TBM

ORIGINAL RESEARCH

Illustrating idiographic methods for translation research: moderation effects, natural clinical experiments, and complex treatment-by-subgroup interactions

Ty A Ridenour, PhD,[1,2] Andrea K Wittenborn, PhD,[3] Bethany R Raiff, PhD,[4] Neal Benedict, PhD,[2] Sandra Kane-Gill[2]

[1]RTI, International, Research Triangle Park, NC, USA
[2]University of Pittsburgh, Pittsburgh, PA, USA
[3]Michigan State University, East Lansing, MI, USA
[4]Rowan University, Glassboro, NJ, USA

Correspondence to: T Ridenour
tridenour@rti.org

Cite this as: *TBM* 2016;6:125–134
doi: 10.1007/s13142-015-0357-5

Abstract

A critical juncture in translation research involves the preliminary studies of intervention tools, provider training programs, policies, and other mechanisms used to leverage knowledge garnered at one translation stage into another stage. Potentially useful for such studies are rigorous techniques for conducting within-subject clinical trials, which have advanced incrementally over the last decade. However, these methods have largely not been utilized within prevention or translation contexts. The purpose of this manuscript is to demonstrate the flexibility, wide applicability, and rigor of idiographic clinical trials for preliminary testing of intervention mechanisms. Specifically demonstrated are novel uses of state-space modeling for testing intervention mechanisms of short-term outcomes, identifying heterogeneity in and moderation of within-person treatment mechanisms, a horizontal line plot to refine sampling design during the course of a clinic-based experimental study, and the need to test a treatment's efficacy as treatment is administered along with (e.g., traditional 12-month outcomes).

Keywords

Trajectory analysis, State-space modeling, Translation, Prevention

Implications

Practice: Idiographic clinical trials need to be employed in pilot studies, research that requires high external validity and clinic-based studies, and to round out knowledge garnered from randomized clinical trials.

Policy: Idiographic clinical trials provide an assortment of rigorous techniques for translational research, especially pertinent to early stages of testing knowledge gained in one type of translation for use in another type.

Research: Idiographic clinical trials need to be used for small populations, rare diseases, when funds are sparse, and for treatment experiments conducted within "real-world" clinical settings.

Fishbein et al. [1] outline the translation process from basic science to widespread policy (their Table 1). An essential bridge for using knowledge that is garnered at one translation phase to advance practice at a later translation phase (and vice versa) is the underlying mechanisms which promote intervention outcomes [1]. To illustrate, highly valuable evidence for back translation is obtained in type 2 translation research when hypothesized intervention mechanisms are not associated with treatment outcomes, thus revealing an error in our basic knowledge [2]. Translation science, like intervention research in general, employs a somewhat limited segment of the research methodologies that are available. The purpose of this manuscript is to demonstrate idiographic techniques that offer wide applicability for translation science in a non-technical manner for researchers who could employ them in collaboration with statisticians.

Predominant contemporary methods that are used to test mechanisms are mediation and moderation analyses [3, 4]. Within nomothetic studies (which derive population- or subgroup-average estimates) such as randomized clinical trials (RCTs), mediation and moderation techniques are highly adept, elegant, and flexible. Their limitations include requiring large samples, reliance on waves of data usually spaced by at least 6 months (long after treatment has ended rather than while treatment is delivered), and they assume that the same mediation process occurs for all members of a population. For a number of common and important intervention research scenarios, alternative techniques are preferred. Pilot studies provide formative knowledge early in each translation stage but are typically too limited in sample and funding for nomothetic methods. Even so, the critical contributions of such studies merit using rigorous techniques. Additional scenarios include treatments for rare diseases (which affect about 25 million US citizens) [5], emerging illnesses (e.g., Ebola), genetic microtrials, hard-to-reach or small populations (e.g., Native American

tribes), clinical trials in third world countries, and policy studies in which "subjects" consist of organizations, clinics, or states. Idiographic clinical trials (ICTs), also termed N-of-1 or case studies, use within-subject experimental designs to focus intensively on individual-level data over short time periods (e.g., 1 week to a few months) and could be used in any of the aforementioned research scenarios [6].

Traditional N-of-1 or case studies were used to inform clinician's decision-making for individual clients/patients whereas ICTs add rigorous data analysis and may use medium-sized samples to improve generalizability of results. ICTs overcome biases of relying on visual inspection [7, 8], are able to utilize small samples coupled with rigorous methods, mediation testing investigates within-person processes over time, do not assume the same mediation processes occur across persons, can study outcomes and mediators as treatment is being administered, provide high external validity, and are often conducted in "real-world" settings rather than highly controlled laboratory settings. Hence, ICTs have much to offer translation research, which often must be conducted in applied real-world settings, emphasize external validity over internal validity, and use small N (as subjects may be clinics, communities, or states).

Presented herein are secondary analyses of data from recent ICTs in psychology and medical clinics to illustrate the flexibility and potential uses of ICTs in translational pilot studies when traditional nomothetic methods could not be used. Detailed statistical treatise is not included, because such resources are available elsewhere [9–12]. Study 1 presents a replication pilot study ICT in which moderated mediation is tested (type 3 translation). Study 2 illustrated use of ICT methods within natural clinical experiments and provided a technique for monitoring sampling progress that can inform making adjustments to recruitment strategy "on the fly" for quality control (type 3). Study 3 presents a pilot study ICT testing a novel dissemination venue (type 4 translation) and shows the sophistication of research questions that ICTs can test. Full reports of each study can be requested from the corresponding author.

