**Title: Borrowing strength from clinical trials in analysing longitudinal data from a treated cohort: Investigating the effectiveness of acetylcholinesterase inhibitors in the management of dementia**

**Authors:**

**Ruth Knight**[1,2*]

**Robert Stewart**[3,4]

**Mizanur Khondoker**[5]

**Sabine Landau**[2]

[1] Oxford Clinical Trials Research Unit, Centre for Statistics in Medicine, University of Oxford, Oxford, UK

[2] Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College, London, UK

[3] Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

[4] South London and Maudsley NHS Foundation Trust, London, UK

[5] Norwich Medical School, University of East Anglia, Norwich, UK

* Corresponding author

Address: Botnar Research Centre, Windmill Road, Oxford, OX3 7LD

Phone number: +441865227982

Email address: ruth.knight@csm.ox.ac.uk

**WORD COUNT:** 3319

**Abstract (218 words)**

Background: Healthcare professionals seek information about effectiveness of treatments in patients who would be offered them in routine clinical practice. Electronic medical records (EMRs) and randomised controlled trials (RCTs) can both provide data on treatment effects; however, each data source has limitations when considered in isolation.

Methods: A novel modelling methodology which incorporates RCT estimates in the analysis of EMR data via informative prior distributions is proposed. A Bayesian mixed modelling approach is used to model outcome trajectories amongst patients in the EMR dataset receiving the treatment of interest. This model incorporates an estimate of treatment effect based on a meta-analysis of RCT as an informative prior distribution. This provides a combined estimate of treatment effect based on both data sources.

Results: The superior performance of the novel combined estimator is demonstrated via a simulation study. The new approach is applied to estimate the effectiveness at 12 months after treatment initiation of AChEIs in the management of the cognitive symptoms of dementia in terms of MMSE scores This demonstrated that estimates based on either trials data only (1.10, SE=0.316) or cohort data only (1.56, SE=0.240) over-estimated this compared to the estimate using data from both sources (0.86, SE=0.327).

Conclusion: It is possible to combine data from EMRs and RCTs in order to provide better estimates of treatment effectiveness.

Key words: Randomized controlled trial; Electronic medical record; Bayesian modelling; Dementia; Cognition; Acetylcholinesterase inhibitors

**Key Messages**

- Data on a treated cohort from an EMR and data from RCTs can be combined to provide estimates of treatment effects that are less biased and more generalisable than those from either data source alone
- This holds true even if both are biased in the same direction
- Estimates from either EMRs or RCTs alone over-estimate the effects of acetylcholinesterase inhibitors in terms of MMSE scores at 12 months after treatment initiation
- It is possible to combine data from observational and randomized data sources even when the observational data is not comparative
- A concerted effort to assemble routine EMR data in a form that can be used to improve real-world inferences from RCTs is required

## Introduction

Healthcare professionals seek knowledge of the effectiveness of treatments in patients who would be offered them in routine clinical practice. Electronic medical records (EMRs) provide potentially valuable representative longitudinal data on treatment outcomes in routine clinical practice (1, 2); however, the absence of an adequate control group can often limit estimates of treatment effects. On the other hand, randomised controlled trials (RCTs) should provide an unbiased estimate of treatment effect in the population in which they are conducted (3), but may lack generalisability to patients who will be given the treatment in routine practice (4, 5). Combining data from both sources may help provide estimates that are both unbiased and generalisable.

Development of methods to combine data from both randomised and observational data sources is an ongoing area of research with a variety of methods developed in recent years (6-8). An early, influential approach was the confidence profile method (9), a direct application of Bayesian modelling which emphasizes a case-specific modelling approach. Meta-analysis is popular (e.g. 10, 11); however, these methods tend to combine aggregate-level data only and require comparative data from both sources. Similarly using a Bayesian model to incorporate aggregate-level data from one source as an informative prior distribution when analysing the other (e.g. 12, 13) also requires comparative data from both sources. Cross-design synthesis (14, 15) combines individual-level data from observational studies with aggregate-level data from RCTs, and involves the adjustment of individual study results for biases, followed by the combination of results within and across designs. The clinical applications of such methods have been limited, due to methodological complexity and individual-level data requirements. There is need for further research in this area, particularly in regard to methods that use EMR data which is a growing source of information. Many of the existing methods depend on having comparative data from an observational study rather than cohort data from an EMR.

In this paper, we propose a novel methodology for combined modelling that provides an estimator of treatment effectiveness which overcomes both the lack of an adequate control group in EMR data and

the lack of generalisability in RCT data. A Bayesian approach combines these data sources incorporating RCT estimates as part of an informative prior distribution.

The motivating clinical question for this work is the estimation of the effectiveness of acetylcholinesterase inhibitors (AChEIs) in managing the cognitive symptoms of dementia. Dementia is a major health concern, affecting 47 million sufferers in 2016, predicted to rise to 131 million by 2050 (16). There is currently no cure for most forms of dementia; however, AChEIs are often prescribed to manage cognitive symptoms (17). These drugs have been prescribed in routine clinical practice for several years, and one source of pseudonymised data on their use is the South London and Maudsley (SLaM) Biomedical Research Centre (BRC) case register (18). This EMR has been used to provide follow-up on a treated cohort of patients with a wide variety of comorbidities who receive a range of concurrent medications (19). The most commonly applied measure of cognition used in routine dementia assessment and care is the Mini-Mental State Examination (MMSE (20)) generating scores ranging from 0 to 30 with higher scores indicating better cognition. There can be situations where a patient is not able to complete all items of the MMSE for reasons unrelated to their cognition (e.g., vision impairment, mobility restrictions etc.). In this case the score may be expressed as being out of a different total (e.g., 24/29). In the remainder of this paper we will refer to the number of questions asked of a patient as the denominator and the number answered correctly as the numerator. The effects of AChEIs have also been investigated in a large number of RCTs and we recently conducted a systematic review and meta-analysis of this data (21). Synthesis of both sources of evidence offers the promise of a better estimate of the effectiveness of these treatments in routine clinical practice.

