

RESEARCH ARTICLE

Using the Bayesian credible subgroups method to identify populations benefiting from treatment: An application to the Look AHEAD trial

Anna Coonan^{1*}, Patrick Schnell², Joel Smith³, John Forbes¹

1 Graduate Entry Medical School, The University of Limerick, Limerick, Ireland, **2** Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH, United States of America, **3** Nuffield Department of Population Health, The University of Oxford, Oxford, United Kingdom

* Anna.Coonan@gmit.ie



OPEN ACCESS

Citation: Coonan A, Schnell P, Smith J, Forbes J (2020) Using the Bayesian credible subgroups method to identify populations benefiting from treatment: An application to the Look AHEAD trial. PLoS ONE 15(4): e0231241. <https://doi.org/10.1371/journal.pone.0231241>

Editor: Rebecca A. Krukowski, University of Tennessee Health Science Center, UNITED STATES

Received: June 11, 2019

Accepted: March 18, 2020

Published: April 21, 2020

Copyright: © 2020 Coonan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Legal requirements prohibit public sharing of the dataset. Data are available upon request from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Interested researchers can find information and submit a formal data request at https://repository.niddk.nih.gov/pages/overall_instructions/.

Funding: This work was supported by the Health Research Board HRB-RL-2013-11 and the National Institute for Health Research Biomedical Research

Abstract

Traditionally, subgroup analyses are used to assess whether patient characteristics moderate treatment effectiveness with general disregard for issues of multiplicity. Using data from The Action for Health in Diabetes (Look AHEAD) trial in the United States, we aim to identify a subgroup where all of its types of members experience a treatment benefit defined as reducing the likelihood of a major cardiovascular event under an intensive lifestyle and weight-loss intervention. We apply the credible subgroups method to a Bayesian logistic model with a conservative prior that is sceptical of large treatment effect heterogeneity. The covariate profiles for which there is sufficient evidence of treatment benefit are, coarsely, middle-aged women, in poor subjective general health and with moderately to poorly controlled diabetes. There is at least 80% posterior probability that the conditional average treatment effect is positive for all covariate profiles fitting this description, which account for 0.5% of trial participants. Conversely, the covariate profiles that are likely to be associated with no benefit are middle aged and older men in excellent subjective general health, with well-controlled diabetes. These profiles apply to less than 2% of trial participants. More information is required to determine treatment benefit or no benefit for the remainder of the trial population.

1. Introduction

In general, healthcare providers aim to use information about a person to prevent, diagnose and treat disease, or to promote wellness. This relies on the ability to translate results from trials into clinical practice where the quantity of interest is usually the average treatment effect. The experimental approach to the estimation of average treatment effects involves random assignment of treatment followed by a comparison of the average outcomes of treated to non-treated individuals who serve as controls [1]. However, average treatment effects only provide

Centre at the University of Oxford NIHR-BRC-1215-20008. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

a partial explanation of treatment effects particularly when they vary across experimental units.

In randomised controlled trials, average treatment effects tend to be generalised to a wider population and as such the results will be taken to apply equally to all patients in the trial. Participants, however, will vary significantly in baseline predictive and prognostic characteristics, and therefore each patient can report a unique effect with respect to a given treatment. This treatment effect heterogeneity can be masked by the calculation of average effects, leaving potential benefitters and non-benefitters of a given treatment unidentified. Producing results on population averages without recognizing the underlying heterogeneity could lead to welfare losses and increases in healthcare inefficiencies and costs [2, 3].

The standard approach for investigating treatment effect heterogeneity is to estimate conditional average treatment effects, or treatment effects among homogeneous groupings of experimental units. These homogenous subgroups are defined by baseline covariates. In order to assess how causal effects vary across subgroups, researchers predict conditional average treatment effects which are average effects conditioned on certain values of observed covariates for which treatment effectiveness varies [4]. Conditional average treatment effects are particularly useful when the researcher wants to explore the heterogeneity of treatment effects in various subgroups defined by a single covariate.

Estimating heterogeneous treatment effects, or systematically different outcomes associated with treatment due to moderating patient characteristics, is typically relegated to subgroup analyses that rely on simple treatment-covariate interaction terms [5]. In clinical trials, however, researchers typically aren't concerned with just one subgroup but rather subgroups across a series of standard covariates such as age, ethnicity, and sex. Multiple interaction tests can be misleading as they are often underpowered, suffer from multiplicity errors and do not allow for simultaneous inference [6, 7]. Moreover, subgroups tend to be prespecified to avoid perceptions of data-mining or overfitting the data and therefore restrict the ability of the researcher to test new unforeseen hypotheses [8].

There is growing recognition in the field of health economics that more nuanced estimates of patient-specific heterogeneous treatment effects can be estimated and conditioned on a variety of observable and unobservable risk factors. Some methods are derived from the computer science discipline such as machine learning methods where the focus is on predicting patient-specific treatment effects [9–11]. Other methods build on a Bayesian framework where the parameter of interest is unknown but the model allows the researcher to vary the degree of expected treatment effect heterogeneity through prior selection [6, 12, 13]. All of these methods provide insight into the reliability of incorporating heterogeneity in treatment response.

Our work builds on a Bayesian framework to classify personalised treatment effects, or average outcomes under treatment and control for every covariate point. Personalised treatment effects are used to classify covariate points into credible subgroups that allow for inferences about the types of patients expected to benefit or not benefit from an intervention, while accounting for the multiplicity arising from testing a hypothesis at every covariate point. This study is the first major application of the Bayesian credible subgroups method to such a large clinical dataset. Previous research and our own are discussed in the context of a large multicentre randomised controlled trial called the Action for Health in Diabetes (Look AHEAD). In the Look AHEAD trial, insignificant treatment effects for the whole sample and all prespecified subgroups were reported, most likely resulting from latent patient heterogeneity [14].

In 2001 the Look AHEAD trial set out to model the long-term effects of an intensive lifestyle intervention (ILI), one specifically aimed at weight loss on cardiovascular morbidity and mortality among patients with type 2 diabetes who were also overweight or obese [14, 15]. The lifestyle intervention included the following: individual and group counselling sessions on

diabetes control, aids for calorie management with the option of meal replacement supplements, and supports to increase physical activity. Patients in the control group received diabetes support and education (DSE) counselling only. Prior to the Look AHEAD trial, the effects of lifestyle interventions on cardiovascular morbidity and mortality were mixed at best. Some studies reported no cardiovascular effect and others reported minimal benefits after two decades of follow-up [16–19]. The effects of weight loss on reduced cardiovascular risk factors in the short-run were well-documented, but the Look AHEAD trial set out to identify if significant effects translated to reductions in a primary composite cardiovascular disease (CVD) endpoint over a long-term follow-up. In September 2012 the Look AHEAD trial prematurely ended after a futility analysis concluded the weight loss intervention did not reduce the rate of primary cardiovascular events [14]. The median follow-up was 9.6 years. Data are available upon request from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). All participants in Look AHEAD provided written consent to allow their information to be used in future research.

A possible explanation for a lack of statistical significance in Look AHEAD was the presence of differing effects for different individuals or patient heterogeneity [14]. Secondary analyses were conducted to see if general health benefits resulted from the lifestyle intervention. The documented health benefits range from weight loss, increased physical fitness, better diabetes control, and reduced urinary incontinence [20–22]. A small number of exploratory studies have examined the original CVD endpoint finding positive treatment effects for subgroups defined by subjective general health and diabetes control [23] and percentage weight loss in the first twelve months [20]. Using different methods, these were all attempts to model treatment heterogeneity in Look AHEAD.