Each of the three studies used one of two analysis techniques, thus illustrating their wide utility. Both techniques analyze time series data, which consist of many repeated observations over a much shorter time period than traditional nomothetic studies (e.g., daily observations for 1.5 months). Trajectory analysis estimates a trend over time in an outcome variable and tests for hypothesized causes or correlates of variability in the outcome (e.g., intervention, person characteristics). In contrast, state-space modeling investigates observation-to-observation (e.g., day-to-day) associations among multiple variables over time within individuals (e.g., to test whether change in one variable precedes change in another or vice versa). Next, these techniques are tersely summarized but also cite more complete references.

ANALYTIC TECHNIQUES

Mixed model trajectory analysis–Mixed model trajectory analysis (MMTA), also termed hierarchical linear modeling, quantifies an individual's outcomes as time series observations at level 1, while aggregates of individual and group data are analyzed at level 2 [6, 9, 10]. Maximum likelihood estimation and common fit statistics tested whether competing predictors, treatment arms, and error covariance structures improved fit of a model to observed data using SAS 9.3. Model parameters were obtained using restricted maximum likelihood estimation. Some fit statistics are not reported herein to conserve space but are available upon request from the corresponding author. Autoregressive, heterogeneous autoregressive, autoregressive moving average, and Toeplitz error covariance structures, each with lag 1, were tested. The Kenward-Roger adjusted F test reduced risk of type I error.

State-space modeling–Theorized treatment mechanisms can be tested by comparing how associations among mechanisms and outcomes differ between the baseline phase (before treatment) versus during the treatment phase using within-subject experiments. Unified structural equation modeling (USEM) is a form of state-space modeling which in some ways resembles SEM but models observation-to-observation changes (e.g., day-to-day or session-to-session) *within individuals* [13]. Chow et al. [11] review consistencies and differences between USEM and SEM including their relative strengths and weaknesses with regard to investigating intraindividual change and interindividual variability.

USEM was modeled separately by individual and study phase (e.g., as if each participants' data were from a separate sample) and aggregated. To acquire fit statistics and parameter estimates aggregated across entire samples or subgroups, their corresponding parameters were fixed to be equal. The best fitting model was then tested for heterogeneity among study phases and/or subgroups. Tests of moderation (e.g., differences between genders or treatments) were conducted by comparing the fit of two models: fixing parameters to be equal across all participants versus estimating parameters separately per level of the moderator (e.g., males vs females). Competing models were tested using Akakie's I nformation Criterion, Brown-Cudeck criterion, and likelihood ratio χ^2 (for nested models) with AMOS 19.

One limitation to ICTs is the need to delineate certain analytic techniques specifically for intensive, within-person time series data analysis. For example, the fit statistics used herein for state-space model testing have established literatures and examples in the literature within time series data and ICT contexts [11, 13]. Yet, alternative statistics may be more appropriate, offer greater power (in the case of LR χ^2), or provide requisite conservative adjustments of which the field is not yet aware. Some recent advances in idiographic statistics include a bootstrapping technique for modeling mediation analysis using time series data [14], power analysis for randomized ordering of competing

treatments [15], and using "moving windows" methods and non-parametric statistical testing to identify distinct within-person patterns of behavior to forecast a person's change from one pattern to another [16].

Study 1: moderated mediation for testing program effectiveness in a novel population

About 16 % of US adults experience major depression, half of whom receive treatment for depression, and of those, 42 % do not receive adequate care [17]. Less evidence exists regarding treatment of depression in men than women, yet men are more likely to experience symptoms of committing suicide, abusing substances, and engaging in risky or aggressive behavior [18]. Men are also less likely to initiate mental health services, and men with depression who seek help often go undiagnosed and untreated [19].

Depression and marital distress frequently co-occur, and evidence suggests that each may influence the other by way of multiple mechanisms [20]. Couple therapy appears to be the most effective depression treatment when relationship distress is co-occurring [21]. However, as noted earlier, evidence is largely limited to women with depression [20].

Originally, this pilot study tested the efficacy of emotionally focused therapy (EFT) for reducing depression in men and women who were also experiencing relational distress [21]. EFT aims to enhance emotional bonds and responsiveness between members of a couple via novel interactions designed to enhance emotional attachment and encourage sharing of emotions. EFT has efficacy in women for enhancing marital satisfaction and reducing depressive symptoms [20]. Cross-lag model mediations between relationship satisfaction and depression were analyzed within individuals (see Fig. 2, later) [13, 22]. Herein, MMTA was conducted twice: first for outcomes as treatment was administered and again with 12-month outcomes added to explore differences between their results [23].

METHODS

Wittenborn et al. [24] pilot tested the efficacy of EFT versus usual care (UC). Per the IRB-approved protocol, 15 couples seeking couple therapy were recruited from a large, mid-Atlantic metropolitan area. Inclusion criteria were heterosexual couples in committed partnerships, both aged 18 years or older, either having a Beck Depression Inventory-II score of 20–30 (mild to moderate depression), and both reporting relationship distress. Exclusion criteria in either partner were ongoing domestic violence, suicidality, or substance use disorder; history of psychosis or bipolar disorder; receiving other psychotherapy; or initiating medication for depression less than 2 months before screening. Eligibility screening was conducted by telephone. At intake, couples completed consent forms, assessments, and a video-taped interaction and were randomly assigned to EFT or referral to UC for depression. Study-funded treatment ended at week 17 of participation. Data were first collected before the research team was aware of ICT methods. Fortuitously, data were collected at baseline, after each session, and at 12 months post-baseline.

Beck Depression Inventory-II–Beck Depression Inventory-II (BDI-II) measures depression severity in adults with an internal consistency of α=0.92 [25]. BDI-II has good convergent validity, correlating with other measures of depression, and construct validity based on factor analysis.

