ANALYSIS OF SEQUENTIAL MEDIATORS IN THE RELATIONSHIP BETWEEN MISTT INTERVENTIONS AND PHYSICAL AND MENTAL HEALTH OF STROKE PATIENTS

By

Michael Bridges

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biostatistics—Master of Science

2021

ABSTRACT

ANALYSIS OF SEQUENTIAL MEDIATORS IN THE RELATIONSHIP BETWEEN MISTT INTERVENTIONS AND PHYSICAL AND MENTAL HEALTH OF STROKE PATIENTS

By

Michael Bridges

The Michigan Stroke Transitions Trial (*MISTT*) is a pragmatic, un-blinded, 3-group randomized controlled trial conducted in 3 Michigan hospitals and designed to compare patient recovery under one of two interventions, (1) social worker case management (*SWCM*) and (2) social worker case management in addition to an online stroke recovery resource (*VSSP*), relative to usual care (*UC*). Using a difference-in-differences approach, and comparing outcome measures at 90-days post-discharge to outcomes at 7-days post-discharge, *MISTT* found significant positive outcomes in PROMIS physical health ($p = 0.002$) and Patient Activation ($p = 0.06$) in the *VSSP* treatment arm relative to *UC*.

We hypothesized that emotional support and patient activation acted as sequential mediators in the pathway between randomly assigned treatment and physical and mental health. We estimated the direct and indirect effects of the interventions using an adapted version of the weighted approach. Multiple imputation was used to account for missing observations and bootstrapping was used to construct standard errors. We found no statistically significant ($p < 0.05$) mediation effects. That said, we observed a sizable positive natural direct effect of the *VSSP* treatment relative to usual care on patient physical health (+1.40, 95% CI: –0.56, 3.35). In addition, there appeared to be a negative partial natural indirect effect of the *SWCM* treatment on both mental and physical health which acted through patient activation, not emotional support.

We report no mediated effects of statistical significance; however, some sizable effects bear further study. In particular, neither of our hypothesized mediators appeared to fully explain the positive effect of *VSSP* treatment on physical health shown in the *MISTT* primary results; and the lack of change in mental health found in *MISTT* might be explained by a decrease in patient activation found in the *SWCM* treatment arm.

ACKNOWLEDGEMENTS

I would like to thank Dr. Zhehui Luo for her unique ability to effectively communicate her deep statistical knowledge (and the patience and kindness to do so frequently); Dr. Mat Reeves for being an excellent mentor, boss, and friend; and Dr. Joseph Gardiner for his SAS expertise. I would also like to thank the MSU department of Epidemiology and Biostatistics, which has been a massive part of my life over the past several years, and to which I will always be grateful. Finally, I would like to thank Colleen for continuing to laugh at (most of) my jokes through 14 months of quarantine.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

A stroke represents a potentially life-altering event that might see a patient admitted briefly to a medical facility and discharged a few days later with little or no additional follow-up. However, the road to recovery continues long after discharge; Bambhroliya et al.[2] tracked a cohort of over two million US adult stroke patients and found that over 10% of them were readmitted to a medical facility within 30 days of discharge, a percentage that varied slightly based on stroke subtype. One of the most critical problems recovering stroke patients face is a lack of information and access to support resources. Hare et al.[5] conducted a qualitative study of stroke survivors that documented their recovery experiences. Several themes emerged from these sessions, but none more definitively than the "general consensus in all groups and interviews", among both patients and caregivers, that they needed better access to stroke recovery resources that provided information about "not only living with a stroke and the problems that might arise from it, but also … wider issues such as adaptations to property, benefits advice, appropriate exercise, points of contact, opportunities to network, surviving a stroke, and preventing further strokes".[5] In short, patients and caregivers alike wanted access to recovery resources to help them solve problems they might not have anticipated while in the hospital. There is a notable gap here that requires sound research on efficient ways to connect recovering patients with reliable resources following hospital discharge.

The Michigan Stroke Transitions Trial

The Michigan Stroke Transitions Trial (*MISTT*) is a pragmatic, un-blinded (open) randomized controlled trial that was conducted in three Michigan hospitals. The protocol has been described elsewhere[13], but in brief, participants were enrolled following hospital admission for an acute stroke event and were randomly assigned to one of three treatment groups upon their discharge home. The three treatment arms were: (1) usual care (*UC*), (2) social worker case management (*SWCM*), and (3) social worker case management in addition to access to the *MISTT* website (*VSSP*), which contained information and support services relevant to stroke recovery. Participants were interviewed seven days

after discharge, and again at 90 days after discharge. Pre-treatment covariate data is available from the initial enrollment, and all outcome data from the two interviews. *Table 1* below shows the baseline covariate data for our sample by randomized treatment arm.

Table 1: Patient pre-treatment covariates. Note that p-values are the result of one-way ANOVA for continuous variables, and chi-square tests for categorical variables. Percentages are column percentages, except in the *N Missing* column, in which percentages of the total sample size that are missing values are displayed

		UC (Trt.1)	***SWCM*** (Trt.2)	***VSSP*** (Trt.3)	***p***	**Total**	**N missing**
N		87	88	90	--	265	0
Sex	*Female*	42, *48.3%*	48, *54.6%*	41, *45.6%*	0.47	131, *49.4%*	0
Age	Median, *SD*	66, *13.9*	66, *13.0*	68, *12.7*	0.52	67, *13.2*	0
	<55	20, *23.0%*	15, *17.1%*	17, *18.9%*	0.54	52, *19.6%*	
	55-75	47, *54.0%*	44, *50.0%*	51, *56.7%*		142, *53.6%*	
	>75	20, *23.0%*	29, *33.0%*	22, *24.4%*		71, *26.8%*	
Race	*White*	66, *75.9%*	68, *77.3%*	75, *83.3%*	0.53	209, *78.9%*	1, *0.4%*
	Black	18, *20.7%*	16, *18.2%*	10, *11.1%*		44, *16.6%*	
	Other	3, *3.5%*	4, *4.6%*	4, *4.4%*		11, *4.2%*	
Site	*Sparrow*	51, *58.6%*	50, *56.8%*	52, *57.8%*	0.97	153, *57.7%*	0
	St. Joe's	22, *25.3%*	25, *28.4%*	26, *28.9%*		73, *27.6%*	
	U-M	14, *16.1%*	13, *14.8%*	12, *13.3%*		39, *14.7%*	
Stroke severity	*Mild*	66, *75.9%*	60, *68.2%*	64, *71.1%*	0.14	190, *71.7%*	0
	Moderate	12, *13.8%*	21, *23.9%*	23, *25.6*		56, *21.1%*	
	Severe	9, *10.3%*	7, *8.0%*	3, *3.3%*		19, *7.2%*	
Discharge Destination	*Home*	49, *56.3%*	32, *36.4%*	36, *40.0%*	0.03*	117, *44.2%*	0
	Inpatient Rehab	29, *33.3%*	47, *53.4%*	48, *53.3%*		124, *46.8%*	
	SNF/SAR	9, *10.3%*	9, *10.2%*	6, *6.7%*		24, *9.1%*	
Health Insurance	*Private*	37, *42.5%*	27, *30.7%*	28, *31.1%*	0.58	92, *34.7%*	4, *1.5%*
	Medicare	42, *48.3%*	48, *54.6%*	51, *56.7%*		141, *53.2%*	
	Medicaid	7, *8.1%*	12, *13.6%*	9, *10.0*		28, *10.6%*	
Patient married**		32, *36.8%*	38, *43.2%*	35, *38.9%*	0.91	105, *39.6%*	81, *30.6%*
Education Level**	H.S. or less	22, *25.3%*	21, *23.9%*	21, *23.3%*	0.89	64, *24.2%*	81, *30.6%*
	College or more	36, *41.4%*	43, *48.9%*	41, *45.6%*		120, *45.3%*	
Caregiver consented to the trial		58, *66.7%*	57, *64.8%*	54, *60.0%*	0.64	169, *63.8%*	*NA*
Caregiver lives with patient***		43, *76.8%*	41, *78.9%*	36, *70.6%*	0.60	120, *75.5%*	10, *8.3%*
Days from hospital admit to final discharge (median, *SD*)		8, *11.7*	13, *9.7*	12.5, *11.2*	0.46	12, *10.9*	0

*Note significant difference in discharge destination between randomized treatment arms. Discharge destination describes whether patients were discharged directly home, to an inpatient rehab facility, or to a sub-acute rehab facility. Discharge destination is an indicator of a patient's wellness at the time of hospital discharge, and is thus likely linked to the severity of the stroke and the recovery trajectory of the patient.

Table 1 (cont'd)

**Marital status and education level were measured at the 7-day interview, while most other covariate and demographic data was measured at the initial enrollment. ***The count of patients with a live-in caregiver is only taken from within the consented caregiver population. As a result, all percentages in this row have a denominator that only includes consented caregivers.

Among the above covariates in *Table 1*, only discharge destination shows a significant difference between the three randomization arms; with this one exception, randomization was adequate at distributing the measured pre-treatment covariates equally across treatment groups. To further scrutinize the effectiveness of the randomization, we can examine various primary and secondary patient-reported outcome measures (PROMs) including quality of life scales collected at the seven-day interview. These scales are all based on a series of questions answered by the patient or, in some cases, by proxy (usually the patient's primary caregiver). It is important to note that these scores were all measured in an interview that occurred after participants had been randomized to their treatment groups, learned the treatment group to which they were assigned, and, in a few cases, already been visited by a social worker. As a result, these scores are not perfect reflections of baseline values, but are still valuable because they were taken prior to treatment beginning in earnest. *Table 2* below displays the seven-day PROMs (both primary and secondary) of participants in the trial.

Table 2: Primary and secondary outcome measures collected at the seven-day interviews by treatment group. P-values were calculated using one-way ANOVA.