Behavioural responses and treatment uptake are notoriously heterogeneous with varying degrees of weight loss being reported in the first year [24]. In the instance of the Look AHEAD trial it is likely that patient heterogeneity was able to impact treatment effectiveness at two entry points: the effect of the intervention on behaviour and uptake, and the effect of behaviour and weight loss on the primary CVD outcome [25]. In Baum et.al. [23] a machine learning algorithm based on decision-trees was used to identify two baseline patient characteristics believed to be instigators behind the trial-wide outcome neutral effects. One covariate was HbA_{1c}, or glycated haemoglobin, which is a measure of diabetes control. The second instigator of treatment effect heterogeneity was a subjective general health indicator derived from the SF-36 questionnaire. These covariates were determined to be the most important variables for defining the range of patient heterogeneity. Defined by HbA_{1c} and subjective general health, two subgroups with positive average treatment effects were identified. However, the HbA_{1c} and subjective general health (SF-36) compositions within these two subgroups were contradictory. One subgroup reported a significant absolute risk reduction defined by the baseline covariates excellent subjective general health ($\text{SF-36} \geq 48$) and well-controlled diabetes ($\text{HbA}_{1c} < 6.8\%$). After participants with an expected adverse treatment response were removed from the analysis, a second subgroup with significant risk reductions was defined by the baseline covariates $\text{HbA}_{1c} \geq 6.8\%$ or both $\text{HbA}_{1c} < 6.8\%$ and $\text{SF-36} \geq 48$. These results led to conclusions stating 85% of Look AHEAD participants averted a cardiovascular event which would be a large overestimation of treatment benefit compared to previous Look AHEAD studies [14, 20, 25, 26].

Machine learning methods, like those used in Baum et.al. [23], are useful for detecting and analysing heterogeneity, but they do not directly address the question of primary interest to the current study: specifically whom does treatment benefit? Previously-applied machine learning methods do not adjust for multiplicity or provide simultaneous inferences, defined as a situation where all types of individuals within a subpopulation satisfy a common statement

[6]. The most relevant statement here is that all covariate points in the subgroup correspond to a positive personalised treatment effect. We use a Bayesian credible subgroups method for simultaneous inference that through careful choice of priors on treatment-covariate interactions, allows us to tune our *a priori* belief in the strength of treatment effect heterogeneity. Our model includes a conservative prior that is sceptical of large treatment effect heterogeneity but still allows for strong evidence from the posterior distribution to overwhelm the prior information.

The inferential goal of this analysis is different from previous secondary Look AHEAD studies that confirm the overall treatment effect is neutral and any subsequently identified subgroups are defined by estimated average treatment effects. These estimates continue to apply assumptions of treatment effect homogeneity within all levels of selected covariates. Instead, we aim to capitalise on inherent patient heterogeneity in order to identify the subgroup consisting of only types of patients expected to benefit, not a subgroup of any types patients that together yield a positive average effect. We explore the mechanisms behind the statistical insignificance of the lifestyle intervention. This is done by examining the possibility of a treatment-benefiting subgroup defined by the baseline covariates sex, age, blood glucose percentage (HbA_{1c}) and a measure of subjective general health (SF-36).

This article is structured as follows: In Section 2 we present the methodology and selection of priors. Section 3 describes the data from the Look AHEAD and defines the descriptive statistics. Section 4 presents the results from the credible subgroups analysis with varying credible levels. Section 5 concludes with a discussion and suggestions for future research.

2. Methodology

Data provided by the NIDDK repository was fully anonymised before we accessed them. Our research received ethical approval from University of Limerick application number 2017_12_19_EHS. All participants in Look AHEAD provided written consent to allow their information to be used in future research.

This analysis uses the Bayesian credible subgroups method to construct two bounding subgroups for a benefiting subgroup B that is unknown. The benefiting subgroup is defined to consist of all covariate profiles with a conditional average treatment effect above a threshold set to zero ($\delta = 0$), intended to capture all positive treatment effects regardless of magnitude. Inferences about the subgroup B are made by constructing bounding subgroups, D and S, similar to the procedure for making inferences about a scalar quantity by constructing interval estimates. The bounding subgroups are constructed with the goal that one, the exclusive credible subgroup (D), contains only covariate profiles with a positive treatment effect. The other, the inclusive credible subgroup (S), is intended to contain all covariate profiles with a positive treatment effect. In the construction of these bounds, treatment effects are estimated and tested at all points throughout the covariate space, thereby introducing multiple testing issues. However, by estimating a subgroup where all of its members satisfy some criteria, such as having a treatment effect above some threshold of significance, this statement of simultaneity accounts for multiplicity due to performing a hypothesis test at each covariate point. The result is a decreased risk of conducting a family wise type I error.

The method of credible subgroups is a two-step regression classification method. The first step involves the estimation of a generalised linear with either linear or nonlinear specifications. The second step constructs a simultaneous credible band of some level around the regression surface. Covariate points with a lower bound above some threshold form the exclusive credible subgroup (D), and those where the upper bound is greater than the threshold become the inclusive credible subgroup (S). As an example of the first step, estimation of a

standard linear model includes predictive and prognostic covariates generally described as follows:

$$E[Y_i|x_i, z_i, t_i] = x_i'\beta + t_i z_i'\gamma \quad (1)$$

Y is the response, x and z are vectors of prognostic and predictive covariates where predictive covariates are those interacted with treatment allocation. β and γ are the effect parameters and $t \in \{0,1\}$ the treatment allocation. Additionally, the model includes a conservative prior on γ that represents the belief that treatment-covariate interactions are unlikely by shrinking coefficients towards zero, but that still allows for strong evidence of treatment effect heterogeneity from the posterior distribution to overrule this. In instances where the response is binary, Eq 1 is estimated with a logistic link as follows:

$$\text{logit}^{-1}(p_i) \quad (2)$$

$$(p_i) = \text{Prob}[Y_i = 1|x_i, t_i, z_i]$$

The second step, classification, comes from the posterior distribution of patients' personalised treatment effects. These are used to classify all combinations of covariates in the covariate space according to a credible subgroup pair: the exclusive credible subgroup (D) and the inclusive credible subgroup (S). Personalised treatment effects provide a clinically useful interpretation of the expected difference in outcomes for patients that share the same baseline characteristics.

In our analysis the covariate space is comprised of age, sex, subjective general health measured by the SF-36 general health questionnaire, and diabetes control measured a glycated haemoglobin percentage (HbA_{1c}). All of the covariates included in this paper are standard medical and demographic variables included in studies examining behavioural treatments for type 2 diabetes and the confounding effects of obesity [14, 17, 18, 20, 23, 27]. The model of the personalised treatment effect is the difference in the expectation of the outcome (y) conditional on the predictive covariates (z), as follows:

$$\Delta(z) \equiv E[Y|x, z; t = 1] - E[Y|x, z; t = 0] \quad (3)$$

From the posterior distribution of personalised treatment effects the covariate space is divided into three parts based on a decision rule used to define the credible subgroup pair (D,S). This is comprised of the covariate profiles with lower bounds for treatment effects above some threshold ($\delta > 0$) that define subgroup D, those in a complement subgroup S^c with upper bounds for treatment effects at most equal to the threshold ($\delta = 0$) and an uncertainty region $S \setminus D$ (S remove D) where more information is required for classification. The trichotomy of the covariate space is illustrated in Fig 1. The defining property of the credible subgroups is as follows:

$$P(D \subseteq B \subseteq S|\text{data}) \geq 1 - \alpha, \quad (4)$$

where D and S are subsets in the covariate space with a bound on the posterior probability that these two sets bracket the benefiting subset B. Formally, D is the exclusive credible subgroup because the posterior probability that D contains only z for which $\Delta(z) > \delta$, and simultaneously, S is the inclusive subgroup that contains all z such that $\Delta(z) > \delta$, is at least $1 - \alpha$. The credible level of the subgroup pair ($1 - \alpha$) is also an approximate lower bound on the frequentist probability that $D \subseteq B \subseteq S$. In this study, D is the reported estimate with posterior probability $1 - \alpha$ that the reported subset characterises exclusively covariate combinations with an expected treatment effect above the threshold ($\delta = 0$).