Dyadic Adjustment Scale–Satisfaction subscale–The Satisfaction subscale of the Dyadic Adjustment Scale (DAS) assessed each partner's relationship satisfaction [26]. Greater scores indicate greater relationship satisfaction, scores range from 0 to 50, and the Satisfaction subscale has demonstrated reliability and validity in varied populations.

RESULTS

MMTA tested the hypothesis that EFT was associated with greater satisfaction and less depression than UC specifically in men. Three participants failed to provide 12-month data, so analyses of weeks 0 to 17 were reconducted for only participants who provided long-term outcomes. MMTA results differed considerably between analyses that exclude versus include the 12-month outcomes (Fig. 1). The best fitting model for satisfaction during weeks 0 to 17 for the entire sample was 30.5726+0.2894 (per week)−0.3030 (per week for UC) (Fig. 1a); results changed little for only those who also provided long-term outcomes: 30.1643+0.2407 (per week)−0.3358 (per week for UC). In contrast, when 12-month outcomes were included, the model was 29.5112+0.0949 (per week)+3.7920 (per week for UC). The two models suggest opposite conclusions regarding EFT; the short-term model suggests greater satisfaction with EFT whereas the long-term model suggests greater satisfaction with UC.

Outcomes for men's depression also differed between the short-term and long-term models but not to the same degree (Fig. 1b). The short-term model predicting depression was 17.0116–0.6116 (per week)+0.2548 (per week for UC); again, results were similar for only those who also provided long-term outcomes: 18.2032–0.5226 (per week)+0.2188 (per week for UC). The long-term model was 14.6737–0.2540 (per week)+0.1335 (per week for UC). Both models implied that men's depression is lessened to a greater degree with EFT compared to UC.

The theoretical Granger causality USEM in Fig. 2 guided tests to clarify the sequence(s) of change among men's depression and relationship satisfaction. Each participant's model parameters appear in columns 2–6 of Table 1. Table 2 presents fit statistics for (1) whole-sample-aggregated estimates versus (2) treatment-specific parameters (the latter better fit observed data). A third, exploratory, model (3) tested the fit of subgroups of participants based on the technique of Zheng et al. for K-cluster analysis of their model parameters [13]. Four

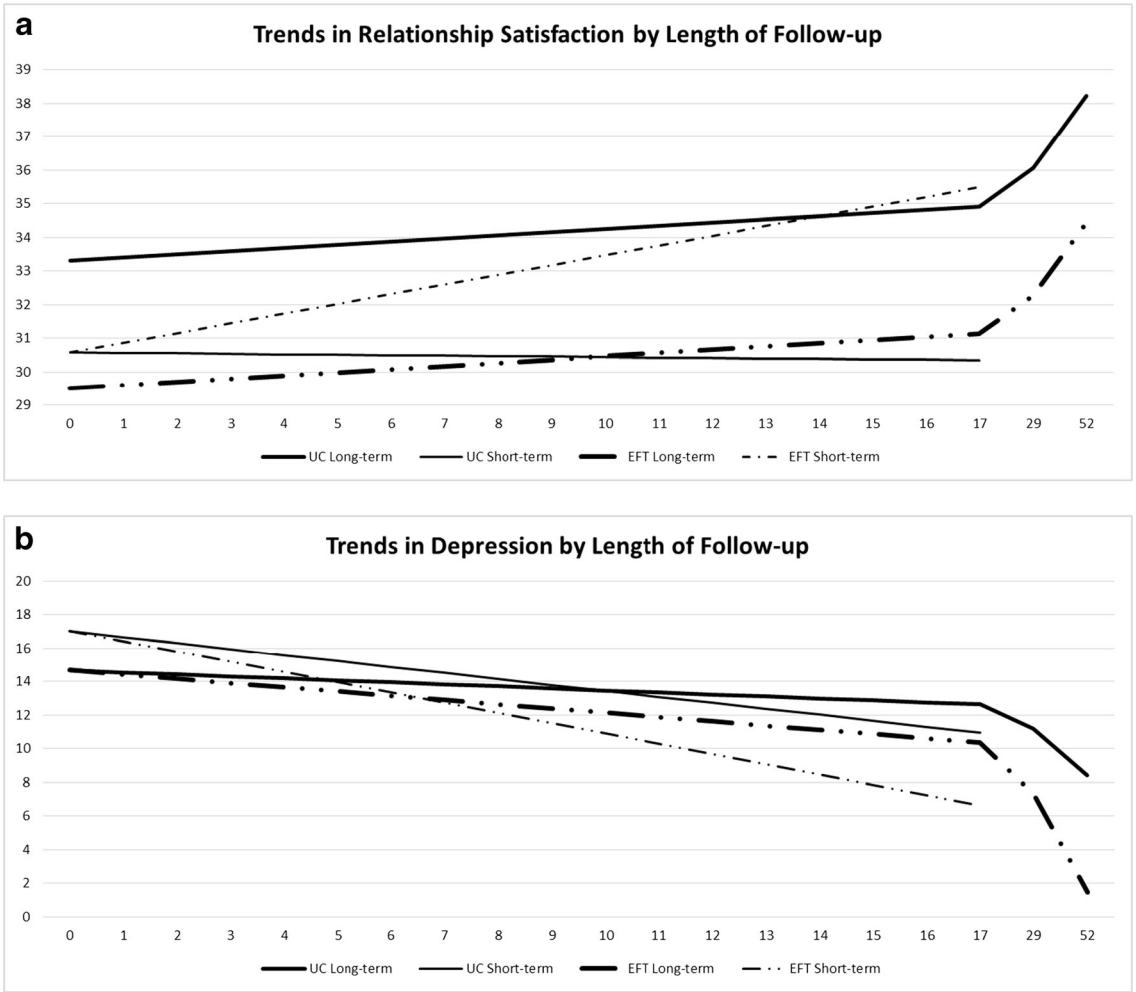

Fig. 1 | Trajectories of outcomes using data from study weeks 0 to 17 versus 0 to 52

clusters emerged, summarized in the first, seventh, and eighth columns of Table 1. This third model provided the best fit, but cluster membership was not associated with treatment (Table 1). A cluster-by-treatment interaction may even better account for observed data, however,sample size precluded such modeling.