	PROM	**Total** Median *(SD)*	***UC*** (1) Median *(SD)*	***SWCM*** (2) Median *(SD)*	***VSSP*** (3) Median *(SD)*	***p-value***	**N (%) missing**
MISTT Primary Outcomes	**Global 10 Physical Health**	42.0 (*5.8)*	42.1 (*5.6)*	42.1 (*5.9*)	41.6 (*5.8)*	0.06	45 (*17.0%)*
	Global 10 Mental Health	45.6 (*8.3)*	46.0 (*7.9)*	44.8 (*8.4*)	44.8 (*8.7)*	0.79	45 (*17.0%)*
	Patient Activation	60.6 (*16.1)*	63.1 (*16.7)*	55.6 (*15.1*)	60.6 (*16.4)*	0.25	55 (*20.8%)*

Table 2 (cont'd)

MISTT Secondary Outcomes	**Patient Emotional Support**	62.0 (*7.8)*	62.0 (*7.7)*	62.0 (*7.9*)	62.0 (*7.9)*	0.92	87 (*32.8%)*
	PHQ9 Depression	4 (*5.1)*	4 (*5.5)*	4 (*5.0*)	5 (*5.0)*	0.93	72 (*27.2%)*
	NeuroQoL Anxiety	50.0 (*8.5)*	50.0 (*9.2)*	48.0 (*7.8*)	51.6 (*8.4)*	0.22	85 (*32.1%)*
*Global-10 Physical Health, Global-10 Mental Health, Patient Emotional Support, and NeuroQoL Anxiety are all T-score metrics with a mean of 50 and a standard deviation of 10; Patient Activation is measured on a scale of 0 to 100; and PHQ9 depression is measured on a scale of 0 to 27.							

Note that among the scores displayed in *Table 2*, all one-way ANOVA p-values were above the 0.05 threshold for statistical significance, although Global-10 physical health is only slightly above ($p = 0.06$). Thus these scores, like the distributions of the pre-treatment covariates, appear to confirm the effectiveness of randomization in balancing the measured covariates and baseline quality of life outcome scores. However, even if covariates are distributed unevenly following randomization, they can be controlled for in analysis. The real goal of randomization is to ensure that *unmeasured* covariates are evenly distributed between our three treatment arms. While this exchangeability is impossible to prove, the covariate distributions and baseline seven-day interview scores do not indicate any cause for alarm when we assume exchangeability in the analysis.

Ultimately *MISTT* outcomes were reported using a difference-in-differences model. Despite randomization, small, non-significant differences between treatment arms existed at the seven-day interview (see *Table 2* above), and the difference-in-difference approach was selected because it is able to measure changes in outcomes over the course of the trial relative to the baseline seven-day values. *Table 3* shows the difference-in-differences results for primary outcomes reported in *MISTT* (which account for the 7-day baseline value), as well as the results of one-way ANOVA tests of differences in 90-day outcomes by treatment arm (which do not account for the 7-day baseline value). Pairwise differences in the 90-day outcomes by treatment arm relative to the control group are also shown. In a difference-in-differences analysis outcomes are shown as changes in an intervention group over the course of the trial

relative to the changes exhibited in the control group (usual care) over the course of the trial. Based on the difference-in-differences results there was a significant improvement in physical health shown in the *VSSP* treatment arm (social worker case management + access to online stroke support recovery resources). Additionally, there were increases significant at the 10% level (p = 0.06) for patient activation in the *VSSP* treatment arm relative to the referent group. When baseline values are discounted and we look solely at the 90-day outcomes, the only significant result was related to PAM; PAM scores were significantly lower in the SWCM group relative to UC.

Table 3: The primary outcomes of *MISTT*: (1) one-way ANOVA p-values testing for differences in 90-day outcome values by treatment arm; (2) pairwise differences in 90-day outcome by treatment arm, where both the *SWCM* and *VSSP* results are shown in reference to the Usual Care arm; and (3) difference-in-difference results, in which each treatment's change in outcome scores between the 7-day and 90-day interviews is compared to the change exhibited in the usual care group.

Outcome Measure	**One-Way ANOVA** for 90-day values by treatment	**90d outcome by Treatment Arm** (reference group = *UC* (*Trt.1*))		**Difference-in-difference** (reference group = *UC* (*Trt.1*))	
		SWCM (*Trt.2*)	***VSSP*** (*Trt.3*)	***SWCM*** (*Trt.2*)	***VSSP*** (*Trt.3*)
Global-10 Physical Health	$p = 0.45$	0.249 (–1.716, 2.286) $p = 0.78$	1.246 (–0.755, 3.247) $p = 0.22$	2.006 (–0.719, 4.730) $p = 0.15$	4.492 (1.760, 7.223) $p = 0.002$
Global-10 Mental Health	$p = 0.42$	–1.520 (–4.288, 1.248) $p = 0.28$	0.190 (–2.579, 2.958) $p = 0.89$	–1.375 (–5.077, 2.323) $p = 0.46$	1.054 (–2.666, 4.773) $p = 0.58$
Patient Activation Measure (PAM)	$p = 0.01$	–5.700 (–11.458, 0.058) $p = 0.05$	3.351 (–2.428, 9.129) $p = 0.26$	–1.800 (–9.197, 5.599) $p = 0.63$	6.998 (–0.379, 14.374) $p = 0.06$

As shown in *Table 3*, there were significant effects on patient health and recovery that were the result of *MISTT* interventions, but the mechanism through which these effects occur is not clear. Further analysis is required to elucidate this mechanism to improve more targeted interventions in the future. The purpose of this paper is to examine the pathways through which *MISTT* interventions act on physical and mental health.

Proposed Model and Justification

We propose that the *MISTT* treatments causally influence Global-10 physical and mental health through a sequentially mediated pathway involving emotional support and patient activation. Given that physical health and mental health are strongly correlated (Pearson correlation coefficient = 0.72) and there might be unmeasured common causes between the two outcomes, they will be included in the same conceptual model. We assume this because a patient's mental and physical health are obviously highly correlated, and are likely being simultaneously affected by other unmeasured variables that are not included in our models. We will include a set of pre-treatment covariates that are all plausible potential confounding variables (in the relationships between mediators and outcomes), as well as several important baseline health measures that were procured at the seven-day interview. Note that discharge destination is included in the model below as having a potential influence on treatment arm. This is due to the imbalance in the distribution of discharge destinations among treatment arms, as shown earlier in *Table 1*. Discharge destination refers to the destination of discharge for patients in *MISTT* – while the trial interventions did not begin until after patient were discharged home, many patients were not discharged directly home, but first discharged to an inpatient rehab or sub-acute rehab facility. Such a discharge is important because it is obviously related to the patient's health state at the time of hospital discharge. Discharge destination may therefore be causally influenced by the pre-treatment covariates, and it causally influences the seven-day baseline health measures. We are assuming that discharge destination does not influence the 90-day health measures (ie. the mediators and outcomes) directly, but rather has influence through the mediators of treatment arm and seven-day health measures.

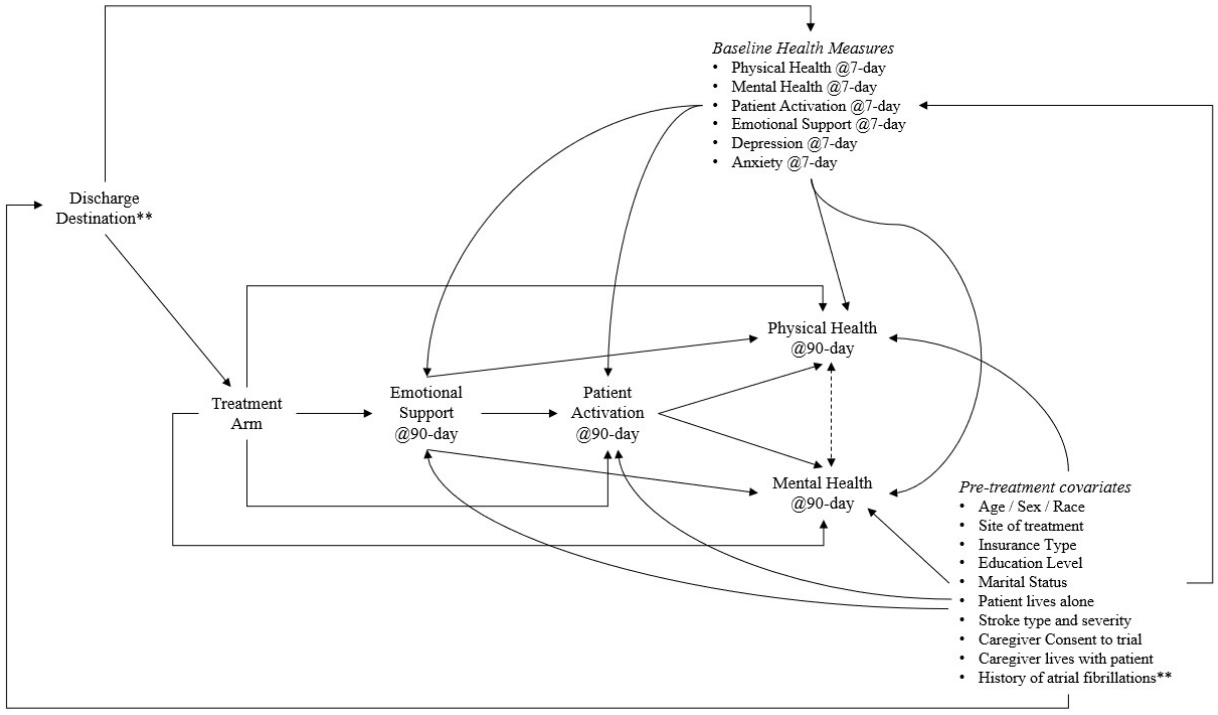

Figure 1: Proposed underlying causal model. Solid lines represent causal associations. Dashed lines represent potentially confounding relationships.