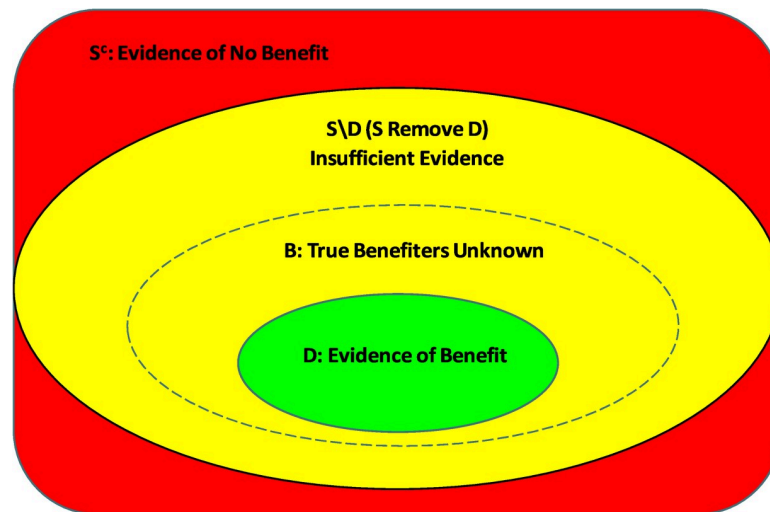


Fig 1. Trichotomy of predictive covariate space. Visualisation of the trichotomy of the covariate space based on the credible subgroup pair (D,S) relative to the true benefiting subgroup (B), adapted from Schnell et.al. [6].

<https://doi.org/10.1371/journal.pone.0231241.g001>

The objective of this analysis is to identify a credible subgroup pair such that, at a credible level of at least 80%, D represents a subgroup that contains *only* types of patients with a positive treatment effect and S represents a subgroup that contains *all* types of patients with a positive treatment effect. Defining an exclusive credible subgroup (D) where all members have a treatment effect above some threshold satisfies properties of simultaneous inference and adjusts for multiplicity [6]. The credible subgroups method allows for inferences that state it is with high probability that the personalised treatment effect ($\Delta(z)$) is positive at all covariate points (z) in the exclusive credible subgroup (D).

We fit a Bayesian logistic regression model with the endpoint likelihood $Y_i | \eta_i \sim \text{Bernoulli}(\text{logit}^{-1}(\eta_i))$ where $\eta_i = x_i' \beta + t_i z_i' \gamma$. We use the vague prior $\beta \sim \text{Normal}(0, 10^3)$ for prognostic effects and a conservative prior $\gamma \sim \text{Normal}(0, 1)$ for predictive effects. In other words, the prognostic effect variances are set to 1000 and those corresponding to predictive effect variances set to 1. A vague prior is selected for prognostic effects preserving the relationship between baseline covariates and the primary endpoint in Look AHEAD; a binary composite variable measuring the presence or absence of at least one CVD outcome. A conservative prior is selected for predictive effects where patient heterogeneity may moderate or mediate the effect of a given treatment. This *a priori* selection reflects the belief that a large amount of treatment heterogeneity is unlikely but still allows for strong evidence from the posterior distribution to overwhelm this belief.

The benefiting subgroup is defined as one for which the ILI treatment is superior to the diabetes support and education (DSE) control. This is in line with current clinical recommendations and the belief that physical activity and controlled caloric intake is beneficial for patients with type 2 diabetes who may be overweight or obese. We set the threshold of a log odds ratio of primary CVD occurrence to $-\delta_1 > 0$. The sign is switched for δ because we want reductions in risk of the primary CVD endpoint. Due to the insignificant results reported in the original trial, we wish to capture treatment benefits of any magnitude in order to avoid automatically eliminating covariate profiles that may not achieve a pre-defined level of interest. The model is fit using the NIMBLE package [28] in the R software [29] with 10,000 Gibbs sampler iterations after 1000 burn-in iterations. Convergence appears to be mixing well and near-immediate from autocorrelation plots for Markov Chains. Parameter autocorrelation plots are illustrated

in the S3 Fig. Our analysis is conducted using only complete cases eliminating 8 observations which renders a sample size of 4893.

3. Data

Look AHEAD was a multicentre randomised controlled trial that tested the effect of an intensive lifestyle intervention focused on diet and exercise on long-term cardiovascular disease outcomes among patients with type 2 diabetes who were overweight or obese [14, 15]. Between August 2001 and April 2004 the trial recruited 5145 participants to undergo randomisation. 2570 were assigned to the intensive lifestyle intervention (ILI) group and 2575 were assigned to the diabetes support and education control (DSE) group. The intervention was modelled on group behaviour programs tailored for obese and overweight patients through individual and group counselling sessions. These sessions continued for an initial 6 months and decreased in frequency throughout a 12-month follow-up. Intervention design was tailored for a weight-loss goal of losing 10% initial body weight in the first year and a total weight-loss of 7% (or more) initial body weight by trial's end. This would be achieved by limiting caloric intake with the option of meal replacement supplements and increasing physical fitness through weekly exercise regimes. The diabetes support and education (DSE) control group provided nutrition and exercise training for a period of four years after randomisation.

The primary outcome was a composite indicator of CVD morbidity and mortality consisting of CVD death, non-fatal myocardial infarction (MI), non-fatal stroke, and hospitalization for angina [14]. The number of CVD events was a concern throughout the trial resulting in the fourth CVD indicator, hospitalisation from angina, being added to the primary outcome two years into the trial [30]. The event rate of the CVD endpoint ended up being less than half the projected rate in the DSE control group, resulting in a very low overall rate of events [14]. Insignificant treatment effects were reported for secondary CVD composite measures and for all prespecified subgroups including age, sex and ethnicity.

Table 1 presents the descriptive statistics for patients in the treated and control groups of the remaining 4901 trial participants after 244 were omitted early in the trial for consent reasons. The characteristics of the patients in the two groups were similar at baseline. The ILI treatment group was on average 58.8 years old, had a BMI of 35.9, weighed 100.9kg and was 41% men. The DSE control group was on average 59.1 years old, had a BMI of 36.0, weighed

Table 1. Characteristics of patients at baseline.

	Intensive Lifestyle Intervention (ILI) (n = 2448)	Diabetes Support and Education (DSE) (n = 2453)	Entire Trial Sample (n = 4901)
Age (years)	58.8 (6.7)	59.1 (6.8)	58.9 (6.8)
Male	1014 (41%)	1016 (41%)	2030 (41%)
History of cardiovascular disease	355 (15%)	335 (14%)	690 (14%)
Diabetes duration (years)	6.7 (6.5)	6.7 (6.3)	6.7 (6.4)
Insulin use	362 (15%)	385 (16%)	747 (15%)
Weight (kg)	100.9 (19.6)	101.2 (18.8)	101.1 (19.2)
BMI (kg/m ²)	35.9 (6.0)	36.0 (5.7)	36.0 (5.9)
HbA _{1c} (%)	7.2 (1.1)	7.3 (1.2)	7.3 (1.2)
HbA _{1c} (> = 6.5%)	1798 (73%)	1867 (76%)	3665 (75%)
Systolic blood pressure (mm Hg)	128.5 (17.2)	129.8 (17.1)	129.1 (17.2)
SF-36 General Health (GH) Questionnaire	47.2 (9.1)	47.3 (8.7)	47.2 (8.9)

Data are in mean (SD) or n (%)

<https://doi.org/10.1371/journal.pone.0231241.t001>

101.2kg and was 41% men. For the entire sample 75% of patients had moderately or poorly controlled diabetes defined as having an HbA_{1c} percentage greater than or equal to 6.5.