## Methods

### *Description of data*

The treated cohort used in this study was extracted from the SLaM BRC case register. Patients were included in the cohort if: (i) they had at least one mention of an AChEI (donepezil, galantamine,

rivastigmine) for which the date of treatment offer (approximated by treatment start date which is coded as the earliest date on any AChEI prescription) could be identified; (ii) they had at least one MMSE score with a denominator ≥24 recorded between 1 year before and 3 years after treatment offer (only a single MMSE score was required for inclusion and this could be before or after treatment initiation); and, (iii) they had received a primary or secondary diagnosis of dementia excluding diagnoses of Parkinson's Disease Dementia and Dementia with Lewy Bodies. For each eligible patient, all MMSE scores recorded between 1 year before and 3 years after treatment were extracted. MMSE scores with a denominator less than 30 were standardised by calculating an adjusted score as numerator divided by denominator multiplied by 30. The treated cohort contained 3134 patients with a total of 13577 scores between them, and covered the period 1st January 2005 to 8th February 2015. A previous systematic review and meta-analysis of trials of AChEIs in managing the cognitive symptoms of dementia forms the RCT dataset (21).

***Estimator of treatment effect based on treated cohort alone***

Each member of the target population, that is patients who receive this treatment in routine clinical practice, can be thought of as having two potential outcome trajectories. The one they would have followed if they were offered treatment and the one they would have followed if they were not. In practice, only the first of these is observed. Using t to denote time, with time 0 being the point of treatment offer, these two outcomes can be summarised as:

$$y_{ij,\,R=1} \mid t_{ij} \sim N(\mu_1(t_{ij}), \sigma^2) \quad \text{if the participant was offered treatment}$$

$$y_{ij,\,R=0} \mid t_{ij} \sim N(\mu_0(t_{ij}), \sigma^2) \quad \text{if the participant was not offered treatment} \qquad (1)$$

where $y_{ij,\,R=1}$ is the outcome for individual i at time $t_{ij}$ if they are offered treatment and $y_{ij,\,R=0}$ is the outcome for individual i at time $t_{ij}$ if they are not offered treatment. For an individual, the effect of treatment offer ($\Delta_i$) at a fixed time, $t=\alpha>0$, is the difference between their outcome at $\alpha$ if they were and were not offered treatment:

$$\Delta_i(\alpha) = y_{ij,\,R=1}\,|\alpha - y_{ij,\,R=0}\,|\alpha \qquad (2)$$

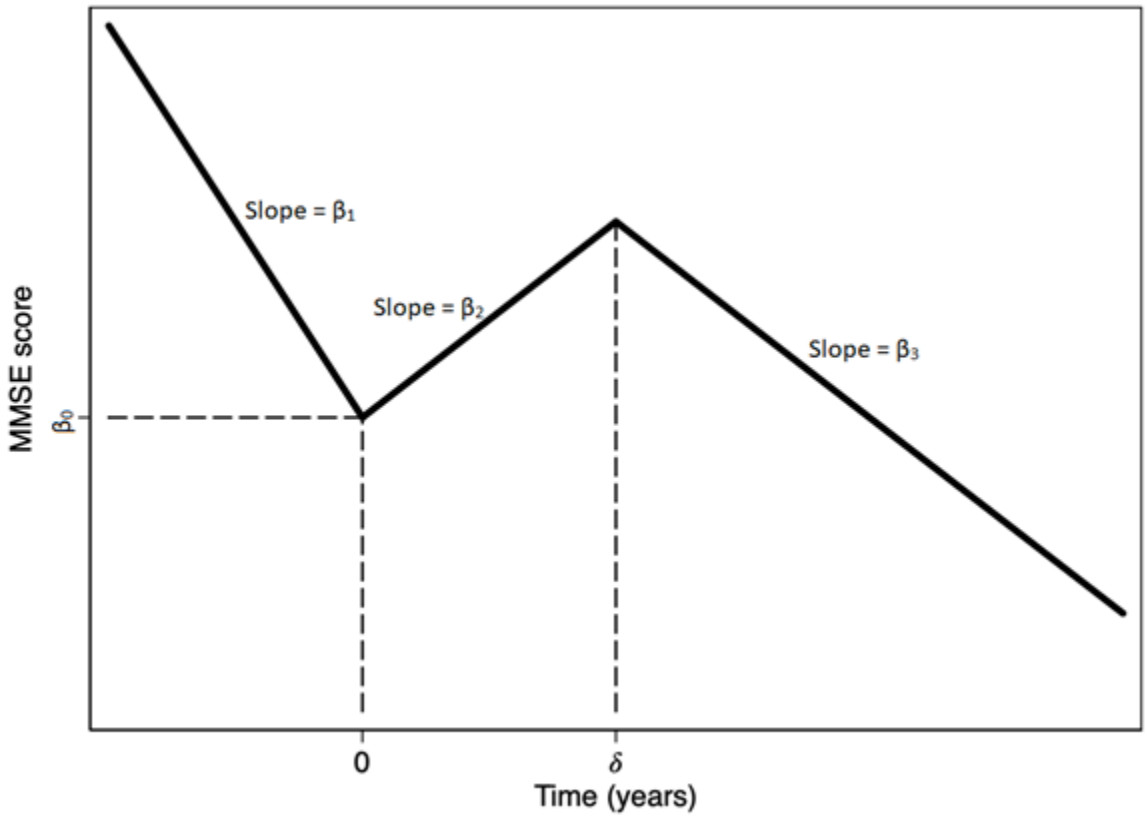
For an individual it is possible to observe only one of the two potential outcomes; therefore, the parameter we are interested in estimating is the average effect of treatment offer at $t=\alpha$ which we call the ARE:

$$ARE(\alpha) = E_i(\Delta_i(\alpha)) = E_i(y_{ij,\,R=1}\,|\alpha - y_{ij,\,R=0}\,|\alpha) = E_i(y_{ij,\,R=1}\,|\alpha) - E_i(y_{ij,\,R=0}\,|\alpha) = \mu_1(\alpha) - \mu_0(\alpha) \qquad (3)$$