The outcome variables used in *MISTT* are patient-reported outcomes, and as such, are a reflection of the patients' own impressions of their health state. Therefore we would expect these measures to be correlated. *Table 4* below illustrates that this is indeed the case.

Table 4: Pearson Correlation coefficients among the measures of interest, from data compiled during the 90-day interview. All correlations are significant at $p < 0.01$.

Variable	*EMO-S*	*PAM*	*PHQ9*	*NEUROQOL*	*G10 PH*	*G10 MH*
EMO-S	—					
PAM	0.34	—				
PHQ9	–0.32	–0.22	—			
NQ-Anxiety	–0.27	–0.21	0.64	—		
G10 PH	0.38	0.43	–0.59	–0.54	—	
G10 MH	0.36	0.45	–0.56	–0.50	0.72	—

Note: EMO-S = Emotional Support. *PAM* = Patient Activation Measure. *PHQ9* = PHQ9 Depression Score. *NQ-Anxiety* = NeuroQoL Anxiety. *G10 PH* = Global 10 Physical Health. *G10 MH* = Global 10 Mental Health.

Additionally, given that the seven-day measurements were taken at the very start of the intervention, we essentially only have data reflecting the results of the trial available at a single time

point, the 90-day interview. Thus we are theorizing causal relationships between related measures taken at the same data point.

Patient Activation Measure (PAM) is a validated measure that tracks "an individual's knowledge, skills, and confidence to manage his/her health" [12]. It is important to note that PAM is not fixed over time, but rather fluctuates based on internal and external factors. Further, it has been shown that "with effective support, individuals can increase their level of activation over time" [12]. PROMIS Emotional support is a t-score that quantifies a study subject's perception of the level of emotional support they have in their life. We believe that there is a plausible pathway whereby treatment arm affects physical and mental health through a sequential mediation pathway of emotional support and patient activation. Thus the use of these two mediators in sequence is important here.

Additionally, our model is unusual in that it features two parallel outcomes. This is important given the strong correlation between physical and mental health ($r = 0.72$, *Table 4*). They are associated with each other, which is represented in *Figure 1* by a dashed line with double-sided arrows. We theorize that there are one or more unmeasured confounders (e.g., genetics) causally affecting both physical and mental health. Further, this requires an additional assumption: that none of the unmeasured confounders causally influencing physical and mental health during the 90-day interview are also causally preceding the pre-and post-treatment confounders. Although the pre-treatment covariates were recorded pre-hospital discharge and the post-treatment confounders were measured at the seven-day interview (and therefore both should temporally precede the unmeasured confounders linking 90-day physical and mental health), it is nonetheless conceivable that there are unmeasured confounders causally influencing 90-day physical and mental health that are also influencing seven-day interview outcomes.

Encoded Assumptions

Exchangeability. This is perhaps the most important of our assumptions. We are assuming that the composition of subjects randomized to each treatment arm is similar. If this assumption holds, there will be no unmeasured confounding between the exposure and any other variable. The distributions of

measured pre-treatment covariates (see *Table 1*) and baseline health measures (see *Table 2*) indicate that randomization was effective at bestowing similar distributions of the measured variables among the three treatments. The only notable exception is discharge destination, which we will need to account for in the analysis. However, while it seems reasonable to accept that none of these measured potential confounding variables are descendants of random treatment assignment, it is not possible to test whether the distributions of all unmeasured covariates are similarly well-distributed. Lastly, while treatment was randomized, given the structure of the study it was not possible to randomly assign the mediator values. Given this, our analysis must control for potential confounders in the relationship between each mediator and subsequent mediators and outcomes.

Positivity. The positivity assumption requires that each treatment group features all values of potentially confounding covariates in the population. Furthermore, it requires that all values of each mediator are represented in all treatment groups, any prior mediator groups, and subgroups defined by confounding covariates. The reasoning here is simple: in order to make inferences about the hypothetical outcomes we should expect based on treatment, mediator, and covariate values, we need to have each unique combination of treatment level, mediator value, and covariate characteristics represented in the data. Random assignment of treatment, and the resulting exchangeability of our treatment groups, should ensure that positivity is not violated.

Consistency. To compare outcomes in a causal study, the estimand of interest is the comparison of a subject's outcome under the treatment they actually received to that same subject's hypothetical outcome had they received a different treatment than they actually did, assuming all other factors remained fixed. Thus, for every subject in *MISTT*, one of the counterfactual outcomes in our comparison is their actual observed outcome, which is being examined relative to the outcome they would have observed had everything else remained equal, but their random treatment arm been different. While we are unable to do this on an individual level, due to the assumption of exchangeability, we can assume that our three treatment arms are composed of similar participants, and thus the mean observed outcomes of

any group reflect the mean counterfactual outcomes of members of both other groups, had they received the treatment in question.

No treatment variation. We assume that within each treatment group, the assigned treatment was equivalent. This is, of course, not true. Variability would be introduced by the effectiveness of different social workers, for example, and a multitude of other factors, and is impossible to completely remove in a pragmatic trial of this scope. However, in its most basic form, and averaged over entire treatment arms, we are assuming three different treatments, applied evenly within treatment group: 1) usual care, 2) social worker case management, and 3) social worker case management as well as access to the *MISTT* website.

Non-interference. We assume that the treatment assignment of one participant does not affect the outcomes of any other study participants. This is less of a concern in *MISTT* as there was no known social interaction between trial members. While it is conceivable that a shared social worker might act as a vector for influence between study subjects, there was no evidence of this in the trial.

Sequential Ignorability. We assume that *(a*) treatment was assigned independently of potential mediating and outcome variables in our model, and *(b*) mediators are independent of the potential outcomes given treatment assignment, pre- and post-treatment covariates, and prior mediators. Given that treatment was assigned randomly, and there is no reason to assume that randomization failed to produce exchangeability, assumption *(a)* is valid. However, assumption *(b)* requires additional thought. The most reliable way to ensure that sequential ignorability holds is to measure mediating variables temporally earlier than outcome variables; in that case, there would not be a plausible way for an outcome to be causally influencing a mediator that was measured at an earlier time point. However, the *MISTT* outcomes data were only measured at baseline (7-day) and the end of the trial (90-day). As a result, there is only one time point from which all outcome data (ie. both mediators and both outcomes) arises. So instead we rely on the plausibility of the mechanism. Even if measured at the same point in time it is more plausible that emotional support causally influences patient activation, which will in turn causally influence simultaneous changes in physical and mental health.

Data Missing at Random. There was a substantial amount of missing data in *MISTT* (see *Table 5* below). Data could not be assumed to be missing completely at random. Participant marital status, for example, was strongly associated with 90-day interview missingness, while race and whether or not the caregiver consent, obtained during enrollment, were both associated with 7-day interview missingness. However, there is no evidence that missingness was associated with treatment arm (*Table 5*). Additionally, there is no evidence that outcome or mediator values were systematically different for those participants with missing observations. We updated the DAG shown in *Figure 1* to include missing data mechanisms (see *Figure A1*).

Table 5: Missingness by treatment arm for important covariates, baseline health measures, and mediator and outcome measures that exhibited high frequencies (and percentages) of missing observations. Note that there is no association between missingness and treatment arm.

Covariate	**Treatment Arm**			**P-value**	**Total N (%)**
	Usual Care N (%)	**Social Worker N (%)**	**Social Worker + Website N (%)**		
*Caregiver lives with patient[1]	2 *(3.5)*	5 *(8.8)*	3 *(5.6)*	0.64	10 *(5.9)*
Education level[1]	29 *(33.3)*	24 *(27.3)*	28 *(31.1)*	0.68	81 *(30.6)*
Marital Status[1]	29 *(33.3)*	24 *(27.3)*	28 *(31.1)*	0.68	81 *(30.6)*
Physical and Mental health @7-day[2]	18 *(20.7)*	14 *(15.9)*	13 *(14.4)*	0.51	45 *(17.0)*
Patient Activation @7-day[2]	21 *(24.1)*	14 *(15.9)*	16 *(17.8)*	0.58	55 *(20.8)*
Emotional Support @7-day[2]	33 *(37.9)*	26 *(29.6)*	28 *(31.1)*	0.45	87 *(32.8)*
Physical and Mental Health @90-day[3]	17 *(19.5)*	16 *(18.2)*	18 *(20.0)*	0.95	51 *(19.3)*
Patient Activation @90-day[4]	21 *(24.1)*	19 *(21.6)*	22 *(24.4)*	0.89	62 *(23.4)*
Emotional Support @90-day[4]	33 *(37.9)*	30 *(34.1)*	31 *(34.4)*	0.84	94 *(35.5)*

1) Pre-treatment covariates. 2) Baseline health measures. 3) Outcomes. 4) Mediators.
*Caregiver lives with patient variable is only available among the patients with consented caregivers.

Identification

We used a version of *Figure 1* that was extended to account for missing data. This more complex DAG is shown in the Appendix *(Figure A1)*. As shown above in *Table 5*, mediators, outcomes, baseline health measures, and select covariates all included missing data. Additionally, a participant's probability of missing their seven-day interview was strongly associated with an increased probability of also missing their 90-day interview (see *Table 6; Table A1*). All 7-d or 90-d interviews generated a physical and mental health t-score; however, some participants opted for an interview by proxy or an abbreviated interview, which collected data on the primary outcomes but little else.

Table 6. Missingness frequencies between the two interviews.

	90-day interview	
7-day interview	Non-missing	Missing
Non-missing	194	26
Missing	20	25

In other words, the missingness mechanisms of mediators, outcomes, baseline health measures, and pre-treatment covariates were all associated with each other. Missingness at the seven-day interview, which includes marital status and education history, are strong predictors of missingness of 90-day measures. A visual illustration of these missing data mechanisms is shown in *Figure A1*, along with the identification assumptions that follow.

Given that we found no evidence to suggest that the data was not missing at random, we can use multiple imputation to simplify identification of effects, while still yielding consistent estimates[16]. After imputing missing data, our identification of effects will simplify to what is shown below.

