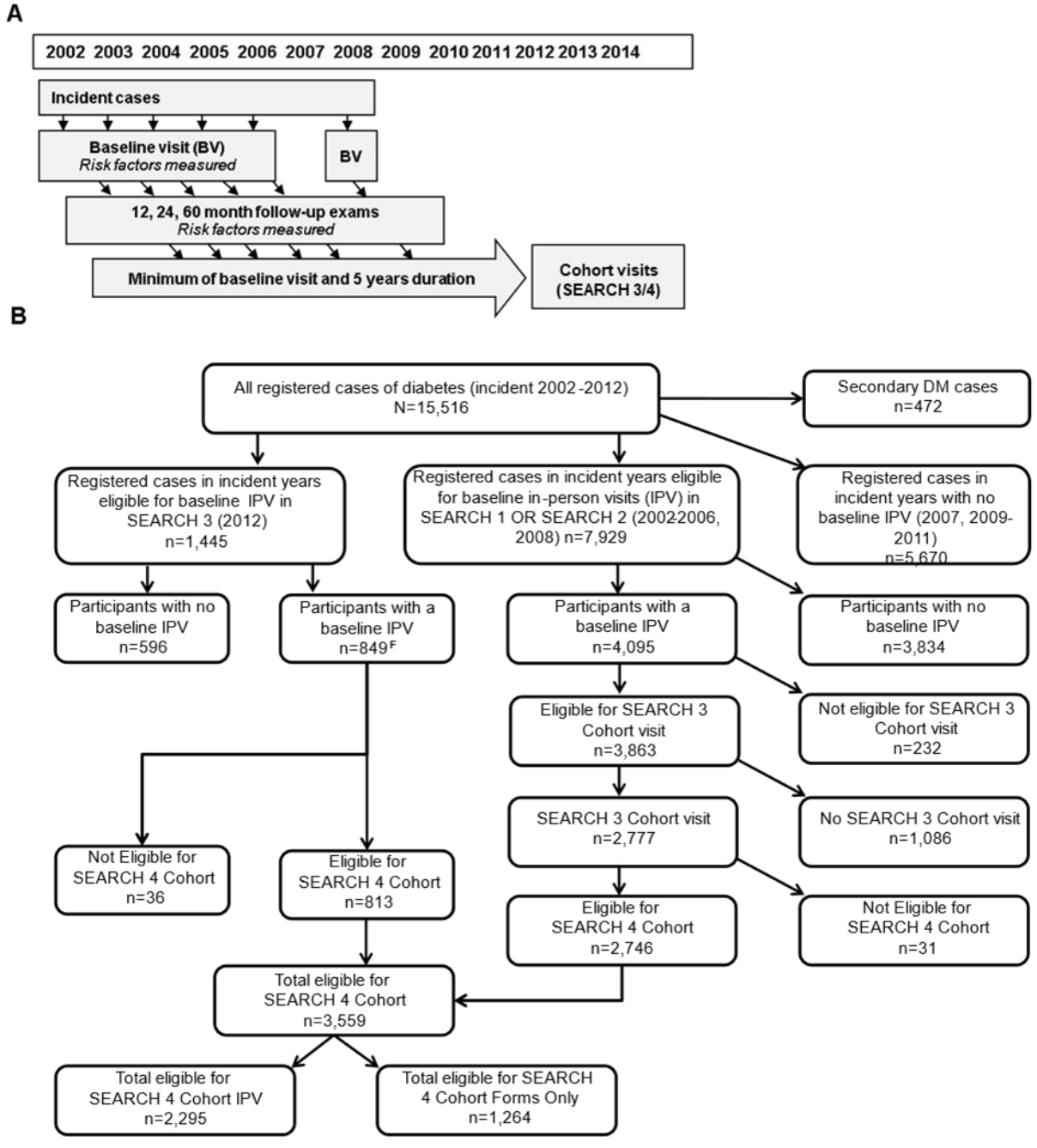
Supplemental Appendix

Supplemental Figures

Figure S1: A: SEARCH Study Design; B: Participants Eligible for SEARCH 3 and 4 'Cohort' visits.