In order to be able to estimate this parameter, appropriate expressions for the average trajectory in the population who are offered treatment and the population who are not are needed. Previous work (22) and non-parametric modelling of the current treated cohort have indicated that a piecewise linear mixed effects model (or alternatively linear spline) with two change points (or knot points), at treatment offer ($t=0$) and at some unspecified time subsequent ($t=\delta>0$), is appropriate to model the trajectory in those who are offered treatment (see Figure 1):

$$\mu_1(t) = \begin{cases} \beta_0 + \beta_1 t, & t<0 \\ \beta_0 + \beta_2 t, & 0\le t<\delta \\ \beta_0 + (\beta_2 - \beta_3)\delta + \beta_3 t, & t\ge\delta \end{cases}$$

$$\mu_1(t) = \beta_0 + \beta_1 t\, 1_{t<0} + \beta_2 \min(t, \delta)\, 1_{t\ge0} + \beta_3 (t-\delta)\, 1_{t\ge\delta} \qquad (4)$$




146

147 **Figure 1:** Piecewise linear model for MMSE trajectories

148 All participants in the cohort were offered treatment, and so an assumption must be made about what
149 would have happened had they not been offered treatment. The assumption made is that they would
150 have continued on their pre-treatment trajectory (A1):

151 $$\mu_0(t) = \beta_0 + \beta_1 t \qquad (5)$$

152 Having made this assumption it is possible to derive an expression for an estimator of treatment effect
153 parameter (ARE). This estimator of treatment effect is denoted $\theta_\alpha$, and may suffer from bias since it
154 relies on assumption A1:

155 $$\theta_\alpha = \mu_1(\alpha) - \mu_0(\alpha) = (\beta_2 - \beta_1)\,\alpha, \qquad \alpha<\delta$$

156 $$(\beta_2 - \beta_3)\delta + (\beta_3 - \beta_1)\alpha, \qquad \alpha\geq\delta \qquad (6)$$

157 Equation (6) can be rearranged to express $\beta_2$ in terms of the other parameters:

158 $\beta_2 = \frac{1}{\alpha}\theta_\alpha + \beta_1,$ $\alpha<\delta$

159 $\frac{1}{\delta}[\theta_\alpha - (\beta_3 - \beta_1)\alpha] + \beta,$ $\alpha\geq\delta$

160 $$\beta_2 = \beta_1 (1_{\alpha<\delta} + \frac{\alpha}{\delta} 1_{\alpha\geq\delta}) + \theta_\alpha (\frac{1}{\alpha} 1_{\alpha<\delta} + \frac{1}{\delta} 1_{\alpha\geq\delta}) + \beta_3 (1 - \frac{\alpha}{\delta}) 1_{\alpha\geq\delta} \quad (7)$$

161 This expression for $\beta_2$ can be substituted into the expected treated trajectory for those who are
162 offered treatment (eqn (4)):

163 $$\mu_1(t) = \beta_0 + \beta_1 (t\, 1_{t<0} + (1_{\alpha<\delta} + \frac{\alpha}{\delta} 1_{\alpha\geq\delta}) \min(t,\delta)\, 1_{t>0}) + \theta_\alpha (\frac{1}{\alpha} 1_{\alpha<\delta} + \frac{1}{\delta} 1_{\alpha\geq\delta}) \min(t, \delta)\, 1_{t>0} +$$

164 $$\beta_3 [(1 - \frac{\alpha}{\delta})1_{\alpha\geq\delta} \min(t,\delta)\, 1_{t>0} + (t- \delta)1_{t\geq\delta} \quad (8)$$

165 This model can be used to estimate the effect of treatment at time $t=\alpha$ ($\theta_\alpha$) based on data from a
166 cohort who were all offered treatment. Random effects on the coefficients can be incorporated to
167 allow variation between patients.

168 $$\mu_1(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})(t\, 1_{t<0} + (1_{\alpha<\delta} + \frac{\alpha}{\delta} 1_{\alpha\geq\delta}) \min(t,\delta)\, 1_{t>0}) + (\theta_\alpha + b_{2i})(\frac{1}{\alpha} 1_{\alpha<\delta} + \frac{1}{\delta} 1_{\alpha\geq\delta}) \min(t, \delta)\, 1_{t>0} +$$

169 $$(\beta_3 + b_{3i})[(1 - \frac{\alpha}{\delta})1_{\alpha\geq\delta} \min(t,\delta)\, 1_{t>0} + (t- \delta)1_{t\geq\delta} \quad (9)$$

170 To fit this model under a Bayesian framework, prior distributions for each of the parameters were
171 determined. In the absence of additional information, non-informative priors should be used (23). For
172 the coefficients, a suitable choice is a normal distribution with zero mean and large deviation. For the
173 residual standard deviation, a suitable choice is uniform on the range zero to one hundred. A suitable
174 prior distribution for a change point parameter, such as $\delta$, is a uniform prior on the range of possible
175 values (24). In this instance, a plausible range is from 0 to 3 since the second change point must come
176 after the first at t=0 and the cohort consists of scores from 0 to 3 years after treatment offer. Suitable
177 vague hierarchical priors are also placed on the random effects. For a single random effect, this is a
178 normal distribution with mean 0 and variance $\sigma_0^2$ which is given a vague prior (U(0,100)). In the
179 presence of two or more random effects, these can be modelled using a multi-variate normal
180 distribution with mean zero. Vague priors are used for the covariance matrix. In the case of two

random effects the constituent parts of the covariance matrix can be given vague priors (U(0,100) for standard deviations and U(-1,1) for the correlation). In the presence of 3 or more random effects an inverse Wishart prior distribution is used for the re-scaled covariance matrix with U(0,100) priors used for the scaling parameters (23).

