Potential Utility of Idiographic Clinical Trials in Drug Development

Ty A. Ridenour, PhD, and **Donald Stull, PhD**, RTI, Research Triangle Park, NC, USA

KEY POINTS

Idiographic clinical trials offer a rigorous alternative to randomized controlled trials when the latter are not feasible due to available sample size, funding, or early phase in clinical testing.

Idiographic clinical trials combine subject-as-own-control designs with hierarchical linear modeling that has been tailored specifically for small sample-intensive, within-person analysis.

Idiographic clinical trials are flexible, have been used for a breadth of settings and clinical outcomes, and can be used to address complex treatment questions including safety, drug dosage, and comparative efficacy.

An innovative methodology — idiographic clinical trials (ICTs) — is introduced as a way to inform randomized controlled trials (RCTs) in terms of RCT planning (eg, sample size, effect size), use in research scenarios when RCTs are not feasible (eg, rare diseases with small populations), or use in applied settings such as clinical practice, where RCT parameters cannot be followed. ICTs can be conducted generally for lower cost with faster completion time than RCTs. ICTs should not be seen as replacements for RCTs, but as a way to help inform RCTs or provide insights for early product development without allocating the resources for an RCT for early evaluation of an asset. The term idiographic clinical trials comes from its emphasis on within-individual processes over time. Compared to RCTs, this approach is adept for early phase clinical trials, pilot studies, and testing whether efficacy from an RCT can be replicated in a specific clinical setting or subpopulation (eg, patients with RCT exclusion criteria). ICTs couple two well-known methodologies to yield rigorous results from small samples: subject-as-own-control experimental designs (eg, crossover designs) with hierarchical linear modeling (or multi-level modeling) refined specifically for small samples.

> ICTs couple two well-known methodologies to yield rigorous results from small samples: subject-as-own-control experimental designs (e.g., crossover designs) with hierarchical linear modeling (or multi-level modeling) refined specifically for small samples.

RCTs

RCTs represent the "gold standard" for evaluating efficacy and safety of pharmaceuticals and biologics for regulatory purposes. RCT key features include randomization, blinding, comparison group(s), and isolation of key findings to treatment conditions, all of which contribute to RCTs' high internal validity. RCTs frequently assess an intervention's effect against alternative interventions or no intervention, while minimizing numerous types of bias. Conversely, RCT disadvantages include use of exclusion criteria (limiting their generalizability); unbalanced attrition (ie, patients in one arm are more likely to drop out, as when those in a usual care arm get sicker sooner and drop out); ethics (eg, it is unethical to give some patients placebo); and investigator discretion (eg, decisions about cross-over may be left to physicians/investigators potentially violating randomization), all of which can reduce an RCT's external validity. Moreover, RCTs typically require large budgets, recruiting hundreds to thousands of participants, and up to 18-month follow-ups per participant.

Hence, RCTs are often a balance between costs, time commitments, internal/external validity, choosing a comparator or no head-to-head comparison, and so on.

In early phase trials when establishing a new compound's efficacy, safety, and potential dosing, an RCT may be too expensive or time-consuming or not even feasible (eg, for rare diseases). If the signal-to-noise ratio indicates treatment is not efficacious, a company could lose millions of dollars. A faster and less-expensive option to an RCT could clearly benefit companies exploring new assets. Rigorous ICTs require rarely more than 50 participants, less than 3 months' duration per participant, can address multiple early trial questions in one sample (safety, efficacy, dosage, differences among subgroups), and offer individuals personalized efficacy (termed "impact"), which is a strong incentive to participate and not attrite, all of which may reduce costs. However, ICTs do have should be some limits and may be only used under under specific circumstances. >

A COMPLEMENTARY ALTERNATIVE

Subject-as-own-control experimental designs (eg, crossover and staggered baseline designs) provide the data collection structure for ICTs. Time series data are collected from each participant during a control (or care as usual) time period/phase and experimental treatment(s) time period/phase(s). Many potential confounds are managed because the same participants provide control and experimental data (rather than randomization). To illustrate, random blood glucose test in patients with type 1 diabetes might be observed repeatedly while receiving standard treatment (control phase) and then again during an experimental treatment (treatment phase) [6]. By randomizing length of control phases among participants and varying their study enrollment dates, potentially confounding factors of practice effects, disease natural history, human development, and historical events are controlled. If efficacy estimates are desired from an ICT, participants ought to resemble the population heterogeneity. By coupling subject-as-own-control designs with statistical techniques such as hierarchical linear modeling that is tailored specifically for intensive within-person analysis, they provide highly flexible, rigorous clinical trials. The analytic techniques account for well-known sources of bias including autocorrelation and limitations of visual inspection [1-4]. Introductory papers to ICTs provide more technical details using study illustrations [5-10].