Supplemental Tables

Table S1: Number of Non-White SEARCH participants included in each modeling step. Multiple outcome regressions are done based on data from two consecutive SEARCH study visits except the last visit; participants were excluded in each regression if they had missing data. Outcomes include hemoglobin A1c (HbA1c), measured at all six SEARCH visits, and four separate diabetes complications measured at the SEARCH cohort visits: Diabetic retinopathy (complication RET), Diabetic kidney disease (complication DKD), Cardiovascular autonomic neuropathy (complication CAN), and Peripheral neuropathy (complication PN).

	SEARCH Study Visits	Sample size
Hba1c Outcome Regression	Cohort 4 Visit	144
	Cohort 3 Visit, Cohort 4 Visit	129
	60-mo Follow-up Visit, Cohort 3 Visit	145
	24-mo Follow-up Visit, 60-mo Follow-up Visit	70
	12-mo Follow-up Visit, 24-mo Follow-up Visit	85
	Baseline Visit, 12-mo Follow-up Visit	112
Baseline Hba1c Value Function Evaluation	Baseline Visit	220
Diabetic Retinopathy Outcome Regression	Cohort 4 Visit	144
	Cohort 3 Visit, Cohort 4 Visit	130
Diabetic Retinopathy Value Function Evaluation	Cohort 3 Visit	259
Diabetic Kidney Disease Outcome Regression	Cohort 4 Visit	135
	Cohort 3 Visit, Cohort 4 Visit	119
Diabetic Kidney Disease Value Function Evaluation	Cohort 3 Visit	229
Cardiovascular Autonomic Neuropathy Outcome Regression	Cohort 4 Visit	139
	Cohort 3 Visit, Cohort 4 Visit	119
Cardiovascular Autonomic Neuropathy Value Function Evaluation	Cohort 3 Visit	244
Peripheral Neuropathy Outcome Regression	Cohort 4 Visit	150
	Cohort 3 Visit, Cohort 4 Visit	128
Peripheral Neuropathy Value Function Evaluation	Cohort 3 Visit	260

Analysis Approach

The aim of the analysis is to use historical data (throughout all SEARCH visits) to estimate the effect of two different estimated treatment policies: π^{white} and π^{nonwhite} on the outcome of patients over a period of time.

1 Notation and Model Covariates

We use the following notation:
Let X_t denotes a set of covariates that affects treatment assignments at time t
Let A_t denotes the treatment at time t
Let Z_t denotes a set of covariates that affects the outcome at time t
Let $Y_t(z,a)$ denotes the outcome of treatment a and covariate z at time t

Given a discrete time point $t = 1,...,T$ ($T = 6$ in this dataset); we observed $\{S_t = \{X_t, Z_t\}, A_t, Y_t(z_t, a_t)\}$ for each subject. For this analysis, the covirates set X_t includes age at diagnosis, sex, SEARCH study site, SEARCH visit, and T1D duration. The covariates set Z_t includes age at diagnosis, sex, SEARCH site, T1D duration, maximum parental education, health insurance type, smoking status, physical activity, screen time, and an indicator variable for non-Hispanic Black (versus other nonwhite). Note that although X_t and Z_t overlaps, X_t is used in the propensity score models in Section 2 and Z_t is used in the outcome regression models in Section 3.

Treatment options at each time point A_t include two aspects of diabetes management: insulin delivery modality and self-monitored glucose frequency. In the primary outcome analysis, Y_t is a measure of glycemic control HbA1c whereas in each of the secondary outcome analyses, Y_t is whether the subject has the following early diabetes complication: Diabetic Retinopathy, Diabetic Kidney Disease, Cardiovascular Autonomic Neuropathy, and Peripheral Neuropathy.