### *2.3 Incorporating RCT data via informative prior distributions*

The assumption on which the estimator $\theta_\alpha$ is based may be biased, patients may not have continued on their pre-treatment trajectory. This is called projection bias, with the projection bias at time = α denoted $\varphi_\alpha$. The true treatment effect, ARE(α) can be calculated as:

$$ARE(\alpha) = \theta_\alpha - \varphi_\alpha$$

$$\theta_\alpha = ARE(\alpha) + \varphi_\alpha \quad (10)$$

This can be substituted into equation (9) to give an expression for the MMSE trajectory in the treated cohort based on the true treatment effect and the projection bias, both at time = α:

$$\mu_1(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})(t\, 1_{t<0} + (1_{\alpha<\delta} + \frac{\alpha}{\delta} 1_{\alpha\geq\delta}) \min(t,\delta)\, 1_{t>0}) +$$

$$(ARE(\alpha) + \varphi_\alpha + b_{2i})(\frac{1}{\alpha} 1_{\alpha<\delta} + \frac{1}{\delta} 1_{\alpha\geq\delta}) \min(t, \delta)\, 1_{t>0} +$$

$$(\beta_3 + b_{3i})[(1 - \frac{\alpha}{\delta})1_{\alpha\geq\delta} \min(t,\delta)\, 1_{t>0} + (t- \delta)1_{t\geq\delta} \quad (11)$$

Data from RCTs can form the basis of an informative prior distribution for ARE(α); however, this is only true for the proportion of the target population who are trial eligible. The model in equation (11) can be expanded to incorporate not only the possibility of different treatment effects in the trial eligible and trial not eligible populations, but also different trajectories within these two populations through the use of $S_i$, which takes value 1 if individual i is trial eligible and 0 otherwise, to denote whether or not the individual is part of the trial eligible population:

$$\mu_1(t) = (\beta_{01}\, 1_{Si=1} + \beta_{02}\, 1_{Si=0} + b_{0i}) + (\beta_{11}\, 1_{Si=1} + \beta_{12}\, 1_{Si=0} + b_{1i})(t\, 1_{t<0} + (1_{\alpha<\delta} + \frac{\alpha}{\delta} 1_{\alpha\geq\delta}) \min(t,\delta)\, 1_{t>0}) +$$

$$(ARE_1(\alpha)\, 1_{Si=1} + ARE_2(\alpha)\, 1_{Si=0} + \varphi_\alpha + b_{2i})(\frac{1}{\alpha} 1_{\alpha<\delta} + \frac{1}{\delta} 1_{\alpha\geq\delta}) \min(t, \delta)\, 1_{t>0} +$$

$$(\beta_{31}\, 1_{Si=1} + \beta_{32}\, 1_{Si=1} + b_{3i})[(1 - \frac{\alpha}{\delta})1_{\alpha\geq\delta} \min(t,\delta)\, 1_{t>0} + (t- \delta)1_{t\geq\delta} \quad (12)$$

where subscript 1 denotes parameters referring to the trial eligible portion of the target population and subscript 2 denotes those for the trial not eligible portion. The trial eligibility parameter is given a Bernoulli distribution:

$$S_i \sim \text{Bin}(1, \pi) \quad (13)$$

where $0 < \pi < 1$ is the proportion of the target population that are trial eligible. The overall treatment effect can be calculated as:

$$ARE(\alpha) = \pi\, ARE_1(\alpha) + (1-\pi)\, ARE_2(\alpha) \quad (14)$$

As before, each of the parameters in the model are given a prior distribution. An informative prior distribution (25) based on meta-analysis of RCTs is used for $ARE_1(\alpha)$. This meta-analysis was performed based on trials identified during a systematic review of the use of AChEIs in the management of dementia [21]. Two steps were followed to convert the meta-analysis results to a suitable informative prior distribution (26); (1) choosing an appropriate distribution; and, (2) using available information from the meta-analysis to provide estimates for the mean and variance. A normal distribution was selected with mean set as the pooled effect estimate from the meta-analysis and standard deviation set as the associated standard error. For other parameters, vague priors as described previously were used. Projection bias, $\varphi_\alpha$, can be both positive and negative and so was given a normal prior distribution with mean 0 and large variance. The trial eligible proportion, $\pi$, is a probability and was thus given a uniform prior on the range 0 to 1.

This proposed combined model relies on two assumptions; first, that there are no treatment effect moderators whose distribution differs between the trial eligible portion of the target population and the trial samples (A3); and second that the projection bias, $\varphi_\alpha$, is the same in the trial eligible and trial

not eligible portions of the target population (A4). These assumptions are weaker than those required when estimating treatment effects based on the treated cohort alone (where A1 is made instead of A3) or trial data alone, where we must assume that the trial and trial eligible portions of the target population do not differ on any characteristics which predict treatment effect (A2) rather than A4.

