. Empirical Analysis

4.1 Primary outcome

Our primary outcome is 'number of prescriptions per 1,000 consults'. This rate is calculated for each prescriber by summing the number of scripts filled for the eight antibiotics of interest and dividing by the number of consults billed to the Medicare for that prescriber over the same period and then multiplying this by 1,000 to express it as a rate per 1,000 consults.

The eight antibiotics of interest are: Amoxicillin, Cephalexin, Amoxicillin with Clavulanic Acid, Roxithromycin, Clarithromycin, Erythromycin, Cefaclor, Trimethoprim with Sulfamethoxazole.

4.2 Balance checks

We will perform a multinomial logistic regression to judge whether observed covariate imbalances are larger than would normally be expected from chance alone. This involves regression of a treatment indicator on the covariates.

The covariates to be included in balance checks are: pre-randomisation prescription rate per 1,000 consults over the previous 12 months, age, sex, and state (with ACT/NSW and Vic/Tas combined).

A p-value of 0.01 or less will prompt a review of the random assignment procedure and possible data-handling mistakes. If the review finds no errors, we will report the imbalance test and proceed on the assumption that the imbalance is due to chance, and report estimates with and without covariate adjustment.

4.3 Treatment Effect

Our primary analysis will be a regression analysis done on an intent-to-treat basis. Our model will be some form of appropriate regression or general linear model.

We will begin our analysis by fitting a linear regression model to the data with antibiotics prescriptions as the response variable, and with treatment groups and pre-randomisation covariates (see Section 4.6) as explanatory variables. The treatment groups and covariates are nominal variables and so will be included in the model as indicator variables. To account for clustering due

to randomising at the clinic level, we will estimate clustered standard errors by including error terms for the clinic (cluster) level and for the individual level.

The above specification is an initial model form that will be adapted as required to produce a model with appropriate fit with model assumptions. Adaptation of the model could include a log transformation of our response variable or transformation of one or more (non-nominal) variables, changing the error distribution, changing the link function or making other changes. All changes to the model form will be made on the basis of purely statistical considerations, seeking consistency of the data with model assumptions (based on diagnostic tests). In choosing an appropriate model form we will not allow consideration of effects on research outcomes.

If we find evidence of month-to-month variation we may use a difference-in-differences estimator instead of differences in means. If we also find evidence of year-to-year variation we may use a difference-in-difference-in-difference analysis, in either case we will present difference-in-means data in an appendix.

4.4 Hypothesis tests

- We will compare all treatment groups to the control group. To avoid problems arising from multiple comparisons we will conduct an initial F-test on the set of all four treatment groups to see if there is evidence of any regression relationship arising from this set. If this test yields evidence of an effect then we will conduct individual t-tests on each treatment group to find which groups show evidence of an effect.
- Any groups with effects that are both statistically and practically different from control will be compared to each other to determine superiority.
- Pre-specified subgroup analyses (see Section 4.7: Subgroups) will be conducted for control vs social norm and control vs delayed prescribing treatments.

4.5 Accounting for matching

Matching is accounted for in the above model by the inclusion of the randomisation block for each prescriber. This nominal variable will filter out any effects that are attributable to differences between prescribers arising from differences in randomisation blocks (for example, socioeconomic differences in populations for different blocks).

Because of the potential for missing data, accounting for matching in our estimation equations may be problematic, as one missing cluster will necessitate dropping the block. If there is a substantial amount of missing data then it may become infeasible to include the randomisation block in the model, in which case we may choose to drop this variable from the analysis. We will make this decision upon assessing the extent of missingness in our final dataset. Regardless of our decision we will report both estimates in an appendix to enable readers to gauge the sensitivity of this decision.

4.6 Covariates

The pre-randomisation covariates included in the above model are the age and sex of the prescriber, and the average prescription rate over the 12 months preceding the trial for each prescriber.

4.7 Subgroups

As part of our primary analysis, we will perform one pre-specified subgroup comparison, comparing treatment effects among high and low prescribers (top 15% compared to the next lowest 15%). If our social norm groups are found to be different from the control group but not different from each other, then these groups will be pooled for this analysis.