2 Multivariate propensity score modeling to estimate the diabetes treatment regimen distributions

The treatment regimen, or distribution of treatments, among the white subgroup was represented by π^{white}, while the treatment regimen observed among the nonwhite subgroup was represented by π^{nonwhite}. Data from the six possible SEARCH study visits were used to model a propensity score to estimate the treatment regimen distributions in both subgroups: white and nonwhite controlling for age, sex, SEARCH study site, SEARCH visit, and T1D duration. The models were fit separately for racial/ethnic subgroups, i.e. the π^{nonwhite} was fit on the nonwhite subpopulation and the π^{white} was fit on the white subpopulation. Multinomial logistic regression was used to fit the probability of treatment options given a linear combination of the aforementioned covariates. Three propensity score models were fit for insulin delivery modality, SMG (before cohort visits), and SMG with CGM use (cohort visits), respectively. Two separate models for frequency of glucose monitoring were fit for the visits (before the cohort visits versus cohort visits) to incorporate the availability of continuous glucose monitoring (CGM) that became available over the duration of the study.

Treatment regimen is a map $\pi : X \to P^m$ from covariates X to a vector of probabilities of size m denoting m different treatment options.

The first step is to estimate a propensity score function $\Psi(x)$ using a multinomial logistic regression fitted to $\{X_i, A_i\}$. Note that $\Psi(x) = (\psi_1(x), ..., \psi_m(x))$ is a vector of probability for m different treatment options.

We fit the propenstity score function on two subpopulations: white and nonwhite. This gives us $\Psi^{\text{white}}(x)$ and $\Psi^{\text{nonwhite}}(x)$ which are the estimate for two different treatment regimens π^{white} and π^{nonwhite}

3 Estimation of the effect of the diabetes treatment regimens on clinical outcomes

The second step is estimating each policy's effect on the outcome of patients over time. This problem is often known as off-policy policy evaluation in the reinforcement learning literature (see [1] and [2] for review of methods) and is also in the tradition of estimating counterfactual quantities within the potential outcome

framework in the statistcal causal inference literature. In this analysis, we take the approach of approximating the action-value function (Q-function) in the context of finite horizon (i.e. the number of discrete time points T is finite). The approach is based on the Fitted Q Iteration (FQI) algorithm [3] which is a model-free off-policy learning for batch mode reinforcement learning in the context of infinite horizon setting. A modification of FQI called Fitted Q Evaluation (FQE [4]) for off-policy policy evaluation has been shown to work well based on empirical study [1]. Details of our approach to estimate the Q function using regression in the finite horizon setting is given below.

The sequential data that we observed $\{S_t = \{X_t, Z_t\}, A_t, Y_t(z_t, a_t)\}$ for each subject can be modeled with a Markov Decision Process (MDP). In our case, the MDP is defined by $< S, A, P_1, P, \mu_y >$

- S is the state space which consist of both X(covariates that affects treatment assignments) and Z (covariates that affects the outcome)
- A is the set of treatment options
- P_1 is the intial state distribution and P is the transition function representing the transition probability $P(s'|s,a)$
- μ_y is the mean outcome model which specifies the mean of $Y_t(s_t, a_t)$

Given a MDP defined by $< S, A, P, \mu_y >$, a trajectory τ is a sequence of states S_t, actions A_t, and Y_t from $t = 1, ..., T$. Our observational data are i.i.d n trajectetories $\{\tau_1, ..., \tau_n\}$. The data generating mechanism is defined by the intial state distribution P_1, transition probability P, a density with the mean outcome μ_y, as well as a treatment regimen π^D that spicifies the conditional probability of the treatment options in the data.

In the context of this problem, we are interested in estimating the counterfactual quantity

$$V^{\pi^T}(s) = E_{P,\pi^T}[\sum_{t=1}^{T} \gamma^{t-1} Y_t | S_1 = s] \tag{1}$$

where γ is the discounting factor and π^T is the target treatment policy (not the treatment policy generating the data π^D)

In our analysis, we need to estimate the following two functions

$$V^{\pi^{\text{white}}}(s) \text{ and } V^{\pi^{\text{non-white}}}(s)$$

Estimating the above functions can be done recursively by the definitons below

$$V^{\pi}(s) := E_{a\sim\pi(s)}[Q_0^{\pi}(s,a)]$$

$$Q_t^{\pi}(s,a) := E_{s'}[Y_t(s,a) + \gamma E_{a'\sim\pi(s')}[Q_{t+1}^{\pi}(s',a')]]$$