DISCUSSION

The single consistent result from analyses was that time-varying relationships between depression and relational satisfaction during EFT (and UC) differ considerably among individuals. Sequences in influence between the two appear to result at least as much from individual differences as treatment type. Also, EFT mechanisms appear to differ between treatment administration compared to long-term (e.g., 12 months) outcomes. Better understanding of these mechanisms and individual differences (and between cluster differences) in the processes connecting relationship satisfaction and depression has potential to improve EFT efficacy for men. Results also suggest that different mediators operate while EFT is delivered compared to sustaining long-term outcomes [23]. Broadly speaking, testing Granger causality can shed light on variance in treatment mechanisms and provides stronger causal inferences than nomothetic studies because results reflect consistent changes within-persons over time.

Study 2: using ICT analyses in a clinical "natural experiment"

Over 70 % of critically ill intensive care unit (ICU) patients experience agitation due to many sources including pain, delirium, hypotension, hypoglycemia, and alcohol withdrawal [27]. A universal goal of critical care practitioners is maintaining a patient's comfort using sedatives. Yet, 40 to 60 % of patients experience inadequate relief from anxiety or over-sedation, and little is known about the patient perspectives of ICU sedation therapy [28, 29].

A barrier to research in the ICU is the infeasibility of randomized trials. ICU patients often cannot be consented for randomization due to delirium, being sedated before arrival, or needing emergent care such as injury - related truama. Also, randomizing patients to sedation management reflects efficacy rather than effectiveness within real-world ICU conditions.

Patients whose ICU stay spans the care of multiple physicians often undergo a within-person natural

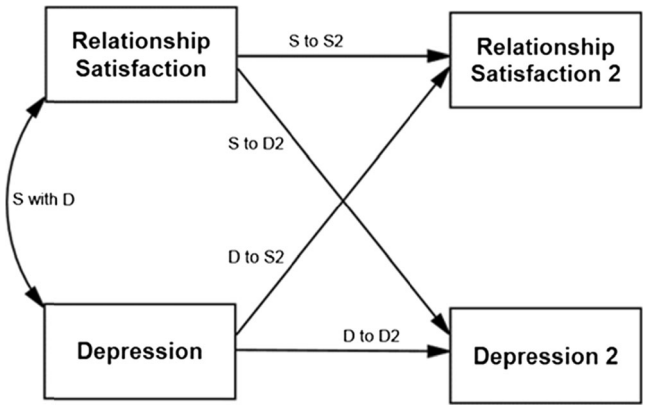

Fig. 2 | Theory-based exploratory model to test Granger causality of the sequence of change in relationship satisfaction and depression

clinical experiment because of variability in sedation practice, based on individual preferences. Additionally, sedation practices vary because patient needs change over time. For example, during extubation, a patient may require a sedative that lacks respiratory depression characteristics. At the ICU in study 2, different physicians administered midazolam, propofol, or dexmedetomidine, allowing comparisons between them in terms of (dis)comfort levels.

One study was available to inform the hypothesis; in 89 surgical ICU patients receiving mechanical ventilation, propofol and dexmedetomidine provided equivalent amnesia and pain control, but dexmedetomidine led to greater discomfort and sleeping difficulties [30]. The present study compared patient comfort among sedatives using MMTA. Herein, data were reanalyzed to demonstrate a technique that could be used in future ICTs to monitor and adjust sampling to better imitate an ideal ICT even when experimental conditions cannot be carefully controlled.

METHODS

Benedict et al. [29] detailed the study methods. Consistent with IRB approval, the 29 participants were from a tertiary care ICU. During usual clinical care, a nurse administered the Sedation-Agitation Scale (SAS) at least once during each 8-h shift [31]. Scores were the absolute value of the difference between a SAS score from the ideal range of 3.0–4.0 (scores in this range=0.0).

Table 1 | Standardized path coefficients for male participants and the four-cluster solution

	ID	Autocorrelation		Cross-lag paths		Common characteristics unique to cluster members	Study arm
		StoS2	DtoD2	StoD2	DtoS2		
Cluster 1	A	−0.02	0.64	−0.30	−0.96	Autocorrelation in depression; Granger causality from depression to satisfaction	UC
	B	−0.05	0.71	−0.13	−0.56		UC
	C	−0.08	0.71	0.04	−0.54		EFT
Cluster 2	D	1.09	0.01	−0.69	0.26	Autocorrelation in satisfaction only; Granger causality from satisfaction to depression and less so the reverse sequence	EFT
	E	0.50	−0.09	−0.83	−0.40		EFT
	F	0.47	0.05	−0.39	−0.28		UC
Cluster 3	G[a]	0.16	0.50	0.31	−0.24	Moderate autocorrelation for depression; small to nil cross-lagged correlations	UC
	H	0.16	0.50	0.00	−0.20		EFT
	I	−0.33	0.33	0.11	−0.23		EFT
Cluster 4	J[a]	0.65	0.64	−0.32	−0.20	Large autocorrelation in depression and satisfaction; moderate to nil cross-lagged correlations	EFT
	K	0.86	0.76	−0.19	0.01		EFT
	L[a]	0.54	0.90	−0.04	−0.45		UC
	M	0.63	0.91	−0.02	0.19		UC
	N	0.55	0.76	0.12	0.05		EFT

