Online Supplementary File

1. Supplementary methods

1.1 Sample size

A sample size calculation was performed assuming a sensitivity of 80% with a 95% confidence interval width of 5% around the best estimate (i.e. sensitivity 0.80, 95% CI 0.75-0.85), and based on an MRH prevalence rate of 30%[1,2]. The nomogram designed by Carley *et al* (2005)[3], based on the work of Buderer *et al* (1996)[4] , was used to determine a sample size of 1500 patients. This calculation was based on the anticipated development of a risk prediction tool from the data collected in the PRIME study. The preliminary methods for the development of a risk prediction tool required a split-sample validation method, where the recruited population would be split to form a 'derivation cohort' for developing a risk prediction model and the other 50% cohort would be the validation sample. More recently 'bootstrap' resampling methods have demonstrated less bias during out-of-sample validation of risk prediction tools in comparison with a split-sample approach[5]. The bootstrap method removes the need to split the sample. On this basis, the sample size required to power the study is approximately half of the original calculation[4]. The avoidance of a type II error is the essence of a power calculation. Conventionally the *p*(type II error) is set at 0.2 such that the researcher desires <20% chance of false negative conclusion[6]. Therefore we chose to calculate a sample size estimation accordingly, assuming a sensitivity of 80%.

1.2 Missing data

To investigate whether missing data may have been missing 'not at random', a sensitivity analysis was conducted following multiple imputation using the pattern-mixture model approach[7]. Using three different values by which the sensitivity pattern was allowed to differ between the observed and the imputed data, the model remained stable.

1.3 Indicator variables

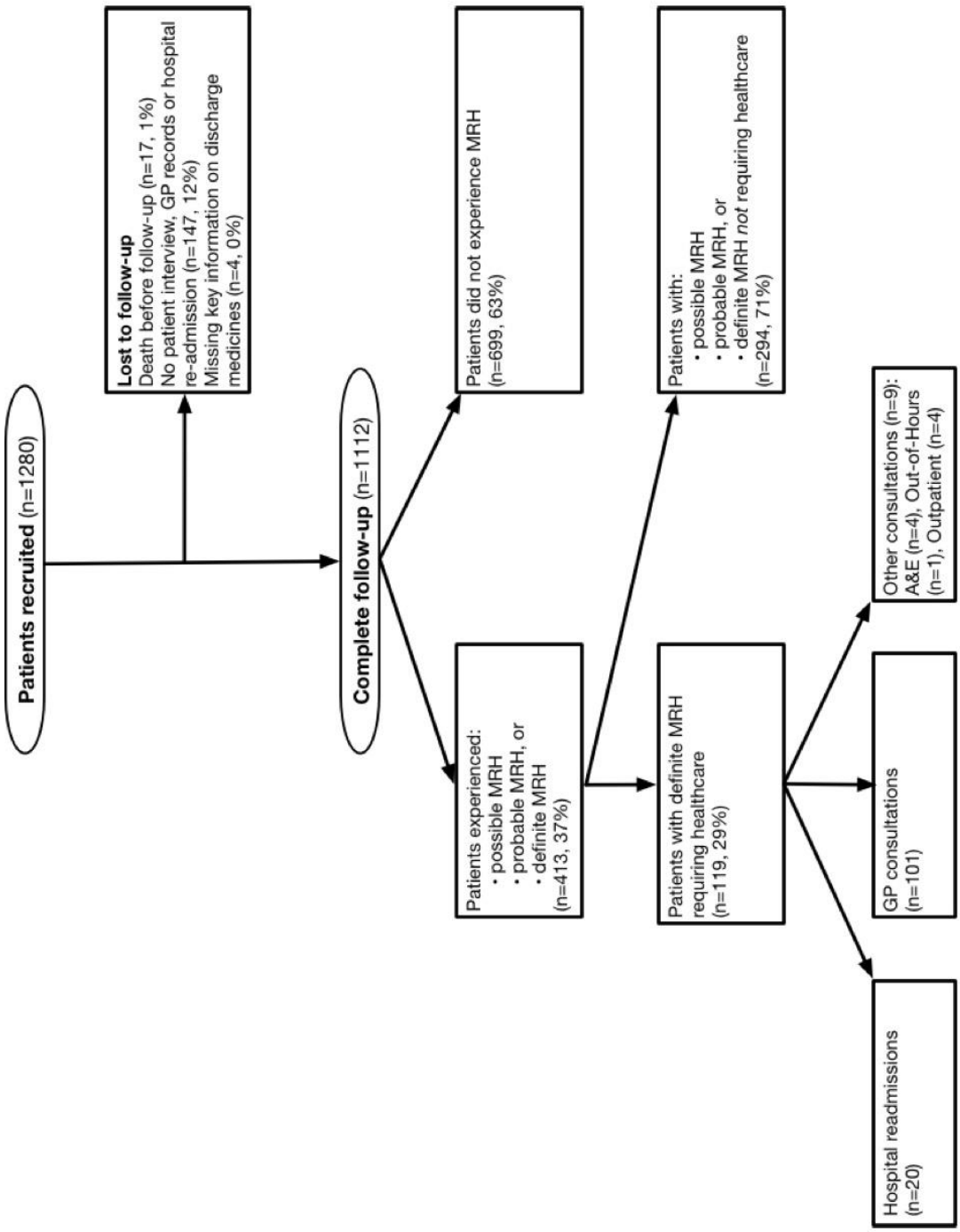
Hand grip strength was modelled as a single variable. Before including it in the model the variable was tested for an interaction with gender in relation to MRH, and this was not found to be statistically significant. The variable was also categorised separately according to gender and a test for interaction did not show any significance. The remaining variables that were not selected did not have indicator variables.