Given $V^{\pi^{\text{white}}}(s)$ and $V^{\pi^{\text{non-white}}}(s)$, we can then estimate the following quantity $v(\pi) = \frac{1}{n}\sum_i^n V^{\pi}(s_i)$ for any population of size n with the initial states s_i. The quantities $v(\pi^{\text{white}})$ and $v(\pi^{\text{non-white}})$ represent the effect of two policies on the outcome in a population of interest.

To estimate the value function $V^{\pi}(s)$ and subsequently $v(\pi)$, we take a regression-based procedure to estimate the Q function above. Based on the observational data of trajectories $\{\tau_1,...,\tau_n\}$ which in our scenario has missing data at some time-points, we approximate a sequence of Q_t^{π} function starting at $t = T$ and iterate backward from $t = T-1$ to $t = 1$ based on data from two consecutive timepoints.

At $t = T$ (last time point), we use regression to approximate the function

$$Q_T(s,a) = E[Y_T(s,a)]$$

Then, at each time-point $t = T-1,...,1$, we first construct the following pseudo-value $\tilde{y}_t$

$$\tilde{y}_t = y_t + \gamma \sum_{a_{t+1}} \psi_{a_{t+1}}^{\pi}(x_{t+1})\hat{Q}^{t+1}(z_{t+1}, a_{t+1}) \tag{2}$$

We then use regression to approximate the function Q_t based on $\{((x_t, z_t, a_t), \tilde{y}_t)_i; i = 1,...,n_t\}$ where n_t is the number of subjects with data available for timepoint t and $t+1$ and in this case $\gamma = 1$ (no discounting; discounting is not necessary since T is finite).

At baseline visit, we can therefore estimate $V^{\pi}(s)$ for each subject and evaluate the treatment regimen outcomes at the population-level by computing the average

of the outcome across a population:

$$v(\pi) = \frac{1}{n}\sum_{i}^{n} V^{\pi}(s_i)$$

Note that in this manuscript, the primary outcome is glycemic control (Hba1c %). Because we set $\gamma = 1$ (no discounting), $V^{\pi^T}(s)$ is just the average Hba1c level across 6 visits under the target treatment policy π^T according to equation (1).

For the secondary outcomes (risk of developing diabetes complications at either cohort visits), we modify equation (1) to

$$V^{\pi^T}(s) = E_{P,\pi^T}[y_1 + (1-y_1)y_2 | S_1 = s]$$

where y_1 is probability of developing complication at cohort visit 1 and y_2 is the probability of developing complication at cohort visit 2. Similarly, for these secondary outcomes, we need to modify equation (3) to

$$\tilde{y_t} = y_t + \gamma(1-y_t)\sum_{a_{t+1}} \psi^{\pi}_{a_{t+1}}(x_{t+1})\hat{Q}^{t+1}(z_{t+1}, a_{t+1})$$

4 Weighted Bootstrap for Inference of $v(\pi)$

We now discuss the inference for the value function $v(\pi)$ and the conference interval for the estimate of $v(\pi)$. In this work, we rely on the multiplier bootstrap or weighted bootstrap to perform inference on $v(\pi)$. Specifically, the weighted bootstrap has the following weights $w_i = \xi_i/\bar{\xi}$ where ξ_i has a standard exponential distribution and $\bar{\xi} = n^{-1}\sum_{i=1}^{n}\xi_i$.