Model parameters of one participant did not statistically fit into any of the clusters but resembled the pattern of cluster 2; they were −0.71, 0.07, 2.23, and −0.09, respectively

S relationship satisfaction, *D* depression

[a] Participant dropped out of the study prior to longer term follow-ups

Table 2 | Fit statistics of three competing subgroupings

Path parameters fixed equal...	χ^2, df	RMSEA	AIC	BCC	LR χ^2, df vs model 1
1. Across all participants	1199.09, 171	0.15	1277.1	1305.0	–
2. Within treatment arms	1183.18, 166	0.15	1271.2	1302.6	15.9, 5
3. Within each cluster	1124.27, 151*	0.15	1242.3*	1284.4*	58.9, 20*

Models 2 and 3 are not nested and thus were not compared using LR χ^2

df degrees of freedom, *RMSEA* root mean square error of approximation, *AIC* Akakie's information criterion, *BCC* Brown-Cudeck criterion, *LR* likelihood ratio

Inclusion criteria were 18 years or older and receiving continuous infusion of midazolam, propofol, or dexmedetomidine for at least 24 h while mechanically ventilated. The sample was 58.6 % male, age $\overline{X}$=49 years (SD=17 years), 65.5 % were surgical patients, and the expected mortality rate was about 15 %. Respiratory failure (24.1 %) and motor vehicle/motorcycle collisions (17.2 %) were the most common admission diagnoses. Dexmedetomidine was rarely used, so analyses compared propofol versus either dexmedetomidine or midazolam.

RESULTS

Length of ICU stay (median=14 days or 15 physician shifts) and days on mechanical ventilation (median=4 days) were extensive. Data were 224 SAS scores. Mean (SD) SAS scores aggregated over all shifts for all participants, for propofol (n=179), midazolam (n=42), and dexmedetomidine (n=8), were 3.78 (0.77), 3.31 (1.1), and 2.98 (0.76), respectively. The MMTA model was SAS=0.275527–0.010276 (no. of shifts)−0.228983 (using propofol). Thus, patient sedation improved the longer patient that was in the ICU and using propofol. Propofol (rather than an alternative) almost corrected patient's initial descrepency (i.e., the intercept) from the optimal SAS range. Demographics were not associated with SAS scores (p>0.05).

DISCUSSION

This study illustrated one real-world scenario that required methods which compliment traditional RCTs. A limitation to relying on a natural experiment to imitate an ICT is that study phases may deviate from carefully controlled conditions. Post hoc, how closely the natural experiment resembled an optimal multiple baseline design was evaluated using the horizontal line plot of Tueller et al. [32] of the schedule of each participant's receipt of study medications (Fig. 3). Although transition points from one medication to another (i.e., lengths of experimental phases) appear random, relatively few participants received multiple medications. Fortunately, the pattern of study phases within the context of this study reinforced the conclusion that propofol is the preferred sedative because the negative effects of dexmedetomidine and midazolam occurred after propofol's beneficial outcomes (i.e., they had to reverse the effect of propofol). Using the horizontal line plot of Tueller et al. [32] one or more times during the recruitment/data collection of a natural experiment, ICT may inform subsequent participant recruitment (e.g., to over-recruit participants having experimental conditions that previously had been under sampled) to strengthen the study design. Benedict et al. [29] used additional data and analysis to present supplemental results from additional patient reports that supported the MMTA results reported herein.

Study 3: using MMTA for complex intervention investigations

Diabetes is the seventh leading cause of death in the USA [33]. Hyperglycemia, which can lead to ketoacidosis, is the primary cause of complications from diabetes including heart disease, stroke, blindness, kidney disease, and nervous system disease. About 215,000 individuals with diabetes are younger than 20 and most have type 1 diabetes (T1D) [34]. Managing blood glucose is especially important for adolescents because of diabetes' chronic course and the potential for cumulative health problems. Glycemic control generally declines in youth with T1D as they transition from childhood to adolescence and then adulthood [35]. Glycemic control reduces or postpones diabetes-related complications; however, adolescents' glucose management is poor on average [36]. Fundamental management of T1D requires taking at least four glucose tests daily. Each additional daily blood glucose

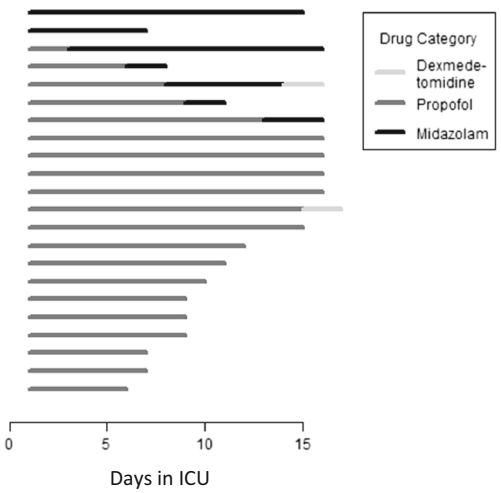

Fig. 3 | Schedule of study patient's medications. Each *horizontal line* represents the duration of a participant's length of study participation in days

test is associated with 0.4 % decreased glycated hemoglobin (A1C) [37] and, in turn, 10.5 % decreased risk of diabetes-related complications [38].

Treatment providers cannot monitor each patient's daily glucose testing. Tested herein was an alternative real-world approach: automated, internet-based monitoring and reinforcement of patient glucose testing. MMTA preliminarily tested its effectiveness in adolescents with T1D who used less than four glucose tests per day. Effects of two interventions with documented efficacy (motivational interviewing (MI) and contingency management (CM)) were controlled. Monetary contingencies were used (e.g., rather than praise or other alternatives) to provide the greatest likelihood of an effect.