### *2.4 Simulation study*

To investigate the properties of the proposed new estimator, a simulation study was conducted. The target population, P, can be split into the trial eligible portion, $P_1$, and the trial not eligible portion, $P_2$. We assumed that $P_1$ could be further split into k mutually exclusive and exhaustive subsets, $P_{1j}$, representing those eligible for each of k trials. Using $Z_i$ as an indicator for trial eligibility, $Z_i$=1 if individual i is trial eligible and 0 otherwise, and setting $P(Z_i = 1) = \pi$, then the treatment effect for each individual ($\Delta_i$, eqn (2)) can be generated using:

$$\Delta_i \mid (Z_i = 1 \cap i \in P_{1j}) \sim N(ARE_{1j}, \omega^2)$$

$$\Delta_i \mid Z_i = 0 \sim N(ARE_2, v^2)$$

$$ARE_{1j} \sim N(ARE_1, \tau^2) \qquad (15)$$

The outcomes for individuals at treatment offer, $\gamma_{0,i}$, can similarly be defined as follows:

$$\gamma_{0,i} = \gamma_{i0,\,R=1} = \gamma_{i0,\,R=0}$$

$$\gamma_{0,i} \mid (Z_i = 1 \cap i \in P_{1j}) \sim N(\Gamma_{01,j}, \omega_0^2)$$

$$\gamma_{0,i} \mid Z_i = 0 \sim N(\Gamma_{02}, \sigma_0^2)$$

$$\Gamma_{01,j} \sim N(\Gamma_{01}, \tau_0^2) \qquad (16)$$

For each individual, these values are first generated and the trajectories under treatment offer are derived. The trajectory under no treatment offer is assumed to be:

$$\gamma_{it,\,R=0} = \gamma_{0,i} + \gamma_1 t + \varepsilon, \qquad t<0$$

$$\gamma_{0,i} + \gamma_2 t + \varepsilon, \quad t \geq 0 \tag{17}$$

where t=0 is the point where treatment would have been offered and the trajectory has two slopes. Equation (2) can be rearranged to show:

$$y_{i\alpha,\, R=1} = \Delta_i + y_{i\alpha,\, R=0} \tag{18}$$

Assuming that the trajectory under treatment offer has a second change point at t=δ and slope $\gamma_3$ thereafter the following expressions for the trajectory under treatment offer can be derived:

$$\text{If } \alpha<\delta: \quad y_{it,\, R=1} = y_{it,\, R=0} + \begin{cases} \frac{1}{\alpha}\Delta_i\, t, & t<\delta \\ \frac{1}{\alpha}\Delta_i\, \delta + (\gamma_3 - \gamma_2)(t - \delta), & t \geq \delta \end{cases}$$

$$\text{If } \alpha \geq \delta: \quad y_{it,\, R=1} = y_{it,\, R=0} + \begin{cases} \frac{1}{\delta}\Delta_i\, t + (\gamma_3 - \gamma_2)(1 - \frac{\alpha}{\delta}), & t<\delta \\ \Delta_i + (\gamma_3 - \gamma_2)(t - \alpha), & t \geq \delta \end{cases} \tag{19}$$

Expressions for the projection bias, $\varphi_\alpha$, and generalisability bias, ζ, are as follows:

$$\varphi_\alpha = (\gamma_2 - \gamma_1)\alpha$$

$$\zeta = ARE_1 - ARE \tag{20}$$

Trial samples for each of the k trials were generated by first assuming that the whole trial sample size, $n_1$, was split evenly amongst the k trials so that $\frac{n1}{k}$ participants from $P_{1j}$ were selected for each trial. For each their outcome at treatment offer was obtained from equation (17). Randomisation, R~Bin(1, 1/2 ), was generated to ensure 50:50 allocation, and their outcome at t=α was obtained using equation (19) if R=1 and equation (17) if R=0. For each participant their change score from time 0 to α was calculated, and these were aggregated to provide an estimate of effectiveness and associated standard error from each trial. A cohort sample was generated by selecting $n_2$ participants from the target population, P; choosing the time points at which observations are recorded for individual i, $t_{sel,i,}$ to ensure variation in location and number of measurements; generate potential outcomes at selected time points using equation (19).

Input values investigated for each of the parameters are summarised in Table 1. Parameters of interest (size of datasets, projection and generalisability biases, and proportion trial eligible (27, 28)) were investigated with the combinations considered determined using a Latin Squares Design (29) in order to reduce the number of parameter combinations considered. Situations where generalisability bias, projection bias, and both generalisability and projection bias simultaneously were present in the data were considered. Other parameters were set to reasonable values, based on example datasets.

Three competing estimators were calculated for each simulated dataset: (i) trials only estimator; (ii) cohort only estimator; and, (iii) combined estimator. Estimators were compared in terms of absolute bias, standard error (SE) and mean squared error (MSE). 500 sets of data were generated for each combination of input values.

**Table 1**: Input parameters used for simulation model

| Parameter | Description | Values |
|---|---|---|
| $n_1$ | Combined size of all trials | 1000, 3000, 10000 |
| k | Number of trials | 4, 12, 40 |
| $n_2$ | Size of cohort dataset | 1500, 3000, 6000 |
| ARE | Average effect of treatment offer in P | 1 |
| $ARE_1$ | Average effect of treatment offer in $P_1$ | 0.5, 1, 1.5 |
| | *Note: values chosen to investigate $\zeta$ = -0.5, 0, 0.5* | |
| $\alpha$ | Time treatment effect measured at (years) | 0.25, 0.5, 1 |
| $\pi$ | Proportion of target population trial eligible | 0.5, 0.7, 0.9 |
| $\Gamma_0$ | Mean intercept in P | 20 |
| $\Gamma_{02}$ | Mean intercept in $P_2$ | 20 |
| $\gamma_1$ | Rate of decline t<0 | -2 |
| $\gamma_2$ | Rate of decline amongst patients not offered treatment | if $\alpha$=0.25: -4, -2, 0 |
| | *Note: values chosen to investigate $\varphi_\alpha$ = -0.5, 0, 0.5* | if $\alpha$=0.5: -3, -2, -1 |