4. Results

The credible subgroups method is used to divide the covariate space into a trichotomy which will identify the profiles of patients expected to benefit from treatment, those expected not to benefit from treatment and those for whom more information is needed. The Look AHEAD trial is used in our analysis and includes an experimental intensive lifestyle intervention (ILI) versus a diabetes support and education (DSE) control. The covariates are believed to be both prognostic and predictive and are defined by baseline measurements for subjective general health measured by the SF-36 questionnaire, age, sex and diabetes control (HbA_{1c}). The end-point of interest is a binary composite indicator measuring the presence of four possible macro cardiovascular events.

The first stage of the credible subgroups method is the estimation of a Bayesian logistic regression with vague priors on prognostic effects and conservative priors on predictive effects. From this stage our objective is to discuss treatment-covariate interactions at the nominal credible level of 95% for which standard subgroup analyses were not powered to detect. The model is fit with 10,000 Gibbs sampler iterations after 1000 burn-in iterations. The posterior mean, standard deviations and credible bounds of effect parameters are reported in Table 2. Significant effects are reported for the treatment-control interaction and individual covariates general health and HbA_{1c}. These results are not adjusted for multiplicity and are considered incomplete because a lack of significance on the age and sex interaction terms does not imply a homogeneous treatment effect across all levels of these covariates. The estimates in Table 2 do not identify the covariate profiles of patients explicitly expected to benefit and therefore do not directly address our research question. The interpretation that we wish to avoid is that the treatment only interacts with general health and HbA_{1c} and that any potential treatment effect exists across all levels of the other covariates.

Fig 2 presents the covariate space which constitutes the exclusive credible subgroup D at the 80% credible level. Credible subgroup membership was determined using the credsubs package [31]. This analysis uses a Bayesian logistic model to estimate posterior mean treatment effects from endpoint likelihoods that follow a Bernoulli distribution. Ellipsoidal contours of the standard errors are centred on a majority of the data points which induce the curved boundaries of the credible subgroup [6]. The covariate profiles classified to the exclusive

Table 2. Posterior summaries of Bayesian logistic parameters.

	Posterior Mean	Posterior Standard Deviation	95% Credible Bounds
Intercept	-2.41*	0.14	-2.18 to -2.63
General Health	-0.07	0.06	0.03 to -0.17
Age	0.43*	0.06	0.34 to 0.53
Sex	0.82*	0.11	0.64 to 0.99
HbA _{1c}	0.39*	0.13	0.17 to 0.61
Treatment	-0.22	0.18	-0.52 to 0.07
General Health × Treatment	-0.19*	0.08	-0.05 to -0.32
Age × Treatment	-0.01	0.08	-0.15 to 0.13
Sex × Treatment	-0.07	0.16	-0.33 to 0.18
HbA _{1c} × Treatment	0.32*	0.18	0.02 to 0.61

Continuous covariates are standardized. Estimates identified as significant according to a 95% credible level are denoted by *.

<https://doi.org/10.1371/journal.pone.0231241.t002>

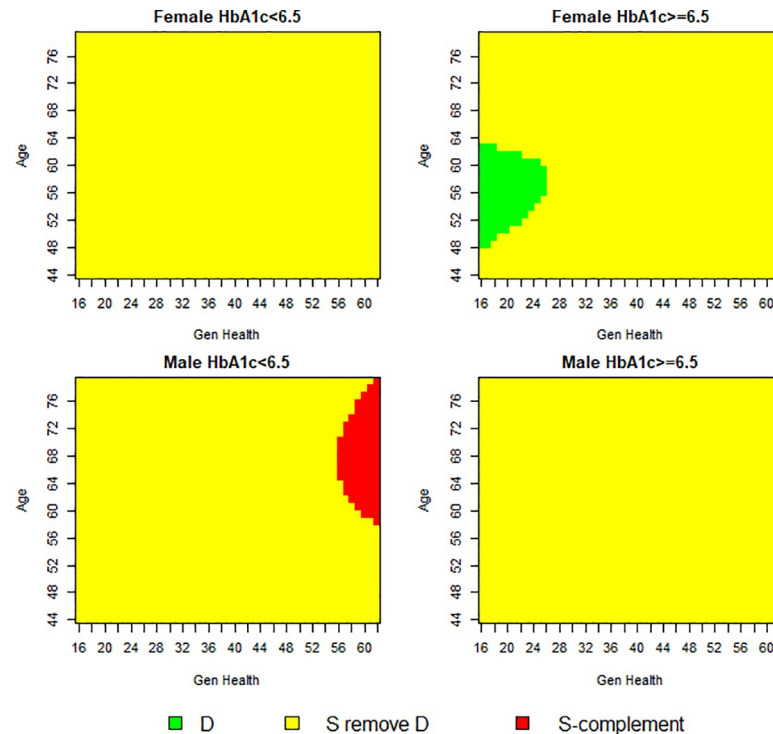


Fig 2. Trichotomy of baseline covariates at 80% credible level. Individual credible subgroup pairs plotted over the covariate space. There is at least 80% posterior probability that all patients with covariate points in D have a positive conditional average treatment effect ($\delta > 0$).

<https://doi.org/10.1371/journal.pone.0231241.g002>

credible subgroup D are middle-aged women in poor subjective general health with moderately to poorly controlled diabetes. There is at least 80% posterior probability that the conditional average treatment effect is beneficial ($\delta < 0$) for all types of patients with covariate points in D, accounting for multiplicity due to testing a hypothesis for each covariate point, and thus promoting the lifestyle intervention as an appropriate treatment for this subgroup. Conversely, the S-complement subgroup identifies the covariate profiles of those who may not benefit, represented by either a null or adverse treatment effect ($\delta \geq 0$). The covariate profiles of patients in this region are middle-aged and older men, in excellent subjective general health with well-controlled diabetes. Therefore, the lifestyle intervention is not to be considered an appropriate treatment for members of this subgroup.

As a comparison a figure of similar threshold and credible level with a vague prior on predictive and prognostic effects is included in the [S1 Fig](#). This figure supports the conclusions presented in [Table 2](#).

The funnel plot in [Fig 3](#) shows the predicted treatment effects across the entire covariate space and their coloured classification according to the credible subgroups pair (D,S). A majority of covariate space is classified as the $S \setminus D$ region where more information is required to classify points as being members of the exclusive credible subgroup. The funnel plot displays the proportion of covariates that are members of the S-complement subgroup and have a treatment effect at most equal to zero. This graph identifies the minority of covariate profiles with sufficient evidence of an expected treatment benefit represented by the green area. The minority of points classified to the exclusive credible subgroup supports the idea that the heterogeneity of patients with these covariate profiles could have been lost in the original trial-wide average estimates. When applying the trichotomy of the covariate space to 4893 observations