We first define (d_{ij}, u_{ij}) for $j = 1, ..., K$ and $i = 1, ..., n$ where K is the finite number of time steps and n is the total number of subjects. Let d_{ij} be the indicator of patient i having data at visit j and let u_{ij} be the data for patient i at visit j i.e. $u_{ij} = (S_{ij}, A_{ij})$, let $h_i = (d_{ij}, u_{ij}, y_{ij}; 1 \leq j \leq k)$. Also, denote the shorthand notation $\tilde{u}_{ij}(a) = (S_{ij}, a)$

Let $\pi(a|X, \theta)$ be the propensity score for treatment option a given covariates X and parameter θ. Let $\hat{\theta}_n$ be the estimate for θ given a dataset, and note that it

can be written as a differentiable function of the empirical process $\mathbb{P}_n$ of the data $\{h_i, i = 1, ..., n\}$

Let $\mathbb{P}_n^*$ be the weighted bootstrap process. Define the weighted bootstrap empirical measure as follow: $\mathbb{P}_n^* f = n^{-1} \sum_{i=1}^n (\xi_i / \bar{\xi}) f(X_i)$

At the K'th (last) time step, it can be shown that the parameter of the estimated Q function has the form

$$\hat{\Psi}_{n_k} = \left(\frac{\mathbb{P}_n d_j u_j u_j^T}{\mathbb{P}_n d_j} \right)^{-1} \frac{\mathbb{P}_n d_j (u_j y_j)}{\mathbb{P}_n d_j}$$

And $\hat{\Psi}_{n_k}$ is a differentiable function of $\mathbb{P}_n h$ and $\hat{\theta}_n$

For $j = 1, ..., K-1$, the parameter of the estimated Q function has the form

$$\hat{\Psi}_{n_j} = \left(\frac{\mathbb{P}_n d_j u_j u_j^T}{\mathbb{P}_n d_j} \right)^{-1} \frac{\mathbb{P}_n d_j u_j (y_j + \sum_{\tilde{a} \in \mathscr{A}} \pi(\tilde{a}, \theta_n)(\tilde{u}_{j+1}(\tilde{a}))^T \hat{\Psi}_{n_{j+1}})}{\mathbb{P}_n d_j}$$

Hence, $\hat{\Psi}_{n_j}$ is a differentiable function of $\mathbb{P}_n h, \hat{\Psi}_{n_{j+1}}, ..., \hat{\Psi}_{n_k}$ and $\hat{\theta}_n$

Specifically, at baseline,

$$\hat{V}_{n_1}(x_{i0}) = \sum_{\tilde{a} \in \mathscr{A}} \pi(\tilde{a}, \theta_n)(\tilde{u}_{i,j+1}(\tilde{a}))^T \hat{\Psi}_{n_1}$$

And finally,

$$\hat{v}_{n_1}(\theta_n) = \mathbb{P}_n \hat{V}_{n_1}(x_{i0})$$

Since K is finite, and by the chain rule, we obtain that $\hat{v}_{n_1}(\pi_{\theta_n})$ is a Hadamand differentiable functional of an empirical process contained in a Donsker class and thus

$$\sqrt{n}(\hat{v}_{n_1}(\theta_n) - v_0(\theta_n)) \rightsquigarrow_d \mathscr{N}(0, \sigma^2) \tag{3}$$

$v_0(\theta_n)$ is the expected value function at baseline of future patients from the same

population receiving the same propensity for treatment (reflected in θ_n). Since $\hat{\theta}_n \to_p \theta_0$, combined with the smoothness of $\sqrt{n}(\hat{v}_{n_1}(\theta) - v_0(\theta))$ over θ in a neighborhood of θ_0, the covergence in (3) follows (See lemma 13.3 of [5]).

Also, suppose $\hat{v}^*_{n_1}(\theta_n)$ is the weighted bootstrap version, we obtain

$$\sqrt{n}(\hat{v}^*_{n_1}(\theta_n) - \hat{v}_{n_1}(\theta_n)) \rightsquigarrow_d \mathcal{N}(0, \sigma^2) \tag{4}$$

conditioned on data, via similar arguements as above replacing $\mathbb{P}_n$ with $\mathbb{P}^*_n$ combined with theorem 2.6 and 2.9 of [5]

5 Imputation of Missing Longitudinal Values of HbA1c

This section describes the multiple imputation method used to estimate the Non-White subgroup and White subgroup HbA1c outcome as well as their confidence intervals. Specifically, Table 2 reported that the population mean of the longitudinal average of HbA1c is 9.2 (95% confidence interval = (8.9, 9.4)) for the Non-White Subgroup and 8.2 (95% confidence interval = (8.1, 8.3)) for the White Subgroup. Note that no imputation was performed for the analysis described in previous sections where we estimate the outcomes under the two different treatment regimens.