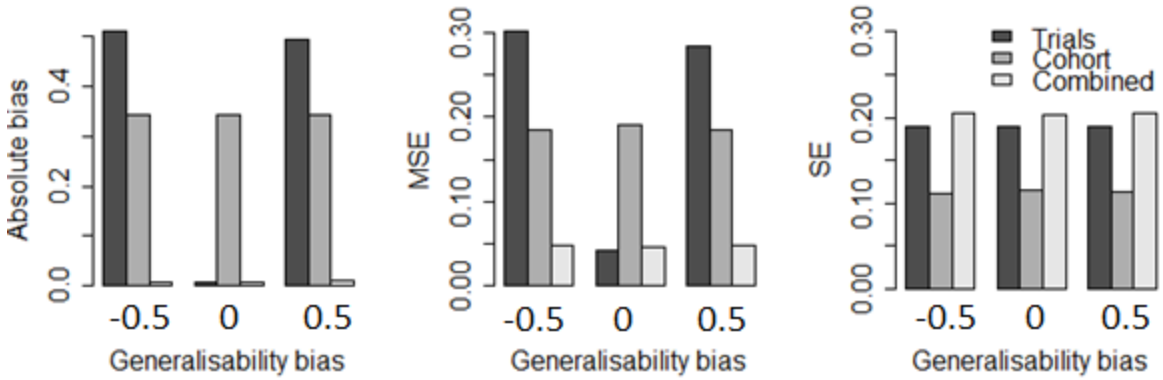
| | | |
|---|---|---|
| | | if α=1: -2.5, -2, -1.5 |
| $\gamma_3$ | Rate of decline when t≥δ in patients offered treatment | -1.9 |
| δ | Second change point (time in years) | 0.3 |
| σ | Residual standard deviation | 2 |
| $\sigma_0$ | Intercept standard deviation | 4 |
| $v^2$ | Total variation in $\Delta_i$ in $P_1$ and in $P_2$ | 1 |
| $\tau_1^2$ | Between trial variation in $P_1$ | 0.25 |

282

283 **3. Results**

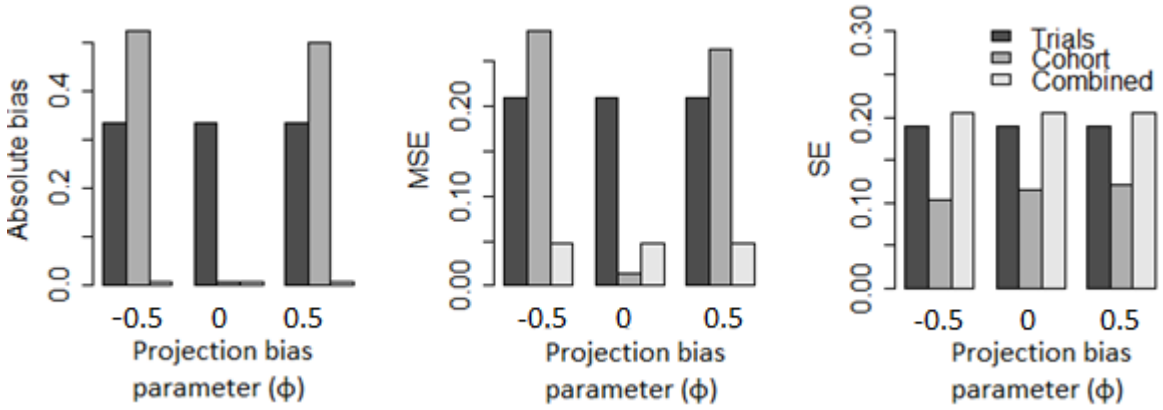
284 ***3.1 Simulation study***

285 The impact of varying the generalisability bias parameter (ζ) at α=0.25 and averaged across all other
286 parameters is summarised in Figure 2a. In the presence of generalisability bias (non-zero ζ) the
287 combined estimator is less biased and has a smaller MSE than the trials only estimator. When there is
288 no generalisability bias (ζ=0), that is when the RCT estimate is representative of the whole population,
289 the combined estimator remains unbiased but the MSE is slightly larger than the trials only estimator.
290 Similarly when varying the projection bias, the combined model has lower bias and MSE than the
291 cohort only estimator in the presence of projection bias and remains unbiased, but with slightly larger
292 MSE when there was no projection bias (see Figure 2b). Variation in the sizes of the two datasets and
293 the trial eligible proportion did not impact on bias estimates, however, increasing the proportion trial
294 eligible or either of the sample sizes led to a reduction in the MSE of the combined estimator. Tables
295 summarising outputs from all combinations of inputs considered in the simulation study are provided
296 in the supplementary material.




297

298 **Figure 2a:** Impact of generalisability bias on performance estimators at α=0.25




299

300 **Figure 2b:** Impact of projection bias on performance of estimators at α=0.25

301 ***3.2 Illustration using real data***

302 Having demonstrated the favourable performance of the combined model in the simulation study, it
303 was applied to the motivating clinical question. The recent systematic review of AChEIs in the
304 management of dementia identified four trials estimating their effects at 12 months after treatment
305 offer (30-33). Five overall eligibility criteria for these trials were established: (i) baseline age between
306 40 and 94 years; (ii) baseline MMSE between 10 and 26; (iii) diagnosed with Alzheimer's or Alzheimer's
307 and cerebrovascular disease; (iv) participant has a reliable/responsible caregiver; and, (v) participant
308 does not have another major psychiatric disorder. Participants in the treated cohort meeting these
309 criteria were identified. Estimates of treatment effect based on the trials only were calculated as part

of the systematic review. Estimates based on the treated cohort only and using the new combined model were calculated. All three estimators suggested a modest but significant effect in favour of treatment (see Table 2). The estimate based on the combined model was lower than those based on either of the two data sources alone, demonstrating that the combined model can quantify both generalisability and projection bias even when they are both in the same direction.