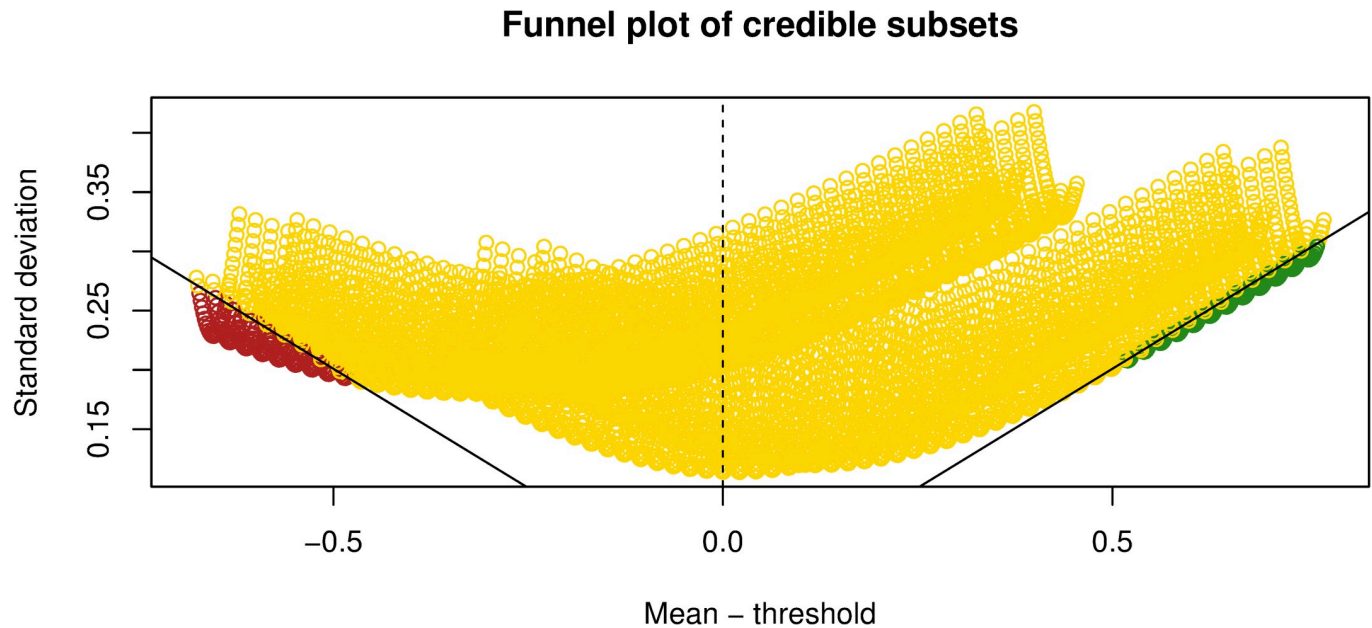


Fig 3. Exclusive (green) and inclusive (yellow) credible subsets at 80% credible level. Funnel plot of predicted treatment effects covering entire covariate space and credible subgroup classification (D,S).

<https://doi.org/10.1371/journal.pone.0231241.g003>

in the Look AHEAD trial, there is sufficient evidence of benefit for less than 0.5% of the Look AHEAD population. Likewise, there is sufficient evidence of an expected null or adverse effect for less than 2% of the trial population.

Fig 4 shows the contours of the maximum credible level that allows assignment of a covariate point to the exclusive credible subgroup. Contours with negative values indicate assignment to the complement of the inclusive credible subgroup at the credible level of the modulus. The 80% contour aligns with the previously identified boundary of the exclusive credible subgroup shown in Fig 2. In selecting a credible level there is a balance to be struck between estimating credible subgroups that minimally misclassify a covariate profile and choosing a credible level that is too conservative to render clinical application. Without specification from regulatory agencies, combined with the fact that the intervention is a behavioural lifestyle intervention, this balance must be weighed by individual healthcare professionals. This figure identifies the highest level of confidence to be 80% which renders a nonempty credible subgroup. As the maximum credible level is relaxed, a greater portion of the covariate space is classified to either the exclusive credible subgroup or the complement of the inclusive credible subgroup.

In Fig 5 we set the treatment effect threshold to zero but relax the credible levels from 80% to 50%. A figure illustrating the trichotomy of the covariate space at a 65% credible level is included in the S2 Fig. As expected, when the credible level is relaxed, the proportion of covariate points allocated to the exclusive credible subgroup (D) and S-complement subgroup (S^c) increase. As the credible level is relaxed, the covariate composition of the exclusive credible subgroup widens to include both men and women for a wider range of ages. The two most important covariates for differentiating patient benefiter from non-benefiter are having a low HbA_{1c} percentage and having good subjective general health. The similarity between these two covariates is to be expected as good control over one's diabetes, which is a physical health indicator, could translate to patients' positive subjective evaluations of their own health. This conclusion supports previous research that has found HbA_{1c} and subjective general health

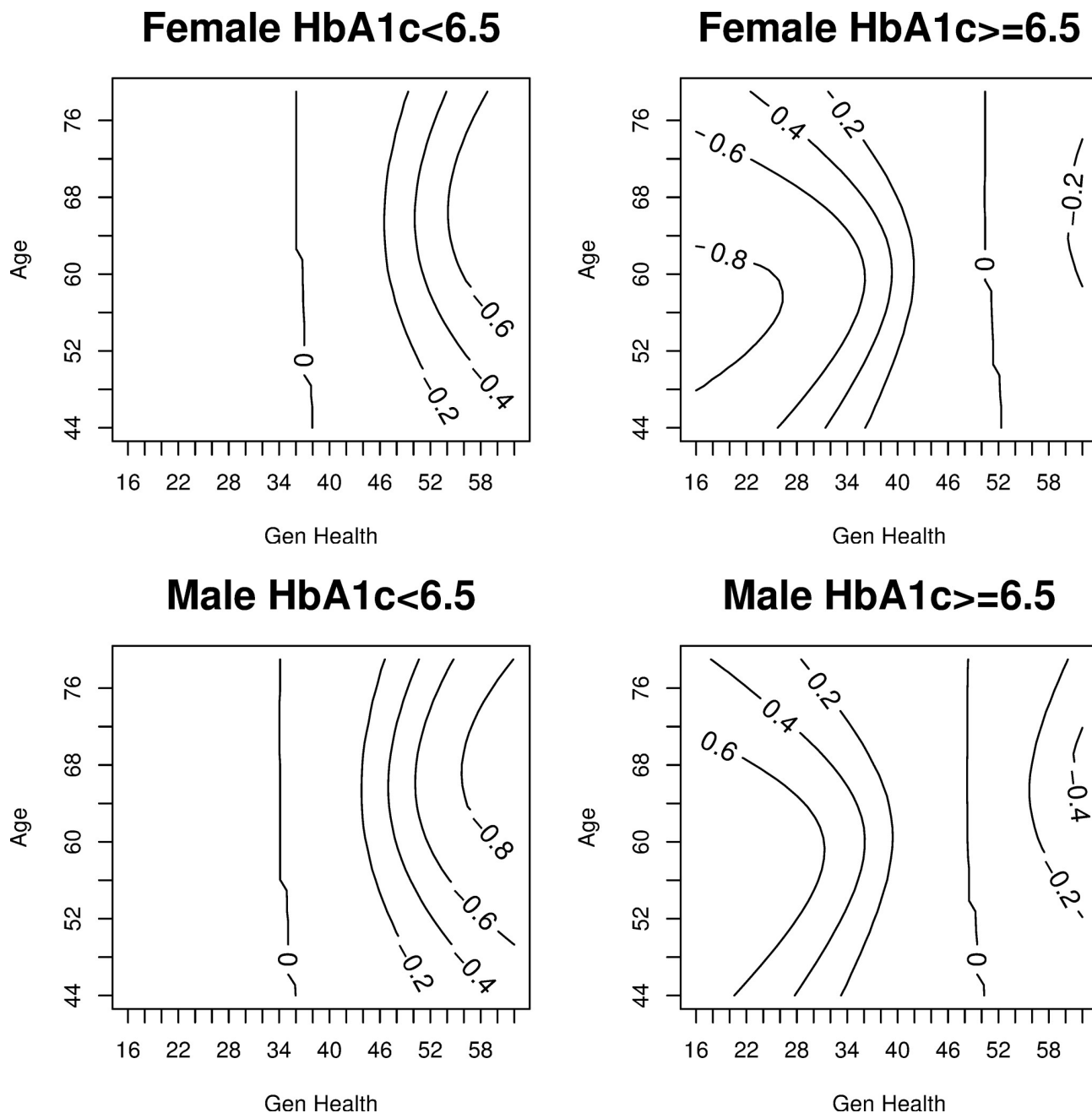


Fig 4. Maximum credible level contours for exclusive credible subgroup inclusion Credible levels that yield nonempty credible subgroups (D,S).