Because some longitudinal HbA1c values for each subject are missing, to estimate longitudinal average we need to impute some values. Out of 978 subjects, 7.7% has HbA1c values for all 6 visits , 17.8% has values for 5 visits, 21.6% has values for 4 visits, 20.9% has values for 3 visits, 16.6% has values for 2 visits, and 15.5% has values for 1 visit. Multiple imputation allows us to account for the additional uncertainty of the unobserved (true) values of these longitudinal HbA1c values. We use the 2l.pan method [6] from the mice R package [7] to perform the imputation (for review of various imputation methods for longitudinal data, see [8]). This method conducts imputation of missing values by sampling from a linear two-level hierachical model with homogeneous within-subject variance. Two-level hierachical model is appropriate in this case because the HbA1c values observed over time are nested within each subject. The following baseline covariates are included in the model for imputation: age at diagnosis, sex, SEARCH site, smoking status, maximum parental education, and health insur-

ance type. Note also that imputation is conducted separately for Non-White and White subgroup.

The standard multiple imputation procedure is followed. The details of the 5 steps process is given below.

- Impute the missing longitudinal HbA1c values using 2l.pan method separately for the two subgroups.
- For each imputed dataset, calculate the subgroup mean $\hat{\mu}_j$ and variance $\hat{\sigma}_j^2$ of the longitudinal average of the HbA1c values.
- Repeat the imputation 10 times to obtain $\hat{\mu}_j, \hat{\sigma}_j^2$ for $j = 1, ..., 10$
- Average the values of the mean estimate from the imputations to get a point estimate. $\tilde{\mu} = \frac{1}{10}\sum_j \hat{\mu}_j$
- Calculate the total variance estimator (for constructing confidence interval) as follow

$$\tilde{\sigma}^2 = \frac{1}{9}\sum_{j=1}^{10}(\hat{\mu}_j - \tilde{\mu})^2 + \frac{1}{10}\sum_{j=1}^{10}\hat{\sigma}_j^2$$

References

[1] Cameron Voloshin, Hoang M. Le, Nan Jiang, and Yisong Yue. Empirical study of off-policy policy evaluation for reinforcement learning. *ICML*, 2019.

[2] Christoph Dann, Gerhard Neumann, and Jan Peters. Policy evaluation with temporal differences: A survey and comparison (extended abstract). *Journal of Machine Learning Research*, 15, 2014. ISSN 23340843.

[3] Damien Ernst, Pierre Geurts, and Louis Wehenkel. Tree-based batch mode reinforcement learning. *Journal of Machine Learning Research*, 6:503–556, 2005. ISSN 15337928.

[4] Hoang Le, Cameron Voloshin, and Yisong Yue. Batch Policy Learning under Constraints. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*,

pages 3703–3712, Long Beach, California, USA, 2019. PMLR. URL http://proceedings.mlr.press/v97/le19a.html.

[5] Michael R Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer, 2008.

[6] Joseph L Schafer and Recai M Yucel. Computational Strategies for Multivariate Linear Mixed-Effects Models With Missing Values. *Journal of Computational and Graphical Statistics*, 11 (2):437–457, 2002. doi: 10.1198/106186002760180608. URL https://doi.org/10.1198/106186002760180608.

[7] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, Articles*, 45(3):1–67, 2011. ISSN 1548-7660. doi: 10.18637/jss.v045.i03. URL https://www.jstatsoft.org/v045/i03.

[8] Md Hamidul Huque, John B Carlin, Julie A Simpson, and Katherine J Lee. A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Medical Research Methodology*, 18(1): 168, 2018. ISSN 1471-2288. doi: 10.1186/s12874-018-0615-6. URL https://doi.org/10.1186/s12874-018-0615-6.