1.4 Bootstrap correction of model optimism and validation

Bootstrap is a statistical method for internal validation of risk prediction models and corrects measures of predictive performance (e.g. c-statistic) for model optimism. It is a resampling method that is used to randomly generate data (data for subsets of patients) from the original (master) dataset with replacement (patients can be selected multiple times). The bootstrap sample is the same size as the original sample. One hundred samples of the derivation data were bootstrapped and in each sample a prediction model developed and used to compute an estimate of model optimism through the following steps. First, the 'bootstrap model' was fit to the same bootstrap sample to obtain a quantitative measure of apparent performance. The bootstrap model was then fit to the original sample to obtain an 'out-of-sample' estimate of performance. The difference between the apparent performance and the out-of-sample performance defines the optimism of the model. This process was repeated 100 times, after which the final model optimism is the average of the optimism values calculated in the bootstrap iterations. This average is then subtracted from the apparent model performance measure, to obtain an optimism-corrected model performance. The linear shrinkage factor for the estimated beta-coefficients of the predictors in the model are derived from the average of the calibration slopes over the bootstrap iterations. 'Shrunk' coefficients were calculated by multiplying the original regression coefficients with the shrinkage factor (value between 0 and 1). The intercept was re-estimated based on the shrunken beta coefficients producing the final model equation. This method of coefficient shrinkage corrects for selection bias that may occur during backwards elimination (as predictors with larger coefficients by chance are more likely to be selected than predictors that by chance had smaller coefficients).[8]