<https://doi.org/10.1371/journal.pone.0231241.g004>

(SF-36) are two of the most important covariates for determining treatment effect heterogeneity in the Look AHEAD trial [23]. The HbA_{1c} composition of the exclusive and inclusive credible subgroups is largely consistent across sex, age and general health covariates.

At the 80% credible level, the application of the covariate profiles of the exclusive credible subgroup, suggests evidence of treatment benefit for less than 0.5% of the Look AHEAD population. Applying the covariates that constitute the S-complement subgroup suggests evidence of a null or adverse treatment effect for 2% of the Look AHEAD population. At the more lenient 50% credible level, the application of the covariate profiles of the exclusive credible

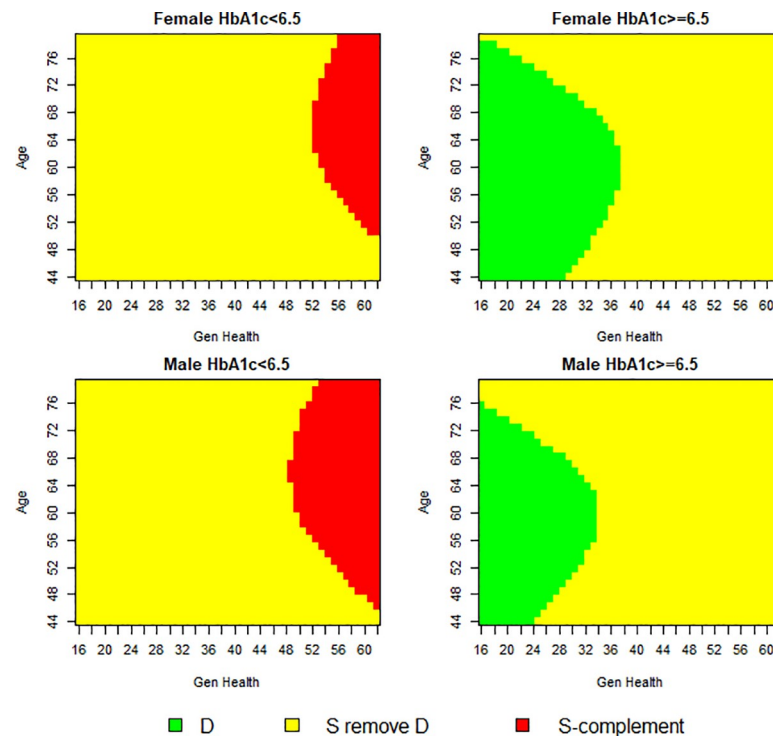


Fig 5. Trichotomy of baseline covariates at 50% credible level. Individual credible subgroup pairs plotted over the covariate space. There is at least 50% posterior probability that all patients with covariate points in D have a positive conditional average treatment effect treatment effect ($\delta = 0$).

<https://doi.org/10.1371/journal.pone.0231241.g005>

subgroup would render a subpopulation of less than 10% of the Look AHEAD trial. While the true benefiting population may be larger, the expansive inconclusive region across two credible levels does not provide supporting evidence of this. Our results across all credible levels differ from previous controversial findings that suggested as much as 85% of the Look AHEAD population benefitted from the ILI based on positive average estimates [23].

5. Discussion

In this analysis of the Look AHEAD trial, we identified a subgroup that benefited from an intense behavioural lifestyle intervention in terms of reduced cardiovascular risk. The aim of the trial was to model the effects of a lifestyle intervention that focused on weight-loss and physical activity on the long-term cardiovascular outcomes of patients with type 2 diabetes who were overweight or obese. After the original trial reported no between group differences in the occurrence of the primary CVD endpoint for patients randomised to the treatment or control, commentators suggested this was the result of latent patient heterogeneity. Recent developments in treatment effect estimation call for a move away from *ad hoc* subgroup analyses and encourage the development of new methodologies designed to capture the richness that patient heterogeneity provides. By fitting a Bayesian logistic regression model, we were able to include prior information of the degree of treatment effect heterogeneity and still allow for strong evidence from the posterior distribution to overrule any assumptions. Using the credible subgroups method, we classified covariate profiles according to a credible subgroup pair that allowed for the identification of a subgroup in which all types of patients exhibit a positive treatment effect. This subgroup was coarsely comprised of the following covariate

profiles: middle-aged women in poor subjective general health with moderately to poorly controlled diabetes ($HbA_{1c} \geq 6.5\%$). There exists at least 80% posterior probability that the treatment effect is positive for patients with these covariate profiles located within the exclusive credible subgroup (D), accounting for multiplicity due to testing a hypothesis for each covariate point. A second subgroup was identified as the S-complement where its members exhibited a null or adverse treatment effect. The covariate profiles of this group were middle-aged and older men, in excellent subjective general health with well-controlled diabetes ($HbA_{1c} < 6.5\%$). Applying the covariate profiles of the exclusive credible subgroup to the Look AHEAD participants rendered an eligible subpopulation of less than 0.5%. Members of this subpopulation would be expected to experience a positive treatment effect.

The covariate composition of the exclusive credible subgroup is consistent with studies from similar trials that reported women and those in poor health experienced the greatest reduction in CVD mortality after participating in a weight loss intervention [17, 32, 33]. Additionally, the identification of a S-complement group supports previous findings that a selection of patients can have null or adverse cardiovascular responses to weight loss interventions [20, 33]. In our analysis, as the credible levels were relaxed, the covariates poor subjective general health and poorly controlled diabetes were consistently the ones to have an expected treatment benefit while the covariates good subjective general health with well-controlled diabetes had an expected null or adverse treatment effect.

The innovation of this paper firstly comes from the inferences made possible by the credible subgroups method: that there is a high posterior probability that all types of patients within the exclusive credible subgroup D experience a positive treatment effect. This addresses a different research question to previous publications that have explored patient heterogeneity but relied on sample average calculations. Our findings compared to those previous demonstrate that it is possible to calculate average estimates that do not capture the responses of minority subgroups with treatment effects that deviate from the mean. Secondly, the credible subgroups method allows us to make statements of simultaneity where all subgroup members experience a treatment effect above some threshold that adjusts for issues of multiplicity due to hypothesis tests in multiple subgroups. This helps to avoid the perception of data dredging which underpins prevalent scepticism of standard subgroup analyses.

Previous research has used exploratory endpoints and *ad hoc* subgroup analysis to identify significant treatment effects reported in secondary studies. Using the original CVD endpoint, our analysis shows that as the credible level is relaxed, the range of the covariate profiles for exclusive subgroup membership expands. There exists a trade-off between statistical confidence and subgroups too narrowly defined to represent any application in medical care. As the credible level is relaxed from 80% to 50%, the proportion of covariate points classified to the exclusive credible subgroup increases. Our research identifies a minority subgroup consisting of covariate profiles with an expected treatment benefit. Based on an 80% credible level, these covariate profiles are used to classify less than 0.5% of the Look AHEAD trial who would simultaneously be expected to experience a positive treatment effect. This contradicts previous research stating 85% of Look AHEAD participants did benefit from treatment as identified by positive average estimates [23].