$M_1(a) \perp A \mid DD$

$M_2(a, m_1) \perp M_1 \mid A, C, Y_7$

$M_2(a, m_1) \perp A \mid C, M_1, Y_7$

$Y_1(a, m_1, m_2), Y_2(a, m_1, m_2) \perp M_1 \mid A, C, M_2, Y_7$

$Y_1(a, m_1, m_2), Y_2(a, m_1, m_2) \perp M_2 \mid A, C, M_1, Y_7$

$Y_1(a, m_1, m_2), Y_2(a, m_1, m_2) \perp A \mid C, M_1, M_2, Y_7$ (direct effect)

$Y_1(a, m_1, m_2), Y_2(a, m_1, m_2) \perp A \mid DD$ (total effect)

The notation above is used to describe the minimum conditioning set based on the assumed DAG. The first line above would read that "M_1 is independent of A conditional on DD." In other words, we can estimate the effect of A on M_1 by adjusting for discharge destination in our model (Note: A = Treatment arm; C = pre-treatment covariates; Y_1 = Global-10 Physical Health @ 90-day; Y_2 = Global-10 Mental Health @ 90-day; M_1 = Emotional Support @ 90-day; M_2 = Patient Activation @ 90-day; Y_7 = Wellness Outcomes @ 7-day; DD = discharge destination).

Causal Mediation Background

In its most basic form, mediation takes the following form (*Figure 2*), where Pathway *c* represents the direct effect of exposure X_i on outcome Y_i, while Pathways *a* and *b* represent the indirect effect of the exposure on the outcome, by way of a mediating variable M_i. The total effect of *X* on *Y* is the

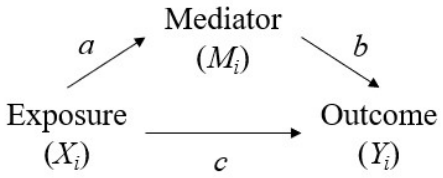

Figure 2: Illustration of the most basic form of a mediating variable, in which X_i has a direct causal relationship with Y_i, but also causally influences Y_i indirectly through M_i.

sum of pathways *a*, *b*, and *c*. The most widely-cited early paper on causal mediation analysis was written by Baron and Kenny[3], who outlined three criteria necessary to define a variable as a mediator: a significant total effect (Equation 1) of the exposure on the outcome; a significant effect of the mediator on the outcome (or indirect effect) (Equation 3); and a significant effect of the exposure on the mediator (Equation 2)[3]. These conditions were to be expressed in three regression equations:

$$Y_i = \beta_{0u} + \beta_{1u}X_i + \varepsilon_{ui} \qquad \text{Equation 1}$$

$$M_i = \alpha_0 + \alpha_1 T_i + \varepsilon_{Mi} \qquad \text{Equation 2}$$

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 M_i + \varepsilon_i \qquad \text{Equation 3}$$

These equations represent Baron and Kenny's three conditions for a mediation effect. M_i is considered a mediator if (1) $\beta_{1u} \neq 0$ (outcome is associated with exposure), (2) $\alpha_1 \neq 0$ (potential mediator is associated with exposure), and (3) $\beta_2 \neq 0$ (outcome is associated with potential mediator conditional on exposure). To generate confidence intervals around the mediated effect, Baron and Kenny recommended calculating standard errors using Sobel's formula[3].

The widespread influence of Baron and Kenny's approach cannot be overstated, but as more research in mediation has been undertaken, several critiques have taken hold and the field has advanced beyond their initial approach. First of all, there does not need to be a significant direct effect of the exposure on the outcome – all that is truly required for a mediating effect is a significant indirect effect of an exposure on an outcome[20]. Secondly, the Sobel formula is based on the assumption that the sampling distribution of the indirect effect is normal, when it is in fact that product of two normal distributions[6], and even if this (poor) assumption does not invalidate the use of Sobel's standard errors in this context, bootstrapping has been shown to be a more powerful method for the calculation of standard errors[6,20]. Additionally, in complex, pragmatic applications of mediation analysis, we are often not merely interested in estimating the effect of a single mediator. A flexible approach that allows for the possibility of multiple mediators, operating either sequentially or in parallel, is necessary.

Mediation Analysis in a Randomized Controlled Trial. Lastly, mediation analysis in an RCT is slightly different than its observational counterparts and it bears a quick discussion. First of all,

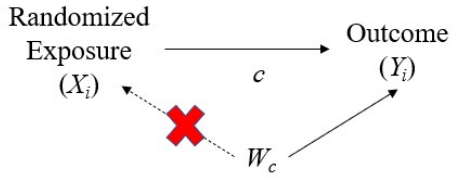

Figure 3: Illustration of the relationship between an exposure and outcome when the exposure is randomized. Note that due to randomization of the exposure, the exposure is independent of potential confounder W_c.

randomization of exposure removes the risk of bias that could be introduced by potentially confounding variables that are unevenly dispersed between treatment arms. Thus in an analysis considering the simple

linear effect of an exposure on an outcome, randomization eliminates the risk of confounding in Equation 2. Consider *Figure 3*, in which the relationship between X and Y is confounded by W_c. In this simple relationship, randomization of X ensures that X is independent of W, and thus that any effect of X on Y is a direct effect.

Now consider *Figure 4*. The relationships between exposure and mediator, mediator and

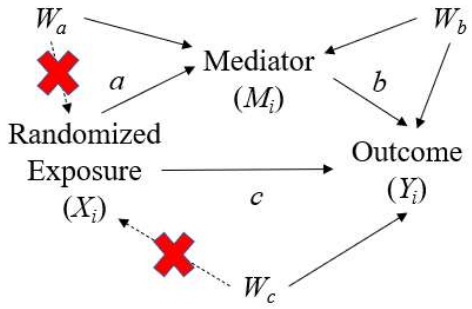

Figure 4: Illustration of a simple mediator relationship in which W_a, W_b, and W_c confound the relationships between exposure, mediator, and outcome. Randomization of exposure X ensures that X is independent of W_a and W_c. However, W_b. must still be controlled for when looking at the effect of mediator M_i on outcome Y_i.

outcome, and exposure and outcome each feature unique confounding variables W_a, W_b, and W_c respectively. Randomization of exposure X ensures that X is independent of both W_a and W_c, or, in other words, removes the potential for confounding in the exposure – mediator and exposure – outcome relationships. However, randomization of exposure does not account for potential confounders in the mediator – outcome relationship, here represented as W_b. Therefore, even when working with a randomized treatment, we still need to account for potential confounders in the relationship between the mediators and outcome.

Natural Effects Model

The natural effects approach to mediation analysis has gained popularity in recent years. It involves controlling the value of the exposure, while allowing the mediator(s) to subsequently take on the values that they naturally would have under the stated value of the exposure[9,11,15]. Direct and mediated

effects can then be calculated based on a process that Steen[15] refers to as "invoking nested counterfactuals." For example, say we are interested in estimating the indirect effect of a binary Treatment A on Outcome Y through a binary mediator M. We could calculate the outcome value when A=1. Then, leaving A=1, and we estimate the effect of changing the value of M from M = [the value that M naturally takes when A=1], to M = [the value M naturally takes when A=0]. In natural effect notation, this natural indirect effect (NIE) is defined as:

$$Y(a=1, M(a'=1)) - Y(a=1, M(a'=0)).$$

Note that this can be extended indefinitely; in a situation with 2 sequential mediators, we could examine the indirect effect of A through mediator 2 as shown below.

$$Y(a=1, M_1(a'=1), M_2(a''=1, (M_1(a'=1))) - Y(a=1, M_1(a'=1), M_2(a''=0, (M_1(a'=1))).$$

Essentially, we are keeping the exposure set to its original level, and mediator 1 set to the level it would have naturally taken under the set exposure level, and thus blocking the effects of treatment and mediator 1 on mediator 2. Then we can calculate the effect of a change in mediator 2 on outcome Y that is a result of the effect of A on mediator 2. In our methods section below, we extend this strategy to include 3 possible treatment levels.

METHODS

Imputation of Missing Data

We performed multiple imputation of missing values using chained equations (ref). Missing values of categorical variables were imputed using the discriminant function method, while continuous variable missing values were imputed using predictive mean matching. These methods were chosen so that imputed values were restricted to values that were already present in the dataset. 25 imputed datasets were created. Auxiliary variables used as predictors of missing data included sex, race, hospital site, health insurance status (private, Medicare, or Medicaid), whether the caregiver consented to the trial, history of atrial fibrillations, discharge destination, severity of stroke, and the number of days between hospital admission for stroke and discharge to home. This process generated 25 datasets of complete data. We then estimated our effects for each of these imputed datasets; final effects reported will be the means of 25 effect estimates and standard errors are estimated by bootstrapping. All analyses were done using SAS statistical software (version 9.4); imputation was done using PROC MI.

Estimation of Effects

Our estimation approach was based on the Natural Effects method described by Steen et al[15]. This technique is explained in detail in the series of steps below. Note that this only describes the methods of the final technique used. The ultimate choice of model is shown below in *Figure 5*. The reasoning behind the decision to pare down the initial conceptual DAG (*Figure 1*) into a simpler working model is described in the Appendix.