Several strengths of ICTs stem from their far smaller samples, shorter durations, and resultant less cost and time compared to RCTs. If an asset shows a strong enough effect, a company could then use a traditional RCT. If the asset effect is not strong, a decision to not pursue that asset means much less cost and time invested compared to using an RCT to reach the same decision. ICTs can often be used when RCTs are not feasible. ICTs can be used frequently in clinical settings where strict adherence to an RCT protocol may not be possible (eg, ICU patients receiving critical care or when every participant requires the treatment), take advantage of natural experiments, or using quasi-experiments that occur during usual clinical care. Two recent examples were comparative studies between medications for emergency care sedation [5,11] and immunosuppression for recipients of liver or kidney transplants [12].

ICT LIMITATIONS

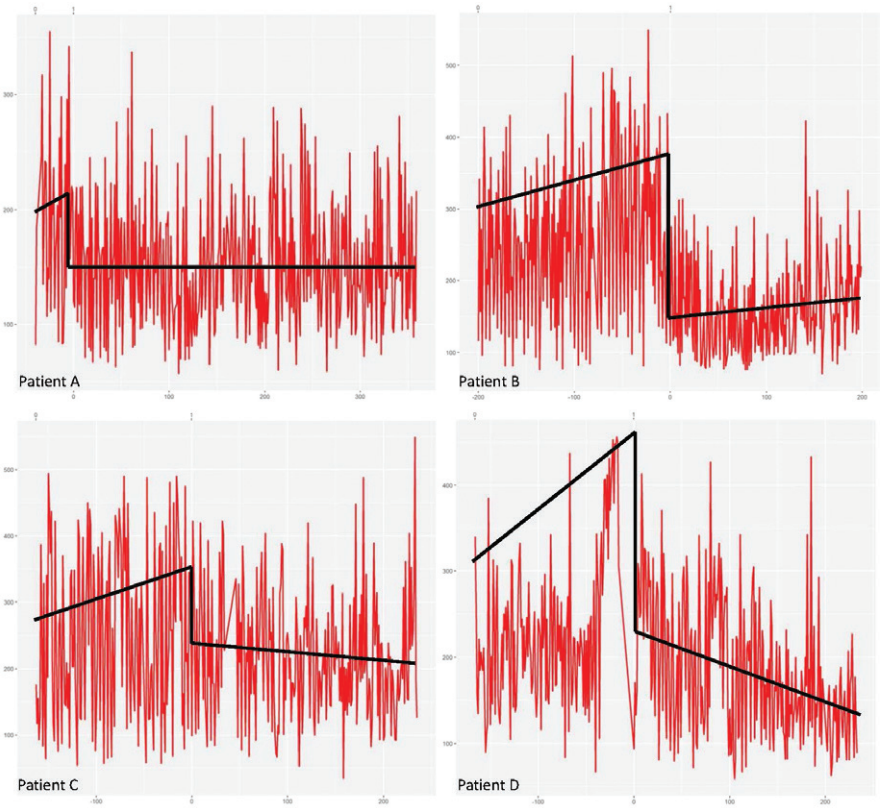
ICTs typically do not provide efficacy for large populations (intensive within-person protocols preclude large samples). Rather, their strengths and limitations provide complementary, patient-centered evidence, much of which can inform subsequent RCTs. ICTs can raise the rigor of early phase trials, orphan drug testing, effectiveness replications of efficacy estimates, and comparative outcomes research involving rare diseases. In addition, ICTs offer limited utility for short-lasting illnesses (common cold, influenza). To illustrate, repeated measurements are usually not feasible during the period of myocardial infarction thus precluding ICTs, whereas ICTs may be ideal for a novel treatment for recovery from myocardial infarction. The following illustrations demonstrate some ICT uses, ranges in complexity, data types, and treatment development stages.

ILLUSTRATION 1: PILOT STUDY OF EFFICACY AND SAFETY

While developing and testing a treatment, ICTs could inform resource allocation, human effort, and time. Erroneous Go/No Go decisions risk (a) costly investments in compounds that end up being unsuccessful or (b) missing lucrative opportunities to develop efficacious medications [13]. ICTs could inform decisions about whether to pursue subsequent clinical phases and provide estimates of effect sizes and patient variability to design them. Illustration 1 demonstrated a pilot ICT that yielded evidence regarding efficacy, within- and between-person variability, and safety. Its small sample illustrates ICTs' potential uses for pilot studies and orphan drug testing.

Diabetic blood glucose is managed in nursing homes by using the sliding scale, which consists of adjusting insulin doses biweekly. Because of large spikes and drops in glucose that occur daily, sliding-scale glucose management often leads to ketoacidosis, unconsciousness, and organ

Figure 1: Manual Pancreas ICT Results Visualized at N=1 Level



Note: X-axis is sequential observations at meals or snack times (4 per day). Y-axis is blood glucose level (mg/dL). The Care-as-Usual phase spans observations prior to "0" (on x-axis); Manual Pancreas was administered thereafter (phases also are indicated by "0" and "1" above each plot).

damage. Contemporary glucose treatments are not used because of cost, potential damage to equipment (eg, glucometers), and a lack of incentives to change. A recently devised algorithm determines bolus insulin dosage based on a patient's blood sugar level and food intake during a meal [6]. Termed "manual pancreas," nurses draw blood to determine glucose levels, enter nutritional values of an anticipated meal, and administer bolus insulin based on the algorithm output. The nursing home where manual pancreas was pilot-tested admitted four patients during the study period.