The findings of this secondary analysis are not meant to overshadow the conclusions from the original Look AHEAD trial. Rather, we demonstrate an alternative method of exploring treatment effect heterogeneity that does not suffer from the limitations of standard subgroup analyses. The credible subgroups method defines subgroups, based on a number of baseline covariates, which consist of all the types of patients expected to benefit or not benefit from a given treatment while accounting for multiplicity. This research applies the credible subgroups method to the Look AHEAD trial to illustrate the trichotomy of the covariate space and

identifies a large number of covariate points located in an uncertainty region. This combined with the restrictive covariate profiles of those who would be classified to the exclusive credible subgroup, explains how so few patients with an expected treatment benefit could have been lost in trial-wide average estimates.

This analysis of the Look AHEAD trial has some limitations. At first glance the output from the credible subgroups method may not be as straightforward to interpret compared to standard subgroup analyses, but the fundamental objective of identifying individual covariate profiles of patients expected to benefit cannot be achieved through simple average calculations. Moreover, the illustrative results are not suitable for readers with colour blindness and therefore primary results are replicated in greyscale and included in the Supplement. Secondly, the credible subgroups method is not designed for variable selection. Instead, the credible subgroups method requires a list of covariates believed to have predictive value when interacted with treatment which requires previous knowledge of intervening factors. Thirdly, the credible subgroups method computationally struggles with increasing covariates due to a reduction in statistical power and the inclusion of many continuous covariates [6]. The possible biases produced by these limitations are similar to those found in standard subgroup analyses. Regarding the biological plausibility of the chosen covariates in particular, all are standard medical and demographic variables in studies examining behavioural treatments for type 2 diabetes and the confounding effects of obesity [14, 17, 18, 20, 23, 27]. We conclude that covariates previously identified as being important for exploring treatment heterogeneity may only explain a small portion of the entire covariate space that is useful for identifying positive treatment effects. In fact, variable selection itself introduces multiplicity problems, and accounting for this source of multiplicity as well would necessitate even more cautious conclusions. In particular, the inconclusive region ($D \setminus S$) would need to be enlarged for any given credible level. Some study of variable selection via Stochastic Search Variable Selection (SSVS) within the regression step of the credible subgroups approach has been done [34], which shows this effect in the sense that $D \setminus S$ is larger when not conditioning on a selected model. As this relates to the comparison with Baum et.al. [23], we have shown that accounting for only some but not all sources of multiplicity does not allow as optimistic conclusions as the previous analysis. This result would be made even more extreme by accounting for additional sources of multiplicity. Finally, point estimates for treatment effects in the selected subgroups D and S -complement are biased away from zero due to selection bias. Thus conclusions drawn from the credible subgroups method should be limited to hypothesis testing, and plots such as Fig 3 should be interpreted only as diagnostics and not as post-selection effect estimates.

There exists much scope for future research in the synergy of the credible subgroups method into adaptive enrichment trial design. The objectives of adaptive trial design and the credible subgroups method are the same: to increase the likelihood of identifying a true subgroup of patients expected to benefit from a given therapy [35]. Adaptive trial design includes prospectively planned analyses that can change the course of an on-going trial based on accumulating the data of subjects within the trial [36]. Adaptive trial design stands to benefit from the credible subgroups method if used throughout interim analyses and at early stage enrollee selection. Sequential trial design permits stopping early for efficacy based on interim analyses [37, 38], which could be defined by the trichotomy of the covariate space using the credible subgroups method. Overwhelming evidence of treatment benefit or harm would allow for early termination while the trial would continue to gain information about the covariate space where classifications are uncertain. Additionally, the credible subgroups method could be used in an adaptive population-enrichment design where biomarkers are identified that would predict response or lack of response to a given therapy [39]. This information could be used in confirmatory trials to limit entry of patients to stage 3 randomisation [40]. In this research,

biomarkers for diabetes and obesity that are related to the effectiveness of the lifestyle intervention would be of particular interest. Future incorporation of the credible subgroups method into adaptive enrichment trial design can magnify the potential treatment effect, reduce costs and increase trial efficiency.

Supporting information

S1 Fig. Trichotomy of baseline covariates at 80% credible level with vague prior on predictive and prognostic effects. Allows for assumptions of treatment effect homogeneity within all levels of selected covariates.

(TIF)

S2 Fig. Trichotomy of baseline covariates at 65% credible level. Individual credible subgroup pairs plotted over the covariate space. There is at least 65% posterior probability that all patients with covariate points in D have a conditional average treatment effect positive treatment effect ($\delta = 0$).

(TIF)

S3 Fig. Parameter autocorrelation plots.

(TIF)

S4 Fig. Trichotomy of the predictive covariate space. Visualisation of the trichotomy of the covariate space based on the credible subgroup pair (D,S) relative to the true benefiting subgroup (B), adapted from Schnell et.al. [6]

(TIF)

S5 Fig. Trichotomy of baseline covariates at 80% credible level. Individual credible subgroup pairs plotted over the covariate space. There is at least 80% posterior probability that all patients with covariate points in D have a positive conditional average treatment effect ($\delta = 0$).

(TIF)

S6 Fig. Trichotomy of baseline covariates at 50% credible level. Individual credible subgroup pairs plotted over the covariate space. There is at least 50% posterior probability that all patients with covariate points in D have a positive conditional average treatment effect treatment effect ($\delta = 0$).

(TIF)

S7 Fig. Trichotomy of baseline covariates at 80% credible level with vague prior on predictive and prognostic effects. Allows for assumptions of treatment effect homogeneity within all levels of selected covariates.

(TIF)

S8 Fig. Trichotomy of baseline covariates at 65% credible level. Individual credible subgroup pairs plotted over the covariate space. There is at least 65% posterior probability that all patients with covariate points in D have a conditional average treatment effect positive treatment effect ($\delta = 0$).

(TIF)

Acknowledgments

Data are available upon request from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) at <https://www.niddk.nih.gov/>.

Author Contributions

Conceptualization: Joel Smith, John Forbes.

Formal analysis: Anna Coonan.

Investigation: Anna Coonan.

Methodology: Patrick Schnell, John Forbes.

Supervision: John Forbes.

Writing – original draft: Anna Coonan.

Writing – review & editing: Anna Coonan, Patrick Schnell, Joel Smith, John Forbes.

References

1. Cameron AC, Trivedi PK. Microeconometrics: methods and applications: Cambridge university press; 2005.
2. Basu A. ESTIMATING PERSON-CENTERED TREATMENT (PeT) EFFECTS USING INSTRUMENTAL VARIABLES: AN APPLICATION TO EVALUATING PROSTATE CANCER TREATMENTS. *J Appl Econ* (Chichester Engl). 2014; 29(4):671–91.
3. Basu A, Jena AB, Philipson TJ. The impact of comparative effectiveness research on health and health care spending. *J Health Econ*. 2011; 30(4):695–706. <https://doi.org/10.1016/j.jhealeco.2011.05.012> PMID: 21696840
4. Heckman J. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *J Human Res*. 1997:441–62.
5. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002; 21(19):2917–30. <https://doi.org/10.1002/sim.1296> PMID: 12325108
6. Schnell P, Tang Q, Walter OW, Carlin BP. A Bayesian credible subgroups approach to identifying patient subgroups with positive treatment effects. *Biometrics*. 2016; 72(4):1026–36. <https://doi.org/10.1111/biom.12522> PMID: 27159131
7. Ruberg SJ, Lei C, Yanping W. The mean does not mean as much anymore: finding sub-groups for tailored therapeutics. *Clin Trials*. 2010; 7(5):574–83. <https://doi.org/10.1177/1740774510369350> PMID: 20667935
8. Olken BA. Promises and perils of pre-analysis plans. *J Econ Perspect*. 2015; 29(3):61–80.
9. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Springer series in statistics New York, NY, USA.; 2001.
10. Athey S, Imbens GW. Machine learning methods for estimating heterogeneous causal effects. *Stat*. 2015; 1050(5).
11. Imai K, Ratkovic M. ESTIMATING TREATMENT EFFECT HETEROGENEITY IN RANDOMIZED PROGRAM EVALUATION. *Ann Appl Stat*. 2013; 7(1):443–70.
12. Green DP, Kern HL. Modeling heterogeneous treatment effects in survey experiments with Bayesian additive regression trees. *Public Opin Q*. 2012; 76(3):491–511.
13. Simon R. Bayesian subset analysis: application to studying treatment-by-gender interactions. *Stat Med*. 2002; 21(19):2909–16. <https://doi.org/10.1002/sim.1295> PMID: 12325107
14. Look AHEAD Research Group. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *N Engl J Med*. 2013; 369(2):145–54. <https://doi.org/10.1056/NEJMoa1212914> PMID: 23796131
15. Look AHEAD Research Group. Look AHEAD (Action for Health in Diabetes): design and methods for a clinical trial of weight loss for the prevention of cardiovascular disease in type 2 diabetes. *Control Clin Trials*. 2003; 24(5):610–28. [https://doi.org/10.1016/s0197-2456\(03\)00064-3](https://doi.org/10.1016/s0197-2456(03)00064-3) PMID: 14500058
16. Schellenberg ES, Dryden DM, Vandermeer B, Ha C, Korownyk C. Lifestyle interventions for patients with and at risk for type 2 diabetes: a systematic review and meta-analysis. *Ann Intern Med*. 2013; 159(8):543–51. <https://doi.org/10.7326/0003-4819-159-8-201310150-00007> PMID: 24126648
17. Li G, Zhang P, Wang J, Gregg EW, Yang W, Gong Q, et al. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. *Lancet*. 2008; 371(9626):1783–9. [https://doi.org/10.1016/S0140-6736\(08\)60766-7](https://doi.org/10.1016/S0140-6736(08)60766-7) PMID: 18502303