METHODS

Raiff et al. [39] detail the study methods. A hybrid participant-as-own-control, with embedded RCT, design tested whether the internet intervention increased frequency of daily glucose tests over and above MI and CM. Participants were randomly assigned to either receive monetary incentives that were contingent on web camera-verified adherence with the guidelines (experimental group) or to a non-contingent group who received incentives independent of their own adherence (a "yoked" control group). Yoked control groups are scientifically rigorous, used to evaluate the relative impact of a contingency that is linked to a novel intervention (controlling for the effect of the reinforcer), and generally show poorer outcomes compared to contingent participants. Randomization was stratified by age (13–15 vs 16–18) and gender. The overall possible total contingency earnings ($220) is comparable to similar CM studies [40].

The multi-phase, ABCA, design consisted of (A) baseline of 5 days, (B) one session of MI, (C) 20 days of web intervention with randomized CM, and (A) 20 days of withdrawal of interventions (termed return to baseline). For withdrawal, participants were instructed to continue using the study meter for 20 days and report their daily glucose testing, and testing records were taken from participant glucometers.

Per the IRB-approved protocol, 52 participants were recruited through an urban outpatient diabetes care center during regularly scheduled medical appointments. Inclusion criteria were ages 13–18, T1D diagnosed at least 1 year prior, using less than four blood glucose tests a day, having internet access at their residence, and completed assent and parental/guardian consent. Participants were loaned a web camera and/or laptop computer (if needed) and given a glucometer and testing supplies. The Bayer Contour USB™ glucometer was used because it can generate reports via a computer's USB port. In this way, investigators verified blood glucose tests.

RESULTS

Randomization was successful, as experimental and non-experimental groups were equivalent in age ($\overline{X}$=15 years, SD=1.5 years), race (28.8 % Caucasian, 50 % Latino, 11.6 % African-American), annual household income (median of about $67 K), age of T1D diagnosis ($\overline{X}$=7 years, SD=3.6 years), and A1C at intake ($\overline{X}$=9.5, SD=1.7). Table 3 presents fit to the data for modeling MI, CM assignment, and the web intervention. Additional effects that were tested (quadratic slope for time, linear slope for MI/baseline, and random effects for time) did not statistically improve model fit ($p<0.05$).

Figure 4 presents the best fitting model as plotted trajectories and an equation. MI was associated with one additional glucose test per day (0.9821). An additional 1.4 glucose tests (1.4080) were associated with the web intervention phase and another two thirds was associated with receiving contingencies (0.6502). During the withdrawal phase, intervention effects gradually declined. Older adolescents responded better to the web intervention (0.4619) than younger adolescents. No age differences were found during baseline, withdrawal, or in response to MI.

DISCUSSION

Results suggested that the web-based monitoring intervention increased daily glucose tests in adolescents with T1D. Similar to past CM findings, efficacy decreased with removal of contingencies. A treatment termination that does not negate treatment effects may improve long-term effectiveness. Alternative reinforcers or more gradual removal of reinforcers are two potential strategies. CM outcomes might be improved in adolescents younger than 16 years with greater parental involvement, use of different contingencies, or alternative venues for monitoring (e.g., smart phones). Results also illustrated back-translation implications regarding developmental differences in outcomes and treatment delivery venue.

Implications

The utility of pilot study ICTs was demonstrated in three distinct translation-oriented studies. The first involved testing treatment generalization to a new population as well as heterogeneity in within-person mechanism pathways toward outcomes (type 2 translation). Study 2 involved type 3 translation in terms of effectiveness of competing treatments within a real-world setting; it also illustrated a technique to monitor how adequately data collection during a natural experiment resembles the needed ICT design (enabling on the fly sampling adjustments). The third study tested an adaptation of a treatment for delivery via a novel venue that could be disseminated to a large patient population. Each study also generated implications for back translation, including

Table 3 | Fit statistics for MMTA models

Model (fixed effects)	AIC	BIC	−2LL	df	LR χ^2, df[a]
1. Intercept, time, MI	6895.7	6905.7	6914.1	6	–
2. Add contingency phase mean effect	6712.0	6724.0	6734.1	7	183.7, 1
3. Add contingency phase slope effect	6653.2	6667.2	6679.0	8	58.8, 1
4. Add contingency phase × age interaction	6646.5	6662.5	6676.0	9	6.7, 1
5. Add mean effect of experimental group	6620.5	6638.5	6653.7	10	26.0, 1

Kenward-Roger correction was used. The only significant random effect was the intercept

AIC Aikake's information criterion, *BIC* Bayesian information criterion, *contingency phase* receiving NS or CS, *LL* log likelihood, *LR* likelihood ratio, *experimental group* participants who received web camera-monitored and contingencies for prescribed frequency of daily glucose tests

[a] For all LR χ^2 tests, the comparison model is the preceding model

*p×0.01

for type 0 translation (revealing heterogeneity in influences between relationship satisfaction and depression, differences in sedating effects among medicines, and developmental differences between younger versus older adolescents).

Each study also demonstrated solutions to traditional barriers to RCTs including small samples, limited funding, pilot studies, and the need for a real-world setting with high external validity. Two ICTs were conducted in clinical settings as treatment was administered to clients/patients, providing greater external validity and more direct implications for effectiveness compared to traditional, highly controlled randomized trials. The flexibility and sophistication that ICTs offer were illustrated, ranging from a natural experiment that occurred as a biproduct of normal clinical procedures to a complex design which controlled for the effects of two treatments while testing for a treatment-by-age interaction.