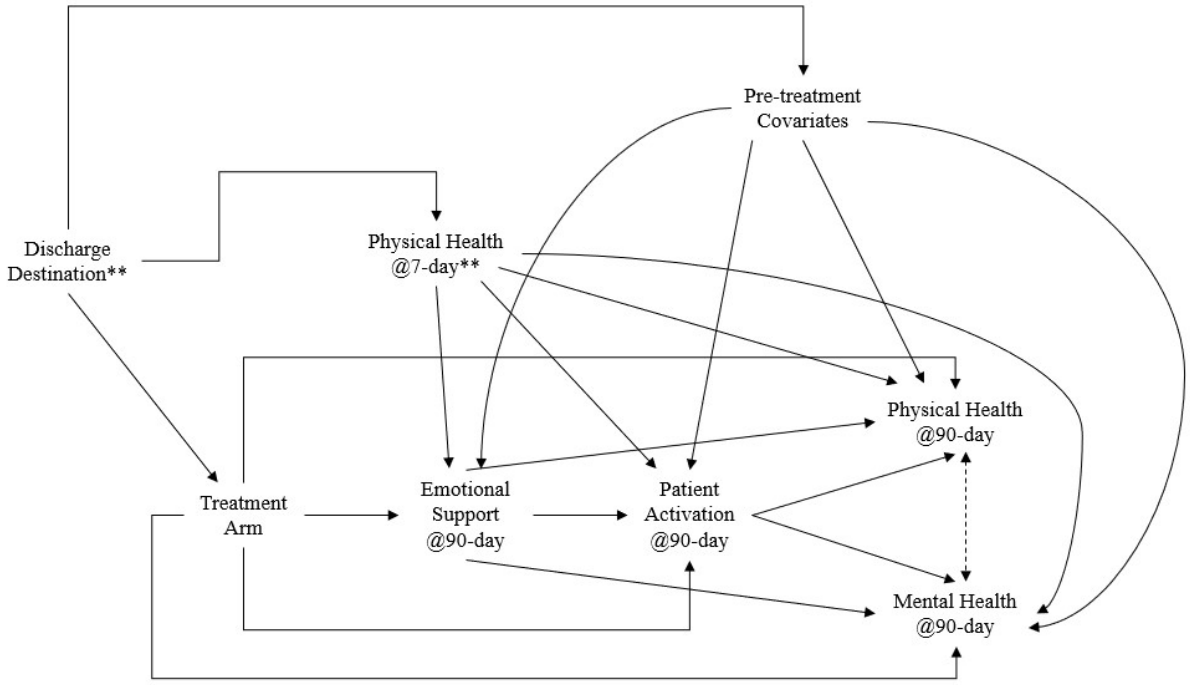

Figure 5: Simplified working DAG that we hypothesized for our final model. Physical health was the only 7-day wellness score that showed any significant difference between treatment group, so the other 7-day wellness scores were eliminated from the model. Additionally, after testing for interaction effects between treatment arm and the mediators, we were able to eliminate interaction effects from the model as well.

1. We fit a model for mediator 1 (emotional support), conditional on treatment arm and potential confounding variables. Given that treatment arm was randomized, we only included discharge destination ($p = 0.03$: abbreviated as D below) and seven-day physical health ($p = 0.06$: abbreviated as Y_7 below) as potential confounders in this model. These two variables were selected empirically due to slight imbalances between randomized treatment arms.

$$E(M_{1i} \mid A = a_i, D = d_i, Y_7 = y_{7i})$$

We stored the estimates (ie. the β's) from this M_1 model for later use. This step is necessary for the later calculation of weights, and can actually be done by modeling either mediator conditional on treatment arm, earlier mediators, and confounding variables. While our primary results display weights calculated by modeling M_1, we replicated the analysis using weights calculated based on an M_2 (patient activation)

model to assess the sensitivity of our results to the weights used. Note that this methods description is based on our use of a working model of M_1; when using M_2 in this step several of the later steps vary slightly.

2. We fit a model for outcome 1, Global-10 physical health (Y_{90P}), conditional on treatment arm, M_1, M_2, pre-treatment covariates (C), and seven-day physical health. We stored the estimates from this model for later use.

$$E(Y_{90Pi} \mid A = a_i, M_1 = m_{1i}, M_2 = m_{2i}, C = c_i, Y_7 = y_{7i},)$$

In this step we considered interactions between treatment arm and both mediators, which were ultimately disregarded because they were non-significant and rendered the model unwieldy (see Appendix: Model Refinement for a description of this process).

3. We replicated step 2 for outcome 2, Global-10 mental health (Y_{90M}), rather than physical health.

$$E(Y_{90Mi} \mid A = a_i, M_1 = m_{1i}, M_2 = m_{2i}, C = c_i, Y_7 = y_{7i},)$$

4. We constructed an extended dataset that required replicating our observed dataset k^2 times, where k is the number of treatment levels. In our case, treatment involved three levels; thus our dataset was replicated 9 times. Three new auxiliary variables were created in this extended dataset: a_0, a_1, and a_2. These will later be used to represent the possible counterfactual values that mediators might take conditional on treatment level and prior mediators. The k^2 replicates are necessary so that each participant will have all possible value combinations of a_0, a_1, and a_2 represented.

In order to construct this extended dataset, we let a_0 take on the value of the observed treatment arm in the first replication; in the second and third replications, a_0 takes on both of the other possible treatment values. This sequence is repeated 3 times for a_0 within each observation. We let a_1 assume the actual treatment value for the first three replicates, and then the other possible treatment values for replicates 4-6 and 7-9. We set a_2 equal to the observed treatment level for all replicates.

An example of this extended dataset is shown below.

patient	replicate	treatment arm	a0	a1	a2	Y1	Y2
1	1	2	0	0	2	33.1	33.9
1	2	2	1	0	2	33.1	33.9
1	3	2	2	0	2	33.1	33.9
1	4	2	0	1	2	33.1	33.9
1	5	2	1	1	2	33.1	33.9
1	6	2	2	1	2	33.1	33.9
1	7	2	0	2	2	33.1	33.9
1	8	2	1	2	2	33.1	33.9
1	9	2	2	2	2	33.1	33.9
2	1	0	0	0	0	41.7	42.4
2	2	0	1	0	0	41.7	42.4
2	3	0	2	0	0	41.7	42.4
2	4	0	0	1	0	41.7	42.4
2	5	0	1	1	0	41.7	42.4

Figure 6: Example of extended dataset to demonstrate how the values of a_0, a_1, and a_2 should be set.

4. We calculated regression weights using the following formula.

$$w_i = \frac{f(M1|A=a1,C)}{f(M1|A=a2,C)}$$

In other words, the numerator is the expected value of a patient's emotional support score (M_1) under counterfactual treatment a_1, given their actual observed covariates. The denominator is the expected value of a patient's emotional support score (M_1) given their actual treatment (a_2) and observed covariates.

5. We imputed the hypothetical outcome under counterfactual exposure level a_0 for each row in the extended dataset, conditional on hypothetical treatment a_0, M_1, M_2, pre-treatment covariates (C), seven-day physical health, and the weights calculated in the previous step. Note that this step was replicated for both outcomes in our study.

$$E(Y_{PHi} \mid A = a_{0i}, M_1 = m_{1i}, M_2 = m_{2i}, C = c_i, Y_7 = y_{7i}, W = w_i)$$

$$E(Y_{MHi} \mid A = a_{0i}, M_1 = m_{1i}, M_2 = m_{2i}, C = c_i, Y_7 = y_{7i}, W = w_i)$$

6. We fit our natural effects model of interest by regressing the imputed outcomes of interest (calculated in step 5) conditional on a_0, a_1, and a_2, weighted by w_1.

$$Predicted\ Y = \beta_0 + \beta_1 a_0^1 + \beta_2 a_0^2 + \beta_3 a_1^1 + \beta_4 a_1^2 + \beta_5 a_2^1 + \beta_6 a_2^2 + \varepsilon$$

7. We calculated effects of interest by combining the β's above. The effects of interest calculations are displayed below for the calculation of the effect of treatment=1 relative to treatment=0.

$$Total\ Effect_{A \to Y}\ (A=1) = \beta_1 + \beta_3 + \beta_5$$

$$Natural\ Direct\ Effect_{A \to Y}\ (A=1) = \beta_1$$

$$Natural\ Indirect\ Effect_{A \to M1 * Y}\ (A=1) = \beta_3$$

$$partial\ Natural\ Indirect\ Effect_{A \to M2 \to Y}\ (A=1) = \beta_5$$

This concludes our stepwise procedure for the analysis of effects. The above analysis was conducted for each of the 25 imputed datasets. The reported estimates will be our mean effect across the 25 imputed datasets.

Generation of Confidence Intervals

From the imputed datasets we calculated the between-estimate variances for each of the effects (b_m) of interest.

$$V_{bw} = \frac{1}{25}\sum_{m=1}^{25}(\bar{b} - b_m)^2$$

We applied bootstrapping (1000 replicates) to each of the 25 imputed datasets, and pooled the results from each dataset. This generated distributions of 1000 estimates from each of the imputed datasets, from which we could extract standard deviations (w_m) from each estimate distribution. The within-estimate variances were calculated for each effect of interest as shown below.

$$V_{wi} = \frac{1}{25}\sum_{m=1}^{25}(se(w_m))$$

The between-estimate and within-estimate variances for each estimate were used to calculate total variance.

$$V_{tot} = V_{wi} + (1 + \frac{1}{25})V_{bw}$$

Thus we were able to generate bootstrapped confidence intervals around each of the effect estimates.

$$b - z_{0.975}\sqrt{\frac{1}{V_{tot}}};\ b + z_{0.975}\sqrt{\frac{1}{V_{tot}}}$$

RESULTS

Effect estimates from our analysis are displayed in *Table 7* and *Table 8* below. Weighting by the ratio of expected values of M_1 (w_1) yielded the estimates in *Table 7*, while weighting by the ratio of expected values of M_2 (w_2) yielded the results in *Table 8*. None of the results obtained in either of these tables differ significantly from the null. However, there were some sizable effects. The natural direct effect of *VSSP* treatment (trt.2) on physical health is positive (1.395, 95% CI –0.560, 3.350) relative to the usual care group. Additionally, the *SWCM* treatment (trt.1) exhibited the same unexpected relationship with mental health in this analysis as in the primary analysis: the total effect of *SWCM* treatment on mental health was negative (–1.563, 95% CI –3.957, 0.830), which appears to largely be the result of the partial natural indirect effect of treatment on mental health through patient activation (M_2), which is also negative (–1.416, 95% CI –3.362, 0.529).