2. Supplementary Results

2.1 Supplementary Figure 1. Patient flow chart



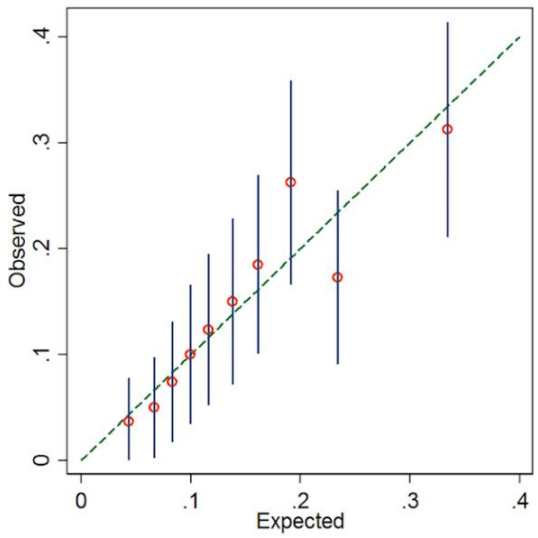
2.2 Patient Recruitment

A total of 1280 patients consented to participate in the study out of 2990 patients screened from five NHS teaching hospitals in England (Royal Sussex County Hospital, Brighton, n=199; Queen Alexandra Hospital, Portsmouth, n=227; Princess Royal Hospital, Haywards Heath, n=219; St Thomas' Hospital, London, n=397; Worthing Hospital, Worthing, n=238). This study had a complex design involving considerable resource related to the follow-up of patients by senior pharmacists to determine MRH, including hospital readmission (25% of patients recruited were readmitted). The sampling strategy was based on the resource secured to conduct both the recruitment (in hospital) and considerable patient follow up in the community (patient/carer interviews, examination of primary care records, prospectively reviewing all readmissions). This resulted in an extended recruitment period as the review of follow up data was time consuming. The implications of this were that patients could not be recruited consecutively and research sites became live at different time periods between 2013 and 2015 as determined by local logistics and research governance. However, our recruitment strategy enabled a substantially higher completion rate (68%) compared to prior research in this area (55%)[9].

2.3 Supplementary Table 1. Excluded candidate variables

Variable	**Reason for exclusion from final regression model of 12 predictors**
Albumin level	>20% missing data
C-reactive protein level	>20% missing data
White Cell count	>20% missing data
Change in accommodation	<10% prevalence in cohort
Hepatic impairment	<10% prevalence in cohort
Cardiovascular drug	Excluded as saturation in cohort (84% prevalence)
New drug added on discharge	Excluded through iterative model building
Opiate drug	Excluded through iterative model building
Anticoagulant drug	Excluded through iterative model building
Abbreviated mental test score	Excluded through iterative model building
Depression on screening score	Excluded through iterative model building
Regular falls in last year (>1)	Excluded through iterative model building
Barthel score	Excluded through iterative model building