It is important to consider disadvantages of ICTs, which include limitations to generalizing results especially when samples are quite small or N=1, the historical tradition in ICTs to not statistically analyze data, and generally being limited to investigations of large effect sizes. As with RCTs, ICTs might identify subgroups of clients/patients but do so by identifying similarities among the individuals (e.g., homogeneous clusters) in terms of trends of outcomes or related characteristics, in a bottom-up manner [13]. Population inferences are usually not the purpose of an ICT. But, for such studies, the importance of replication in ICTs corresponds to the proportion of the population that is contained in the ICT sample size. For a true N-of-1 study that is used to guide a clinical decision for the patient (i.e., population=1), replication is irrelevant. For a rare disease, a small sample (N<50) may represent a large proportion of the population of those with the disease. For drawing inferences about a large population, replicating a single ICT is requisite.

Although the techniques demonstrated here require advanced statistical expertise, they can be widely used for testing innovative interventions, translating knowledge from one stage to another [1], shedding light on treatment mechanisms as treatment is delivered, identifying heterogeneous patterns of within-person mediation, and for

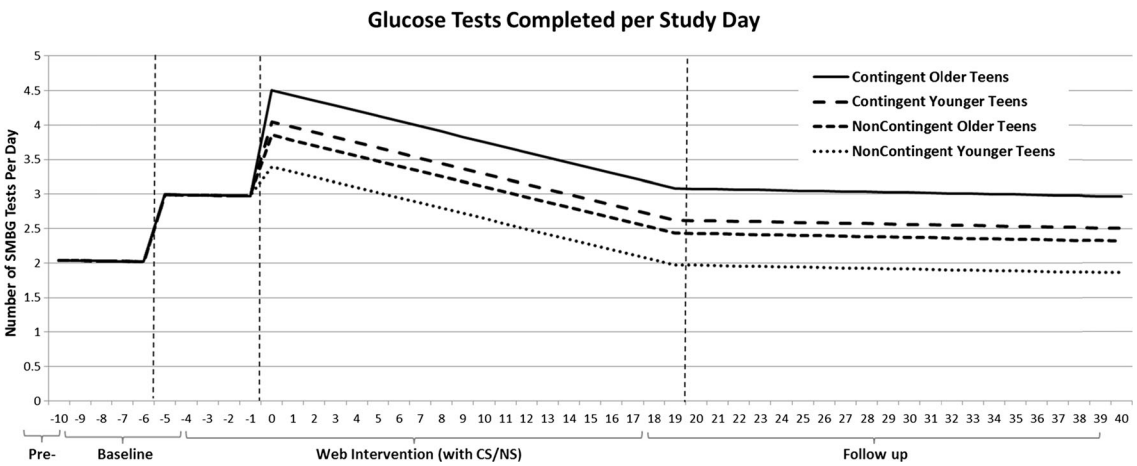

Fig. 4 | MMTA trajectories of daily frequency of glucose testing in adolescents with T1D. The corresponding MMTA equation is frequency of tests=1.9861−0.0055 (per day)+0.9821 (during MI phase)+1.4080 (for web intervention)−0.0693 (per day of web intervention)+0.4619 (older teens during web intervention)+0.6502 (if received contingencies). Each model parameter reached p×0.01 statistical significance

implementation in many situations where a RCT is infeasible. The breadth and depth of evolving ICT methods offer many opportunities for translational prevention research. This is especially true in light of recent health care trends. In 2001, the Institute of Medicine named patient-centered care a central aim of US health care, illustrating the practical need for ICTs which could underlie evidence-based individualized treatment [41]. An important component of the evolving personalized care and medical home model movements is for patients/clients to contribute to clinical decision-making; ICTs provide a format for evidence-based decision-making in which patients/clients and providers have the same information on which to base decisions. Thus, introducing ICT methods for translation scientists and in novel health care contexts, as was accomplished herein, serves an important step toward disseminating them to applied researchers and providers.

Acknowledgments: This was an investigator-initiated study funded by grants from the National Institute on Drug Abuse (P50 05605), National Institute of Child Health and Human Development (R21 061683), an investigator-initiated grant from Hospira, Inc., and Fahs-Beck Fund for Research and Experimentation and the Virginia Tech College of Liberal Arts and Human Sciences Dean's Faculty Fellowship. The funders played no role in the design, conduct, or analysis of the study nor in the interpretation and reporting of the study findings.

Compliance with ethical standards

Conflict of interest: Ty Ridenour, Andrea Wittenborn, Bethany Raiff, Neal Benedict, and Sandra Kane-Gill have no conflict of interests to declare regarding the research reported herein.

Adherence to ethical principles: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee, overseen by the institutional review boards of the institutions where the study was conducted, and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