Table 7: Summary of total, direct, and indirect effects from our analysis using imputed datasets. These effects were calculated using the w_1 weighting scheme, where weights were calculated based on modeling M_1 conditional on treatment. Note that the "usual care" treatment arm is the reference group for all estimates below.

Outcome	**Component**	**SWCM (Trt.2)**		**VSSP (Trt.3)**	
		Estimate	*95% CI*	*Estimate*	*95% CI*
Physical Health T-score	*(TE)* $E(X_{A \rightarrow Y1})^a$	–0.304	–2.372, 1.763	0.876	–1.264, 3.017
	(NDE) $E(X_{A \rightarrow Y1})^b$	0.615	–1.319, 2.548	1.395	–0.560, 3.350
	(NIE$_{M1}$) $E(X_{A \rightarrow M1*Y1})^c$	0.031	–0.933, 0.995	0.138	–0.780, 1.056
	(pNIE$_{M2}$) $E(X_{A \rightarrow M2 \rightarrow Y1})^d$	–0.950	–2.630, 0.730	–0.657	–2.341, 1.028
Mental Health T-score	*(TE)* $E(X_{A \rightarrow Y2})$	–1.563	–3.957, 0.830	0.158	–2.420, 2.736
	(NDE) $E(X_{A \rightarrow Y2})$	–0.182	–2.447, 2.082	0.527	–1.949, 3.004
	(NIE$_{M1}$) $E(X_{A \rightarrow M1*Y2})$	0.035	–1.045, 1.116	0.168	–0.863, 1.199
	(pNIE$_{M2}$) $E(X_{A \rightarrow M2 \rightarrow Y2})$	–1.416	–3.362, 0.529	–0.537	–2.513, 1.440

a. The total effect difference of treatment arm (A) on physical health (Y_1).
b. The joint natural direct effect of treatment arm (A) on physical health (Y_1).
c. The joint natural indirect effect of treatment arm (A) on physical health (Y_1) with respect to emotional support (M_1).
d. The partial indirect effect of A on Y_1 with respect to M_2 (patient activation).

The above results were calculated using a weighting scheme based on emotional support, M_1. The same process was done using a weight calculation based on the expected value of patient activation, M_2.

Those results are displayed in *Table 8* below. Note that the estimates of total effect and natural direct effect are similar between the two weighting schemes. That said, these effects are parsed into indirect effects quite differently depending on the weighting scheme used. For example, the total effect of *SWCM* treatment on mental health is similar for both weighting schemes ($TE_{w1} = -1.563$, $TE_{w2} = -1.562$). However, the model that uses weights based on emotional support (*Table 7*) ascribes most of this effect to an indirect effect running through patient activation, whereas the model using weights based on patient activation (*Table 8*) ascribes most of this effect to an indirect effect running through emotional support.

Table 8: Summary of total, direct, and indirect effects from our analysis using imputed datasets. These effects were calculated using the w_2 weighting scheme, where weights were calculated based on modeling M_2 conditional on treatment and M_1. Note that the "usual care" treatment arm is the reference group for all estimates below.

Outcome	**Component**	**SWCM (Trt.2)**		**VSSP (Trt.3)**	
		Estimate	*95% CI*	*Estimate*	*95% CI*
Physical Health	*(TE)* $E(X_{A \rightarrow Y1})$	–0.319	–2.386, 1.748	0.885	–1.285, 3.055
	(NDE) $E(X_{A \rightarrow Y1})$	0.615	–1.319, 2.548	1.395	–0.560, 3.350
	(NIE_{M1}) $E(X_{A \rightarrow M1*Y1})$	–1.032	–2.791, 0.728	–0.422	–2.239, 1.395
	($pNIE_{M2}$) $E(X_{A \rightarrow M2 \rightarrow Y1})$	0.098	–0.664, 0.859	–0.088	–0.853, 0.677
Mental Health	*(TE)* $E(X_{A \rightarrow Y2})$	–1.562	–3.948, 0.824	0.181	–2.433, 2.794
	(NDE) $E(X_{A \rightarrow Y2})$	–0.182	–2.447, 2.082	0.527	–1.949, 3.004
	(NIE_{M1}) $E(X_{A \rightarrow M1*Y2})$	–1.533	–3.588, 0.523	–0.210	–2.357, 1.938
	($pNIE_{M2}$) $E(X_{A \rightarrow M2 \rightarrow Y2})$	0.153	–0.766, 1.071	–0.137	–1.059, 0.785

Differences in weighting scheme lead to small differences in the total effect and natural direct effects of treatment on the outcome. However, the choice of weighting scheme drastically alters the estimates for the indirect effects in all cases.

In addition to analyzing imputed data, we also conducted an available case analysis, which is shown in *Table 9* below. The data is generated by weighting on M_1 (similar to *Table 7*). While the estimates from the available case analysis were generally larger in magnitude than with the imputed data (*Table 7*), the confidence intervals were larger as well. One feature of the available case analysis that was different than in the imputed data was sizable negative values in both treatment arms for the partial indirect effect of treatment on physical health through patient activation ($pNIE_{M2\,(SWCM)} = -1.750$;

$pNIE_{M2\,(VSSP]}$ = –1.578). These partial indirect effects in both treatment arms largely negated natural direct effects of treatment on physical health in both the treatment groups (NDE_{SWCM} = 1.604; NDE_{VSSP} = 2.031). This illustrates a larger point of mediation analysis. There does not necessarily need to be "an effect to be mediated". An effect that appears absent on its first analysis can in reality be the sum of two competing indirect effects.

Table 9: Summary of total, direct, and indirect effects from an analysis in which only available cases were used, and the weighting scheme was based on modeling M_1 conditional on treatment.

Outcome	**Component**	**SWCM (Trt.2)**		**VSSP (Trt.3)**	
		Estimate	*95% CI*	*Estimate*	*95% CI*
Physical Health	*(TE)* $E(X_{A \rightarrow Y1})$	0.013	–2.525, 2.551	0.836	–1.697, 3.370
	(NDE) $E(X_{A \rightarrow Y1})$	1.604	–0.812, 4.021	2.031	–0.314, 4.376
	(NIE_{M1}) $E(X_{A \rightarrow M1*Y1})$	0.159	–1.138, 1.455	0.383	–0.890, 1.657
	($pNIE_{M2}$) $E(X_{A \rightarrow M2 \rightarrow Y1})$	–1.750	–3.994, 0.493	–1.578	–3.770, 0.614
Mental Health	*(TE)* $E(X_{A \rightarrow Y2})$	–2.913	–5.972, 0.145	–0.910	–4.036, 2.215
	(NDE) $E(X_{A \rightarrow Y2})$	–1.146	–4.109, 1.817	–0.383	–3.442, 2.676
	(NIE_{M1}) $E(X_{A \rightarrow M1*Y2})$	0.274	–1.396, 1.944	0.659	–0.960, 2.279
	($pNIE_{M2}$) $E(X_{A \rightarrow M2 \rightarrow Y2})$	–2.041	–4.604, 0.521	–1.186	–3.655, 1.283

DISCUSSION

Ultimately our results did not support our hypothesized sequentially mediated pathways between treatment and physical and mental health. However, while we found no total, direct, or indirect effects of strong statistical significance, our results do suggest the presence of some sizable potential effects. Specifically, the *VSSP* treatment appears to exert a positive direct effect on physical health compared to usual care. On the other hand, the *SWCM* treatment appears to have a negative total effect on patient mental health, an effect which appears is explained by an indirect effect through patient activation.

These were the results that we obtained from the model that we deemed most plausible, which included no treatment-mediator interactions, controlled for discharge destination and 7-day physical health, and calculated weights based on the first mediator, emotional support. That said, it is important to note that many model variations were run in this analysis. In addition to this final model, we tried a weighting scheme featuring weights calculated based on the expected value of M_2, and while the total and direct effects were similar between these strategies, the indirect effects were vastly different. Similarly, we ran a model that featured interaction effects between treatment and mediators, and found, that a) no significant treatment-mediator interaction effects existed in our data, and b) that the inclusion of these non-significant interaction terms made the model unwieldy; our comparison of results in the interaction models between the w_1 approach and the w_2 approach yielded vastly different total, direct, and indirect effects. Additionally, while six seven-day health measures were initially included as potential confounders of effects between M_1, M_2, and the outcomes, we eventually excluded all but seven-day physical health, which showed a marginal imbalance between treatment arms that we felt needed to be accounted for in analysis. The others, which were well-balanced between treatment arms (*Table 2*), were discarded from the model; given that the mediators were measured at the 90-day interview, seven-day measures were not plausible potential confounders in the relationship between mediators and outcomes.

Randomization of treatment arms made this a simpler analysis than it would have been if treatment had been assigned naturally in an observational study. We were able to assume exchangeability

between treatment arms. Ultimately this is why it was logical to favor a weighting scheme based on the expected values of M_1 rather than M_2. There are far fewer assumptions that need to be made when weights are calculated based on the relationship between treatment group (A) and emotional support (M_1), as opposed to the relationship between A and M_2. Because of randomization, we assumed that there were no confounding variables in this relationship besides those that we empirically noticed were different between treatment arms. As it happened, there were two variables that displayed imbalance and were thus exceptions to this theoretical exchangeability, both of which we conditioned on when modeling the relationship of A and M_1.

The first exception to the exchangeability of treatment arms assumption was discharge destination. Discharge destination is an important indicator for a patient's condition at the time of hospital discharge. However, it is also likely strongly associated with factors such as the severity of the stroke, the patient's level of health prior to having the stroke, covariate data related to the home environment (ie. is the primary caregiver physically capable, does primary caregiver live with patient, etc.), and factors such as health insurance coverage, age, or even the practices of the discharging hospital. This makes it a prime confounding candidate. And yet, despite randomization that occurred at the time a patient was discharged home (ie. after a potential stay in SNF or IRF), patients who were discharged directly home were significantly more likely to be randomized to the usual care treatment arm, while patients who were discharged to inpatient rehab were significantly more likely to be randomized to either the *SWCM* or *VSSP* treatment arms. Nevertheless, it was a simple matter to control for this imbalance in our analysis.