**Table 2:** Comparing estimates of the average effect of the offer of AChEI treatment at 12 months after treatment offer in terms of the MMSE

| | Trials only | Cohort only | Trials and cohort |
|---|---|---|---|
| **ARE** | 1.10 | 1.56 | 0.86 |
| **SE(ARE)** | 0.316 | 0.240 | 0.327 |

## 4. Discussion

In this paper we have proposed and evaluated a novel estimator of treatment effectiveness which incorporates data from RCTs to improve the estimates of treatment effect available from analysis of data on a treated cohort. The performance of the novel estimator was compared with that of those based on either data source alone via a simulation study, demonstrating the model to be superior for estimating effectiveness in the presence of bias in one or both of the data sources. The model was applied to estimate the effectiveness of AChEIs 12 months after treatment offer in terms of MMSE scores, and this highlighted an important strength of the new combined model; namely that even when both of the estimates based on a single data source are biased in the same direction the model can identify and account for these biases.

The combined model used a Bayesian framework, allowing the incorporation of data external to the current dataset in the form of informative prior distributions. This model can be considered an example of the type of bias analysis proposed by Greenland (34). In this combination of data sources,

it is important to account for potential differences in study design and data collection features between the sources. The lack of such a mechanism is one criticism of many of the existing techniques (35). Others have suggested that results could be weighted to account for perceived differences in reliability of data sources (12, 13), adjusting point estimates given anticipated bias (36) or discounting the weight of prior information using power priors (37); however, each of these approaches requires substantial subjective judgement about how to make these adjustments which can present challenges. On the other hand, the approach proposed here provides a method by which informative priors can be applied only to the proportion of the population to which they apply.

One limitation of this method is that it does not address potential differences in distribution between the trial eligible portion of the treated cohort and the trial sample (the requirement for assumption A3 to hold). Approaches to account for these differences (e.g., (38)) could be investigated in future as a possible expansion to the model proposed here which would allow assumption A3 to be relaxed and increase the possible applications for the model. In addition, the model does not currently address the possibility that adherence rates may differ between the trial and cohort populations, instead relying on the fact that these are likely to be similar when the trials in question are pragmatic phase III trials. Future work could address this by incorporating adherence in the model; however, this would require careful definition of adherence in both data sources.

A further limitation which may be encountered in applying such a model is the need to be able to identify patient eligibility in an EMR; however, the availability of such data is increasing (39). Techniques such as natural language processing, constructing variables from constituent parts or the use of proxies may be required; these can increase the time and complexity of fitting this type of model. Similarly whilst all trial reports should include details of eligibility criteria as per the CONSORT Statement (40), the implementation of these guidelines has been mixed and there is still need for further improvements (41, 42). In addition, the new combined model relies on assumption A4;

however, this is weaker than assumption A1 which must be made when estimating treatment effects based on only one type of data.

Assumption A4 is analogous to the one made when calculating the linearly extrapolated estimator of treatment effect in Cross Design Synthesis (15). The model proposed here does, however, have advantages over Cross Design Synthesis, in that it uses a treated cohort from routine data rather than a comparative observational study; such data sources are more readily and more widely available. In addition, the approach proposed here uses both data sources directly in estimating the treatment effect.

The approach has been developed for a continuous outcome measure; however, it could be expanded in future for use with in other clinical settings (e.g., where patients are expected to recover permanently as a result of treatment) or other data types (for example binary or time to event data). This would require careful consideration of reasonable assumptions for outcomes under control conditions in the EMR data and then using these assumptions to derive an appropriate model with a treatment effect parameter on which a prior could be placed. Simulation studies would be required to investigate the performance of such an expansion of the model. The modelling methods could also be applied in other conditions and to estimate the effectiveness of other treatments.

In conclusion, in this paper we have proposed a Bayesian mixed model approach to combining data from trials and a treated cohort to estimate treatment effectiveness and demonstrated using a simulation study the superiority of estimates of effectiveness produced by this model compared to those provided by either data source alone. The new model was also applied to estimate the effectiveness at 12 months after treatment offer of AChEIs in the management of dementia as measured using the MMSE. Several possible avenues for future extensions of this model have also been proposed.

## Ethics approval

The CRIS system received ethical approval from the Oxfordshire Research Ethics Committee C as an anonymised data resource.

## Author contributions

All authors contributed to study conception, study design, interpretation of the results, and reviewed the draft manuscript. RK performed the formal analyses and prepared the first draft of the manuscript.

## Sources of funding

This paper represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London as part of a PhD studentship to RK. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

RS is part-funded by: i) the National Institute for Health Research (NIHR) Biomedical Research Centre at the South London and Maudsley NHS Foundation Trust and King's College London; ii) the Medical Research Council (MRC) HDR UK DATAMIND hub; iii) an NIHR Senior Investigator Award; iv) the National Institute for Health Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust.

## Conflict of interest

RS declares research support received in the last 36 months from Janssen, GSK and Takeda. All other authors declare no competing interests.