1. Fishbein DH, Sussman S, Ridenour TA, Herman-Stahl M. Expanding the translational spectrum of prevention science: emerging research to support scaling up proven practices that prevent behavioral health problems. Transl Behav Med. This issue.
2. Dishion TJ, Spracklen KM, Andrews DW, Patterson GR. Deviancy training in male adolescent friendships. Behav Ther. 1996; 27: 373-390.
3. MacKinnon DP. Introduction to statistical mediation analysis. NY: Lawrence Erlbaum; 2008.
4. Edwards JR, Lambert LS. Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. Psychol Meth. 2007; 12: 1.
5. National Institutes of Health. NIH funds research consortia to study more than 200 rare diseases: $29 million awarded to expand NCATS' collaborative Rare Diseases Clinical Research Network. Accessed on Jan 6, 2015 at: www.nih.gov/news/health/oct2014/ncats-08.htm. 2014.
6. Ridenour TA, Pineo TZ, Maldonado-Molina MM, Hassmiller-Lich K. Toward idiographic research in prevention science: demonstration of three techniques for rigorous small sample research. Prev Sci. 2013; 14: 267-278.
7. Franklin RD, Gorman BS, Beasley TM, Allison DB. Graphical display and visual analysis. In: Franklin RD, Allison DB, Gorman RS, eds. Design and analysis of single-case research. Mahwah, NJ: Lawrence Erlbaum; 1997: 119-158.
8. Smith JD. Single-case experimental designs: a systematic review of published research and current standards. Psychol Method. 2012; 17: 510-550.
9. Hedeker D, Gibbons RD. Longitudinal data analysis. Hoboken, NJ: Wiley; 2006.
10. Singer JD, Willett JB. Applied Longitudinal Data Analysis. New York: Oxford: 2003.
11. Chow SM, Ho MHR, Hamaker EL, Dolan CV. Equivalence and differences between structural equation modeling and state-space modeling techniques. Struct Equat Model. 2010; 17: 303-332.
12. Molenaar PC, Huizenga HM, Nesselroade JR. The relationship between the structure of interindividual and intraindividual variability: a theoretical and empirical vindication of developmental systems theory. In: Understanding human development. US: Springer; 2003: 339-360.
13. Zheng Y, Wiebe RP, Cleveland HH, Molenaar PC, Harris KS. An idiographic examination of day-to-day patterns of substance use craving, negative affect, and tobacco use among young adults in recovery. Multivariate Behav Res. 2013; 48: 241-266.
14. Gu F, Preacher KJ, Ferrer E. A state space modeling approach to mediation analysis. J Educ Behav Stat. 2014; 39: 117-143.
15. Liao P, Klasnja P, Tewari A, Murphy SA. Micro-Randomized Trials in mHealth. arXiv preprint arXiv:1504.00238.
16. Bobashev GV, Liao D, Hampton J, Helzer JE. Individual patterns of alcohol use. Addict Behav. 2014; 39: 934-940.
17. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA. 2003; 289: 3095-3105.
18. Centers for Disease Control and Prevention. Web-based injury statistics query and reporting system (WISQARS). Retrieved from: www.cdc.gov/injury/wisqars/index.html 2011.
19. Simpson HB, Nee JC, Endicott J. First-episode major depression: few sex differences in course. Arch General Psychiatr. 1997; 54: 633-639.
20. Wittenborn AK, Culpepper B, Liu T. Treating depression in men: the role of emotionally focused couple therapy. Contemp Fam Ther. 2012; 34: 89-103.
21. Barbato A, D'Avanzo B. Efficacy of couple therapy as a treatment for depression: a meta-analysis. Psychiatr Quart. 2008; 79: 121-132.
22. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. Econometrica. 1969; 37: 424-428.
23. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess. 2009; 21: 486-497.
24. Wittenborn AK, Liu T, Ridenour TA, Seedall RB. Emotionally focused therapy for depression: a rigorous pilot randomized controlled trial. Under review.
25. Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1996.
26. Spanier GB. Measuring dyadic adjustment: new scales for assessing the quality of marriage and similar dyads. J Marriage Fam Ther. 1976; 38: 15-28.
27. Jacobi J, Fraser GL, Coursin DB, et al. Clinical practice guidelines for the sustained use of sedatives and analgesics in the critically ill adult. Crit Care Med. 2002; 30: 119-141.
28. Jackson DL, Proudfood CW, Cann KF, et al. The incidence of suboptimal sedation in the ICU: a systematic review. Crit Care. 2009; 13: R204.
29. Benedict N, Felbinger M, Ridenour TA, Anthes A, Altawalbeh S, Kane-Gill S. Correlation of patient reported outcomes of sedation and sedation assessment scores in critically ill patients. J Crit Care. 2014; 29: 1132.e5-1132.e9.
30. Corbett SM, Rebuck JA, Greene CM, et al. Dexmedetomidine does not improve patient satisfaction when compared with propofol during mechanical ventilation. Crit Care Med. 2005; 33: 940-945.
31. Riker RR, Picard JT, Fraser GL. Prospective evaluation of the sedation-agitation scale for adult critically ill patients. Crit Care Med. 1999; 27: 1325-1329.
32. Tueller S. longCatEDA: Package for Plotting Categorical Longitudinal and Time-Series Data. R package version 0.17. 2014.
33. Centers for Disease Control and Prevention. FastStats—leading causes of death. (n.d.). Accessed September 16, 2014, from www.cdc.gov/nchs/fastats/leading-causes-of-death.htm.
34. Centers for Disease Control and Prevention. National Diabetes Fact Sheet—Publications—Diabetes DDT. (n.d.). Accessed Feb 6, 2012, from www.cdc.gov/diabetes/pubs/factsheet11.htm.
35. Crowley R, Wolfe I, Lock K, McKee M. Improving the transition between paediatric and adult healthcare: a systematic review. Arch Dis Child. 2011 archdischild202473.
36. Berg CA, Wiebe DJ, Suchy Y, et al. Individual differences and day-to-day fluctuations in perceived self-regulation associated with daily adherence in late adolescents with type 1 diabetes. J Pediatr Psychol. 2014; 39: 1038-1048.
37. Haller MJ, Stalvey MS, Silverstein JH. Predictors of control of diabetes: monitoring may be the key. J Pediatr. 2004; 144: 660-661.
38. Stratton IM, Adler AI, Neil HAW, et al. Association of glycaemia with macrovascular and microvascular complications of type 2

diabetes (UKPDS 35): prospective observational study. BMJ. 2000; 321: 405-412.
39. Raiff BR, Barry VB, Jitnarin N, Ridenour TA. Internet-based incentives increase blood glucose testing with a non-adherent, diverse sample of teens with type 1diabetes: a randomized, controlled trial. Under Rev.
40. Raiff B, Dallery J. Internet-based contingency management to improve adherence with blood glucose testing recommendations for teens with type 1 diabetes. J Appl Behav Anal. 2010; 43: 487-491.
41. Institute of Medicine. Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press; 2001.