2.4 Supplementary Figure 2. Calibration plot across tenths of predicted risk of medication harm



3. Supplementary Discussion

3.1 Supplementary Table 2. Summary of prior risk prediction tools

Study	Derivation cohort	ADR inclusion by causality	Outcome measure and verification	Model development	Predictors	Validation	Model performance
O'Mahony (2018) (ADRROP) Tool to predict in-hospital ADR	Retrospective 1687 (number of events NR) Unknown if EPV guidance followed	Probable or definite events based on WHO-UMC criteria	ADR Physician only verified events.	Backwards elimination (p<0.1)	(1)Female gender, (2)Age>70 (3)eGFR<30ml/min/1.73^2 (4)assistance with an ADL, (5)≥4 comorbidities, (6)Liver disease, (7)Number of STOPP criteria drugs (1=3 points, ≥2=6 points), (8)≥1 fall in last year	Prospective Split-sample validation N=530 (number of events NR)	0.59 (0.53-0.65) Calibration NR
Nair et al (2016) (PADR-EC) Tool to predict unplanned hospital admission due to ADR	Prospective 768 (115 events) EPV guidance not met.	Probable (Naranjo score 5-8) or definite events (≥9)	ADR Pharmacist only	Multivariable logistic regression model with score based on predictors p<0.05	(1)Number of antihypertensives (1-2 = 3 points, ≥3 = 5 points) , (2)dementia, (3)renal failure, (4)drug changes in last 3 months, (5)use of anticholinergic medication	Prospective Validation in neighbouring hospital (n=240). 30 events in validation cohort	0.67 (0.56-0.78) Calibration NR
Trivalle (2011) Tool to predict in-rehabilitation ADE	Prospective 505 (152 events) Unknown if EPV guidance followed	At least probable events included. No validated tool used to assess causality.	ADE Physician and pharmacist	Backwards elimination (p<0.05)	(1)Number of drugs (7-9 = 6 points, 10-12 = 12 points, ≥13 = 18 points) (2)Antipsychotic drug (3)Anticoagulant drug	Retrospective Bootstrap internal validation	0.70 (0.65-0.74) Calibration NR
Onder (2010) (GerontoNet) Tool to predict in-hospital ADR	Retrospective 5936 (383 events) EPV guidance met	Probable (Naranjo score 5-8) or definite events	ADR Physician only	Backwards elimination (p<0.1)	(1)≥4 comorbidities, (2)eGFR<60mL/min (3)Heart failure (4)Liver disease (5)Previous ADR (6)Number of drugs (5-7 = 1 points, ≥8 = 4 points)	Prospective Validation in sample (n=483) from 4 European countries with 56 events	0.70 (0.63-0.78) Calibration NR
Tangiisuran (2009) (BADRI) Tool to predict in-hospital ADR	Prospective 690 (86 events). EPV guidance not met.	Possible, Probable or definite events (Hallas algorithm)	ADR Physician and pharmacist	Backwards elimination (p<0.1) and forwards selection	(1)Hyperlipidaemia, (2)Number of drugs (≥8) (3)Hypoglycaemic agent (4)High WCC on admission, (5)Length of stay≥12 days	Retrospective Validation in sample (n=483) from 4 European countries with 56 events	0.73 (0.66-0.80) Calibration acceptable
McElnay (1997) Tool to predict unplanned hospital admission due to ADE	Prospective 929 (149 events). Unknown if EPV guidance followed.	Probable (Naranjo score 4-8) or definite events (≥9)	ADE Unclear	Backwards elimination (p<0.05)	(1)Digoxin, (2)antidepressants, (3)COPD, (4)Angina, (5)gastrointestinal complaint, (6)abnormal potassium, (7)patient belief that medicine influenced hospital admission	Prospective Split-sample validation (n=204). 37 events in validation cohort	C-statistic NR. Calibration NR.

ADR Adverse Drug Reaction; ADE Adverse Drug Event; WCC White Cell Count; NR Not Reported; EPV Events per Variable; eGFR estimated Glomerular Filtration Rate

References

[1] Hanlon T, Pieper F, Hajjar R, Sloane J, Lindblad I, Ruby M, et al. Incidence and predictors of all and preventable adverse drug reactions in frail elderly persons after hospital stay. J Gerontol A Biol Sci Med Sci 2006;61:511.

[2] Gray SL, Mahoney JE, Blough DK. Adverse drug events in elderly patients receiving home health services following hospital discharge. Ann Pharmacother 1999;33:1147–53.

[3] Carley S, Dosman S, Jones SR, Harrison M. Simple nomograms to calculate sample size in diagnostic studies. Emerg Med J 2005;22:180–1. doi:10.1136/emj.2003.011148.

[4] Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Acad Emerg Med 1996;3:895–900.

[5] Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. Stat Methods Med Res 2017;26:796–808. doi:10.1177/0962280214558972.

[6] Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. Emerg Med J 2003;20:453–8. doi:10.1136/emj.20.5.453.

[7] Leurent B, Gomes M, Faria R, Morris S, Grieve R, Carpenter JR. Sensitivity Analysis for Not-at-Random Missing Data in Trial-Based Cost-Effectiveness Analysis: A Tutorial. Pharmacoeconomics 2018;36:889–901. doi:10.1007/s40273-018-0650-5.

[8] Steyerberg EW, Eijkemans MJC, Harrell Jr FE, Dik J, Habbema F. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med Stat Med 2000;19:1059–79. doi:10.1002/(SICI)1097-0258(20000430)19:8<1059::AID-SIM412>3.0.CO;2-0.

[9] Gandhi TK, Weingart SN, Borus J, Seger AC, Peterson J, Burdick E, et al. Adverse Drug Events in Ambulatory Care. N Engl J Med 2003;348:1556–64. doi:10.1056/NEJMsa020703.