## References


1. Smeets HM, Laan W, Engelhard IM, Boks MP, Geerlings MI, de Wit NJ. The psychiatric case register middle Netherlands. BMC Psychiatry. 2011;11(1):106.
2. Stewart R, Davis K. 'Big data'in mental health research: current status and emerging possibilities. Soc Psychiatry Psychiatr Epidemiol. 2016;51(8):1055-72.
3. Dunn G, Emsley R, Liu H, Landau S, Green J, White I, et al. Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: a methodological research programme. 2015.
4. Licht RW, Gouliaev G, Vestergaard P, Frydenberg M. Generalisability of results from randomised drug trials: a trial on antimanic treatment. The British Journal of Psychiatry. 1997;170(3):264-7.
5. Rothwell PM. External validity of randomised controlled trials:"to whom do the results of this trial apply?". The Lancet. 2005;365(9453):82-93.
6. Ades A, Sutton A. Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. J Roy Stat Soc Ser A (Stat Soc). 2006;169(1):5-35.
7. Kaizar EE. Incorporating both randomized and observational data into a single analysis. Annual Review of Statistics and Its Application. 2015;2:49-72.
8. Verde PE, Ohmann C. Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. Research synthesis methods. 2015;6(1):45-62.
9. Eddy DM. The confidence profile method: a Bayesian method for assessing health technologies. Oper Res. 1989;37(2):210-28.
10. Prevost TC, Abrams KR, Jones DR. Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. Stat Med. 2000;19(24):3359-76.
11. Larose DT, Dey DK. Grouped random effects models for Bayesian meta-analysis. Stat Med. 1997;16(16):1817-29.
12. O'Rourke K, Walsh C, Hutchinson M. Outcome of beta-interferon treatment in relapsing-remitting multiple sclerosis: a Bayesian analysis. J Neurol. 2007;254(11):1547-54.
13. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. Stat Methods Med Res. 2001;10(4):277-303.
14. Droitcour J, Silberman G, Chelimsky E. Cross-design synthesis: a new form of meta-analysis for combining results from randomized clinical trials and medical-practice databases. Int J Technol Assess Health Care. 1993;9(3):440-9.
15. Kaizar EE. Estimating treatment effect via simple cross design synthesis. Stat Med. 2011;30(25):2986-3009.
16. Prince M, Comas-Herrera A, Knapp M, Guerchet M, Karagiannidou M. World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future. 2016.
17. Terry AV, Buccafusco JJ. The cholinergic hypothesis of age and Alzheimer's disease-related cognitive deficits: recent challenges and their implications for novel drug development. J Pharmacol Exp Ther. 2003;306(3):821-7.
18. Perera G, Broadbent M, Callard F, Chang C-K, Downs J, Dutta R, et al. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) case register: current status and recent enhancement of an electronic mental health record-derived data resource. BMJ open. 2016;6(3):e008721.
19. Perera G. Predictors of response to acetylcholinesterase inhibitors: an observational case register-based cohort study: King's College London (University of London); 2013.
20. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3):189-98.

21. Knight R, Khondoker M, Magill N, Stewart R, Landau S. A systematic review and meta-analysis of the effectiveness of acetylcholinesterase inhibitors and memantine in treating the cognitive symptoms of dementia. Dement Geriatr Cogn Disord. 2018;45(3-4):131-51.
22. Perera G, Khondoker M, Broadbent M, Breen G, Stewart R. Factors associated with response to acetylcholinesterase inhibition in dementia: a cohort study from a secondary mental health care case register in London. PloS one. 2014;9(11):e109484.
23. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models: Cambridge university press; 2006.
24. Pauler DK, Finkelstein DM. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. Stat Med. 2002;21(24):3897-911.
25. Lunn D, Jackson C, Best N, Spiegelhalter D, Thomas A. The BUGS book: A practical introduction to Bayesian analysis: Chapman and Hall/CRC; 2012.
26. Schlüter P, Deely J, Nicholson A. Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. Journal of the Royal Statistical Society: Series D (The Statistician). 1997;46(3):293-316.
27. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015;16(1):495.
28. Zimmerman M, Chelminski I, Posternak MA. Exclusion criteria used in antidepressant efficacy trials: consistency across studies and representativeness of samples included. The Journal of nervous and mental disease. 2004;192(2):87-94.
29. Wallis WD, George JC. Introduction to combinatorics: Chapman and Hall/CRC; 2016.
30. Hager K, Baseman AS, Nye JS, Brashear HR, Han J, Sano M, et al. Effects of galantamine in a 2-year, randomized, placebo-controlled study in Alzheimer's disease. Neuropsychiatr Dis Treat. 2014;10:391.
31. Karaman Y, Erdoğan F, Köseoğlu E, Turan T, Ersoy AÖ. A 12-month study of the efficacy of rivastigmine in patients with advanced moderate Alzheimer's disease. Dement Geriatr Cogn Disord. 2005;19(1):51-6.
32. Mohs RC, Doody R, Morris J, Ieni J, Rogers S, Perdomo C, et al. A 1-year, placebo-controlled preservation of function survival study of donepezil in AD patients. Neurology. 2001;57(3):481-8.
33. Winblad B, Engedal K, Soininen H, Verhey F, Waldemar G, Wimo A, et al. A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD. Neurology. 2001;57(3):489-95.
34. Greenland S. Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. International journal of epidemiology. 2009;38(6):1662-73.
35. Verde PE, Ohmann C, Morbach S, Icks A. Bayesian evidence synthesis for exploring generalizability of treatment effects: a case study of combining randomized and non-randomized results in diabetes. Stat Med. 2016;35(10):1654-75.
36. O'rourke K, Walsh C. Impact of stroke units on mortality: a Bayesian analysis. European journal of neurology. 2010;17(2):247-51.
37. Ibrahim JG, Chen M-H. Power prior distributions for regression models. Statistical Science. 2000;15(1):46-60.
38. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. American journal of epidemiology. 2010;172(1):107-15.
39. Morgan VA, Jablensky AV. From inventory to benchmark: quality of psychiatric case registers in research. The British Journal of Psychiatry. 2010;197(1):8-10.
40. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. Ann Intern Med. 2011;154(4):291-2.
41. Hopewell S, Dutton S, Yu L-M, Chan A-W, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ. 2010;340:c723.

42. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. Cochrane Database Syst Rev. 2012(11).