The second of these exceptions is seven-day physical health. Participants were randomized to a treatment upon their discharge home, and then interviewed seven days later. By the time of their interview, they had learned their randomized treatment arm and, in some cases, already been visited by a social worker. Nonetheless, we think of this as a baseline measure because the intervention had not begun in earnest at this point, so we would not expect an effect on physical or mental health due to treatment arm. What is interesting is that the participants who were randomized to the social worker + website treatment arm (*VSSP*) had significantly lower seven-day physical health scores than those in the other two

treatment arms. We wouldn't expect a difference of this magnitude in the seven-day values in either direction, but if there were to be an imbalance, it seems more likely that those who knew they were randomized to the most favorable treatment arm to perceive their physical health as better. It is possible that this is an effect of the imbalance of discharge destination between treatment arms – patients discharged from inpatient rehab were more likely to have lower physical health scores in the seven-day measure. Perhaps a visit by a social worker served to highlight to patients the limitations of their own physical ability at that point in time, or just the knowledge of their own randomized treatment arm assignment somehow biased their perception of their own physical health.

The breadth of *MISTT* data available from each interview was a strength of this study. In addition to basic demographics, data was available for a wide variety of validated wellness measures, including institutional support, informational support, anxiety, emotional and behavioral dyscontrol, self-reported modified Rankin score, and depression, to name a few. The sheer volume of data available for a high percentage of the study participants enrolled in the trial make it a gold mine for future mediation analyses. Furthermore, our sequential mediation analysis should provide a useful blueprint for conducting similar studies in the future. Given that no strongly significant mediation effects were found in our analysis, the most logical next step would be to look for mediation effects of other available wellness scores. The *MISTT* primary results reported a significant positive effect of *VSSP* treatment on patient physical health. Our results show that this effect was not mediated by either emotional support or patient activation; however, it is still worth searching for the mechanism of this improvement, and there is data available to do so.

While the volume of data recorded in *MISTT* was a strength, the trial design was a challenge because it was not designed with mediation in mind; had the trial had been designed for the purpose of an eventual mediation analysis, outcome data would have been recorded at intermediary time points between baseline and the 90-day conclusion to the trial. Ideally each mediator would be recorded at its own time point to guarantee sequential exchangeability. However, *MISTT* was meant to be a pragmatic implementation of an ambitious intervention, and as such, opted for two intensive interviews, one at

baseline and one at the conclusion, to measure changes in many different metrics over the course of the entire trial, rather than multiple short interviews done sequentially. In order to work around this particular study design, we used mediators that were measured at the same point in time as the outcomes, and relied on literature and plausibility to assume that it was likely that emotional support causally influenced patient activation which causally influenced physical and mental health. While it is not disqualifying that the sequential mediators are recorded at the same time point, it would strengthen the analysis to have recorded patient mediators at different time points to ensure that sequential exchangeability was not violated.

In conclusion, our analysis demonstrates that in the Michigan Stroke Transitions Trial, the reported positive effect of the *VSSP* treatment on patient physical health was not mediated by either the patient's perceptions of their emotional support or the patient's level of activation. While we ascribe this to a direct effect of treatment on outcome, it is possible that this effect is mediated by a different wellness variable that we have not pinpointed yet. Further investigation could elucidate this pathway, which might help to inform future interventions. Secondly, we found that there was a possible negative effect of the *SWCM* treatment on patient mental health, an effect which appeared to be mediated by patient activation. This effect is only of significance at a 10% level, but it would be worth investigating further if a social worker intervention could in some instances result in lower PAM scores, implying that the patient is less empowered to solve their own health problems. And finally, we hope that our analysis provides a useful blueprint in conducting a sequential mediation analysis using a Natural Effects Model, that incorporates both multiple imputation techniques to account for missing data, and bootstraps these imputed datasets to generate standard errors.

APPENDIX

APPENDIX

Proposed Model Including Missing Data

Table A1: Relevant variables in the analysis regressed against the missingness of the primary outcomes and mediators of our study. Only type 3 test of fixed effect p-values from logistic regressions are displayed below. An asterisk is used to highlight significance with a threshold of p=0.10.

	Variable	90d Emotional Support (M_1)	90d Patient Activation (M_2)	90d Global Health (Y_1, Y_2)[1]
Pre-TRT covariates	Treatment arm	0.84	0.89	0.95
	Patient lives alone	0.49	0.63	0.77
	Sex	0.70	0.92	0.58
	Discharge Destination	0.13	0.46	0.73
	Race	0.28	0.12	0.13
	Insurance status	0.66	0.44	0.67
	Caregiver consents to trial	0.42	0.44	0.41
	History of CAD	0.45	0.23	0.75
	History of afib	0.49	0.17	0.09*
	History of HTN	0.31	0.48	0.57
	History of depression	0.72	0.19	0.26
	History of alcoholism	0.90	0.91	0.63
	Stroke severity	0.05*	0.95	0.71
	Caregiver lives with patient	0.45	0.30	0.37
	Marital Status	0.48	0.90	0.37
	Education History	0.31	0.67	0.46
Pre-TRT covariates missing	Caregiver lives with patient missing	0.36	0.91	0.69
	Marital status missing	<0.001*	<0.001*	<0.001*
	Education History missing	<0.001*	<0.001*	<0.001*
Baseline Health Measures	7d physical health	0.08*	0.15	0.15
	7d mental health	0.02*	0.06*	0.08*
	7d patient activation	0.001*	0.05*	0.02*
	7d emotional support	0.97	0.86	0.69
	7d anxiety score	0.12	0.19	0.12
	7d depression score	0.004*	0.03*	0.02*
Baseline Health missing	7d PH, MH missing	<0.001*	<0.001*	<0.001*
	7d PAM missing	<0.001*	<0.001*	<0.001*
	7d EMS missing	<0.001*	<0.001*	<0.001*

[1]Note that both outcomes are displayed in the same column here – they are both based off the same 10-item questionnaire, and so have the same missingness pattern.

Table A1 is necessary to clarify the way that missing data affects variable relationships in our model. Note that the strongest predictor of missingness of 90-day mediators and outcomes was the missingness of seven-day health measures. Education and marital status, which were also measured in the seven-day interview, are similarly strongly associated. These relationships are displayed in the updated DAG shown in *Figure A1* below.

Figure A1

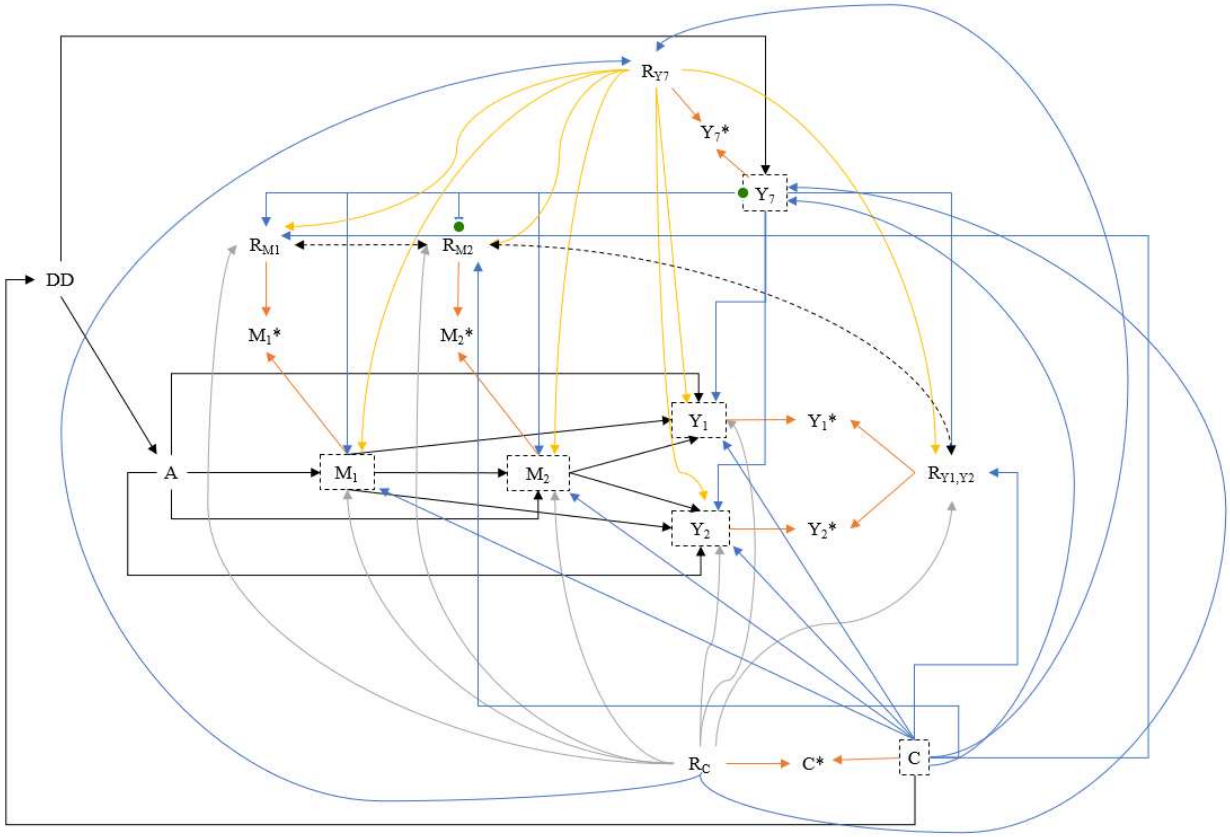

Figure A1: Theorized Causal DAG expanded to include missingness mechanisms. Variable names: A = Treatment Arm; DD = Discharge Destination; M1 = Emotional Support @90-day; M2 = Patient Activation @90-day; Y1 = Physical Health @90-day; Y2 = Mental Health @90-day; Y7 = Health Measures @7-day; C = Pre-treatment covariates; (*) = denotes the observed portion of a variable; R_ = the missingness mechanism of a variable.