We will interact a binary indicator variable indicating whether an individual is a low or high prescriber (relative to our sample of the top 30% of prescribers) with treatment indicators. We will also perform a similar analysis by interacting a continuous prescription percentile variable.

A number of secondary subgroup analyses will also be performed. These will include comparisons of older vs younger prescribers, male vs female prescribers, and comparisons based on socioeconomic status of area and remoteness. These analyses will be considered exploratory.

4.8 Standard Error Adjustments

As recommended by Imbens and Kolesár (2016), we will perform the Bell-McCaffrey adjustment for standard errors and confidence intervals. This combines a bias-reduced heteroscedasticity/cluster robust standard error with a degrees-of-freedom adjustment.

4.9 Timepoints

Our primary analysis will include monthly time points, potentially to 12 months post intervention. We will perform the analyses specified in Section 4.3: Treatments for each month. However, if a treatment group is not different from control by the second month we will cease making comparisons for that group. If no treatment groups are significantly different from control by the second month, primary analysis of the trial will cease.

4.10 Multiple comparisons

As stated above, when conducting tests for effects from multiple categories of prescriber (e.g., when we compare the treatment groups to the control group) we will account for multiple comparisons by first testing for an effect in the *set* of categories, using appropriate statistical tests such as F-tests. If no evidence of an effect is found at this level then we will conclude that there is no effect *even if one or more of the individual categories would have shown an effect in an individual test*. This protocol ensures that we do not implicitly "cherry-pick" categories based on low p-values in tests of effects in the regression model.

5. Reporting

5.1 Deviation from this pre-analysis plan.

If our final report contains analyses that deviate from this plan we will make it clear that these analyses not pre-specified and provide justification for them. Conversely, if we omit pre-specified analyses we will make these available as supplementary material. In either case, deviation from the pre-analysis plan will be driven solely by statistical considerations and will not be influenced by any consideration of differences in findings on the research outcomes that would occur under different

model forms or methods of analysis. If findings on the research outcomes are sensitive to different plausible model forms or methods of analysis then we will report this fact in our analysis.

5.2 Outcome tables

Baseline characteristics and balance

	Control (n =)	Education (n =)	Education + social norm (n =)	Social Norm (n =)	Delayed prescribing (n =)
Baseline prescription rate per 1,000					
Age (years)					
Sex (female)					
SEIFA score					
State SA					
NSW/ACT					
WA					
Vic/Tas					
Qld					
NT					

*Data are means (SD) or numbers (%)

Main outcome table – difference from control

	Control (n =)	Education (n =)				Education + social norm (n =)				Social Norm (n =)				Delayed prescribing (n =)			
	Mean (SD)	Mean (SD)	Adjusted difference (95% CI)	p		Mean (SD)	Adjusted difference (95% CI)	p		Mean (SD)	Adjusted difference (95% CI)	p		Mean (SD)	Adjusted difference (95% CI)	p	
Month 1																	
Month 2																	

Month 3												
…												

Main outcome table – comparison between treatment groups

	Education + social norm		Social Norm		Delayed prescribing	
	Adjusted difference (95% CI)	*p*	Adjusted difference (95% CI)	*p*	Adjusted difference (95% CI)	*p*
Education						
Education + social Norm						
Social Norm						

Note: Treatment groups compared in this table will depend on preceding hypothesis tests.

Subgroup table

	Control (n =)	Treatment group(s) (n =)				
	Mean (SD)	Mean (SD)	Adjusted difference (95% CI)	*p*	*High vs Low treatment difference (95% CI)*	*Interaction p-value*
High prescribers						
Low prescribers						

Note: Treatment groups included will depend on the results of preceding hypothesis tests. Secondary subgroup analyses will be presented in a similar fashion.

6. References

Imbens, Guido W., and Michal Kolesár. 2016. "Robust Standard Errors in Small Samples: Some Practical Advice." Review of Economics and Statistics 98 (4): 701–12. doi:10.1162/REST_a_00552.