18. Uusitupa M, Peltonen M, Lindström J, Aunola S, Ilanne-Parikka P, Keinänen-Kiukkaanniemi S, et al. Ten-year mortality and cardiovascular morbidity in the Finnish Diabetes Prevention Study—secondary analysis of the randomized trial. *PLoS One*. 2009; 4(5):e5656. <https://doi.org/10.1371/journal.pone.0005656> PMID: 19479072
19. Li G, Zhang P, Wang J, An Y, Gong Q, Gregg EW, et al. Cardiovascular mortality, all-cause mortality, and diabetes incidence after lifestyle intervention for people with impaired glucose tolerance in the Da Qing Diabetes Prevention Study: a 23-year follow-up study. *Lancet Diabetes Endocrinol*. 2014; 2(6):474–80. [https://doi.org/10.1016/S2213-8587\(14\)70057-9](https://doi.org/10.1016/S2213-8587(14)70057-9) PMID: 24731674
20. Look AHEAD Study Group, Gregg EW. Association of the Magnitude of Weight Loss and Physical Fitness Change on Long-term CVD outcomes: The Look AHEAD Study. *Lancet Diabetes Endocrinol*. 2016; 4(11):913. [https://doi.org/10.1016/S2213-8587\(16\)30162-0](https://doi.org/10.1016/S2213-8587(16)30162-0) PMID: 27595918
21. Look AHEAD Research Group. Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes: one-year results of the look AHEAD trial. *Diabetes Care*. 2007.
22. Breyer BN, Phelan S, Hogan PE, Rosen RC, Kitabchi AE, Wing RR, et al. Intensive lifestyle intervention reduces urinary incontinence in overweight/obese men with type 2 diabetes: results from the Look AHEAD trial. *J Urol*. 2014; 192(1):144–9. <https://doi.org/10.1016/j.juro.2014.02.036> PMID: 24533998
23. Baum A, Scarpa J, Bruzelius E, Tamler R, Basu S, Faghmous J. Targeting weight loss interventions to reduce cardiovascular complications of type 2 diabetes: a machine learning-based post-hoc analysis of heterogeneous treatment effects in the Look AHEAD trial. *Lancet Diabetes Endocrinol*. 2017; 5(10):808–15. [https://doi.org/10.1016/S2213-8587\(17\)30176-6](https://doi.org/10.1016/S2213-8587(17)30176-6) PMID: 28711469
24. Wadden TA, Butryn ML, Hong PS, Tsai AG. Behavioral treatment of obesity in patients encountered in primary care settings: a systematic review. *JAMA*. 2014; 312(17):1779–91. <https://doi.org/10.1001/jama.2014.14173> PMID: 25369490
25. Gregg EW, Wing R. Looking again at the Look AHEAD study. *Lancet Diabetes Endocrinol*. 2017; 5(10):763–4. [https://doi.org/10.1016/S2213-8587\(17\)30238-3](https://doi.org/10.1016/S2213-8587(17)30238-3) PMID: 28711470
26. Pi-Sunyer X. The Look AHEAD trial: a review and discussion of its outcomes. *Current nutrition reports*. 2014; 3(4):387–91. <https://doi.org/10.1007/s13668-014-0099-x> PMID: 25729633
27. Li G, Zhang P, Wang J, An Y, Gong Q, Gregg EW, et al. Cardiovascular mortality, all-cause mortality, and diabetes incidence after lifestyle intervention for people with impaired glucose tolerance in the Da Qing Diabetes Prevention Study: a 23-year follow-up study. *Lancet Diabetes Endocrinol*. 2014; 2(6):474–80. [https://doi.org/10.1016/S2213-8587\(14\)70057-9](https://doi.org/10.1016/S2213-8587(14)70057-9) PMID: 24731674
28. Nimble Development Team. NIMBLE: An R Package for programing with BUGS models. 0.4 ed2015.
29. R Core Team. R. A Language and Environment for Statistical Computing. Vienna, Austria R Foundation for Statistical Computing; 2019.
30. Brancati FL, Evans M, Furberg CD, Geller N, Haffner S, Kahn SE, et al. Midcourse correction to a clinical trial when the event rate is underestimated: the Look AHEAD (Action for Health in Diabetes) Study. *Clin Trials*. 2012; 9(1):113–24. <https://doi.org/10.1177/1740774511432726> PMID: 22334468
31. Schnell P, Carlin B. credsubs: Credible Subsets R package version 1.0. 2017.
32. Unick JL, Beavers D, Jakicic JM, Kitabchi AE, Knowler WC, Wadden TA, et al. Effectiveness of lifestyle interventions for individuals with severe obesity and type 2 diabetes: results from the Look AHEAD trial. *Diabetes Care*. 2011;DC_110874.
33. Harrington M, Gibson S, Cottrell RC. A review and meta-analysis of the effect of weight loss on all-cause mortality risk. *Nutr Res Rev*. 2009; 22(1):93–108. <https://doi.org/10.1017/S0954422409990035> PMID: 19555520
34. Schnell P. Credible subgroups: identifying the population that benefits from treatment. University of Minnesota Digital Conservancy: <http://hdl.handle.net/11299/188890>; 2017.
35. Bauer P, Bretz F, Dragalin V, König F, Wassmer G. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. *Stat Med*. 2016; 35(3):325–47. <https://doi.org/10.1002/sim.6472> PMID: 25778935
36. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Adaptive Designs for Clinical Trials of Drugs and Biologics. Silver Spring, MD: Food and Drug Administration 2018.
37. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika*. 1977; 64(2):191–9.
38. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979:549–56. PMID: 497341
39. Bhatt DL, Mehta C. Adaptive designs for clinical trials. *N Engl J Med*. 2016; 375(1):65–74. <https://doi.org/10.1056/NEJMr1510061> PMID: 27406349
40. Simon N, Simon R. Adaptive enrichment designs for clinical trials. *Biostatistics*. 2013; 14(4):613–25. <https://doi.org/10.1093/biostatistics/kxt010> PMID: 23525452