Figure A1 above shows the proposed model updated to include missing data. Given the number of variables in the updated model, abbreviations for variables will be included. See the caption above for the complete list of variables and their associated abbreviations. Variables that are partially observed (ie. have missing data) are drawn in a dashed box, with an arrow from this dashed box to the portion of that variable that is actually observed (represented in *Figure A1* with an asterisk *). Each of these incompletely observed variables also has an associated *R* variable, its missingness indicator variable (note that these strategies for representing missing data in a DAG are borrowed from Thoemmes et al. 2015[16]). Black arrows represent the relationships described in our initial conceptual DAG excluding missing data relationships.

From *Figure A1*, we can make the following inferences for the calculation of natural effects.

$M_1(a) \perp A \mid C, R_C, Y_7, DD$

$M_2(a, m_1) \perp M_1 \mid A, C, R_C, Y_7, R_{Y7}$

$M_2(a, m_1) \perp A \mid C, R_C, M_1, Y_7, R_{Y7}$

$Y_1(a, m_1, m_2), Y_2(a, m_1, m_2) \perp M_1 \mid A, C, R_C, M_2, Y_7, R_{Y7}$

$Y_1(a, m_1, m_2), Y_2(a, m_1, m_2) \perp M_2 \mid A, C, R_C, M_1, Y_7, R_{Y7}$

$Y_1(a, m_1, m_2), Y_2(a, m_1, m_2) \perp A \mid C, R_C, M_1, M_2, Y_7, R_{Y7}$ (direct effect)

$Y_1(a, m_1, m_2), Y_2(a, m_1, m_2) \perp A \mid DD$ (total effect)

Model Refinement

We first ran a model that was based on our initial theorized DAG (*Figure 1*). We included potential interactions between treatment and mediators. We explicitly described interactions here to

highlight that, while we need to include interaction terms between auxiliary variables a_0, a_1, and a_2, these terms only interact with each other within each treatment arm.

$$Predicted\ Y = \beta_0 + \beta_1 a_0^1 + \beta_2 a_0^2 + \beta_3 a_1^1 + \beta_4 a_1^2 + \beta_5 a_2^1 + \beta_6 a_2^2 + \beta_7 a_0^1 a_1^1 + \beta_8 a_0^1 a_2^1 + \beta_9 a_1^1 a_2^1 + \beta_{10} a_0^1 a_1^1 a_2^1 + \beta_{11} a_0^2 a_1^1 + \beta_{12} a_0^2 a_2^2 + \beta_{13} a_1^2 a_2^2 + \beta_{14} a_0^2 a_1^2 a_2^2 + \varepsilon$$

The effect estimate calculations for treatment=1 (relative to treatment=0) that include these interaction terms are displayed below.

$$Total\ Effect_{A \to Y}\ (A=1) = \beta_1 + \beta_3 + \beta_5 + \beta_7 + \beta_8 + \beta_9 + \beta_{10}$$

$$Natural\ Direct\ Effect_{A \to Y}\ (A=1) = \beta_1$$

$$Natural\ Indirect\ Effect_{A \to M1*Y}\ (A=1) = \beta_3 + \beta_7 + \beta_9 + \beta_{10}$$

$$partial\ Natural\ Indirect\ Effect_{A \to M2 \to Y}\ (A=1) = \beta_5 + \beta_8$$

This yielded the results below.

Table A2: Estimate results when running our initially hypothesized model, which included terms for potential interactions between treatment and mediators, and accounted for seven-day interview health measures, even when their values were not statistically different between treatment arm. w_1 below refers to a weighting scheme weighted by the expected values of mediator 1 (emotional support), whereas w_2 refers to a weighting approach based on the expected values of mediator 2 (patient activation). Note that the "usual care" treatment arm is the reference group for all estimates below.

Outcome	Effect	w_1		w_2	
		SWCM	***VSSP***	***SWCM***	***VSSP***
Physical Health	*(TE)* $E(X_{A \to Y1})$	–0.341	0.855	–0.384	0.812
	(NDE) $E(X_{A \to Y1})$	0.0481	1.355	–0.029	1.258
	$E(X_{A \to M1*Y1})$	0.126	0.215	–0.812	–0.244
	$E(X_{A \to M2 \to Y1})$	–0.514	–0.715	0.458	–0.202
Mental Health	*(TE)* $E(X_{A \to Y2})$	–1.549	0.203	–1.508	0.243
	(NDE) $E(X_{A \to Y2})$	–0.504	0.290	–0.585	0.213
	$E(X_{A \to M1*Y2})$	0.116	0.236	–1.794	0.269
	$E(X_{A \to M2 \to Y2})$	–1.160	–0.323	0.871	–0.239

After running the above analysis, we noted the discrepancy present in all effects of interest based on whether we were using a weighting scheme based on M_1 (w_1) or M_2 (w_2). This implied a

misspecification of our working model. While Steen's weighting scheme could potentially leave indirect effects sensitive to which mediator was being used to calculate the weights, we did not expect the total effects and natural direct effects to be similarly responsive to weighting changes. We decided some model refinement was needed. First of all, this model included interaction terms between treatment arm and mediator values. An interaction effect was theoretically possible if the treatment applied to each patient acted differently based on mediator levels (ie. a patient's level of personal support or activation). We found no statistically significant interaction effects. Additionally, this model included seven-day interview values of six wellness measures (physical health, mental health, emotional support, activation, depression, and anxiety). Of these six wellness measures, only physical health showed any marginal difference between treatment arms (*Table 2*). Thus we refined the model by removing terms describing interactions between treatment and mediators, all of which were insignificant in our model, and we further refined the model by removing all seven-day health measures besides physical health that were not shown to have a meaningful relationship with treatment group. Ultimately, this simplified our natural effects model to the more streamlined version shown in *Figure 3*.

REFERENCES

REFERENCES

1. Albert, Jeffrey. “Mediation Analysis vis potential outcomes model.” *Statistics in Medicine* (2008) 27: 1282-1304.

2. Bambhroliya, Arvind B; Donnelly, John P.; Thomas, Eric J.; et al. “Estimates and Temporal Trend for US Nationwide 30-Day Hospital Readmission Among Patients with Ischemic and Hemorrhagic Stroke.” *JAMA Network Open* (2018) 1: 4: e181190.

3. Baron, Reuben M.; Kenny, David. A. “The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations.” *Journal of Personality and Social Psychology* (1986) 51: 6: 1173-1182.

4. Brand, Jaap; van Buuren, Stef; et al. “Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data.” *Statistics in Medicine* (2019) 38: 210-220.

5. Hare, R; Rogers, H; Lester, H; et al. “What do stroke patients and their carers want from community services?” *Family Practice* (2006) 23: 1: 131-136.

6. Hayes, Andrew F. “Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millenium.” *Communication Monographs* (2009) 76: 4: 408-420.

7. Hayes, Andrew F. *Introduction to mediation, moderation, and conditional process analysis: a regression-based approach*. The Guilford Press. 2013.

8. Hernan, MA; Robins, JM. *Causal Inference: What if.* Boca Raton: Chapman & Hall / CRC. 2020.

9. Hong, Guanglei. “Ratio of Mediator Probability Weighting for Estimating Natural Direct and Indirect Effects.” In: Proceedings of the American Statistical Association, Biometrics Section. Alexandria, VA: American Statistical Association; 2007: 2401-2415.

10. Imai, Kosuke; Keele, Luke; Tingley, Dustin. “A General Approach to Causal Mediation Analysis.” *Psychological Methods* (2010) 15: 4: 309-334.

11. Lange, Theis; Rasmussen, Mette; et al. “Assessing Natural Direct and Indirect Effects Through Multiple Pathways.” *American Journal of Epidemiology* (2014) 179: 4: 513-518.

12. “Patient Activation Measure.” PAM-10. ©2003-2016 University of Oregon. All Rights Reserved. ©2016 Insignia Health. All Rights Reserved.

13. Reeves, Mathew J; Hughes, Anne K; et al. “Improving transitions in acute stroke patients discharged to home: the Michigan stroke transitions trial (MISTT) protocol.” *BMC Neurology* (2017) 17: 115.

14. Reeves, Mathew J; Fritz, Michele C; Woodward, Amanda T.; et al. “Michigan Stroke Transitions Trial: A Clinical Trial to Improve Stroke Transitions.” Circulation: Cardiovascular Quality and Outcomes (2019) 12: 7: e005493.

15. Steen, Johan; Loeys, Tom; et al. “Flexible Mediation Analysis with Multiple Mediators.” *American Journal of Epidemiology* (2017) 186: 2: 184-193.

16. Thoemmes, Felix; Mohan, Karthika. “Graphical Representation of Missing Data Problems.” *Structural Equation Modeling: a Multidisciplinary Journal* (2015) 22: 4: 631-642.

17. Van Buuren, Stef; Groothuis-Oudshoorn, Karen. “MICE: Multivariate Imputation by Chained Equations in R.” *Journal of Statistical Software* (2011) 45: 3.

18. Van Buuren, Stef. *Flexible Imputation of Missing Data: Second Edition*. Chapman & Hall / CRC. 2018.

19. VanderWeele, Tyler J. “Mediation Analysis: a Practitioner’s Guide.” *Annual Review of Public Health* (2016) 37: 17-32.

20. Zhao, Xinshu; Lynch, John G; Chen, Qimei. “Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis.” Journal of Consumer Research (2010) 37: 2: 197-206.