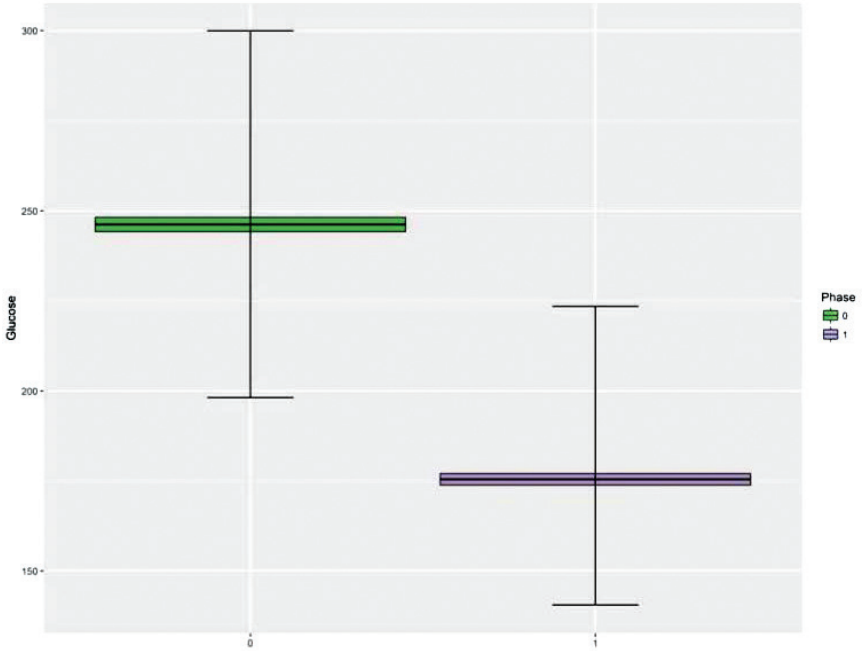
Figure 1 presents participants' modelled trajectories superimposed on observed glucose levels 4 times per day over 100 days. Each participant experienced an instant drop in blood glucose when manual pancreas was initiated, albeit to varying degrees. Variance in blood glucose illustrates how within-person patterns may interfere with interpretation of results and the importance of parsing out autocorrelation to obtain unbiased estimates. Relevant to this study is the circadian rhythm of blood sugar levels, which varies in periodicity among individuals. To test for interactions between manual pancreas and circadian rhythms, analyses were re-conducted separately for each time of day (Table 1). At certain meals, manual pancreas was associated with no improvement. Hence, to avoid safety risks associated with injections, manual pancreas could be skipped for patient A at breakfast and lunch whereas for patient C, manual pancreas could be limited to lunch.

Table 1: Change in Blood Glucose with Manual Pancreas per Time of Day

	Breakfast 7:30am	Lunch 11:30am	Dinner 4:30pm	Snack 8:30pm
Entire Sample	-35.9 (9.8)	-43.3* (194.2)	-59.4 (9.7)	-59.1* (277.9)
Patient A	0.2* (11.1)	1.8* (24.4)	-50.4 (20.2)	-104.2 (19.4)
Patient B	-32.2 (8.8)	-117.3 (23.0)	-156.3 (19.3)	-122.2 (17.0)
Patient C	11.5* (27.5)	-66.6 (26.8)	-35.5* (25.4)	3.0* (27.7)
Patient D	-112.1 (16.0)	26.3* (17.6)	43.5 (17.7)	-57.3 (24.3)

*Note: *Change in glucose was NS (P>.01). Parenthetical values are standard errors.*

Figure 2: Manual Pancreas ICT Aggregate



Note: Center of boxes represent mean blood glucose levels (mg/dL), 95% confidence intervals appear as upper and lower levels of boxes, and whiskers depict the standard deviations.

Figure 2 illustrates how an ICT can inform a subsequent RCT. The box-and-whiskers summary presents care-as-usual versus manual pancreas in terms of mean glucose levels, 95% confidence intervals (tops and bottoms of boxes), and standard deviations (whiskers). Inasmuch as the ICT sample resembles the clinical population of interest, results provide unbiased estimates for RCT planning.

Regarding safety testing, two patients were admitted to an emergency department during care-as-usual phases due to complications from ketoacidosis. During manual pancreas phases, no patient required emergency care. Moreover, clinical staff observed one patient to be far more alert and responsive during manual pancreas, presumably due to lower blood glucose. No health risks were observed related to the manual pancreas; however, repeated use of injections merited caution.

Also illustrated is ICT's provision of person-centered data. Recent movements such as precision medicine and evolving methodologies such as genetic micro trials provide opportunities for ICTs. These methods also can be used for testing mechanisms of outcomes while a treatment is being administered. Far more sophisticated analytics (eg, state-space modeling) are available for testing multivariate processes and outcomes [14].

ILLUSTRATION 2: DIFFERENTIAL/ COMPARATIVE EFFICACY

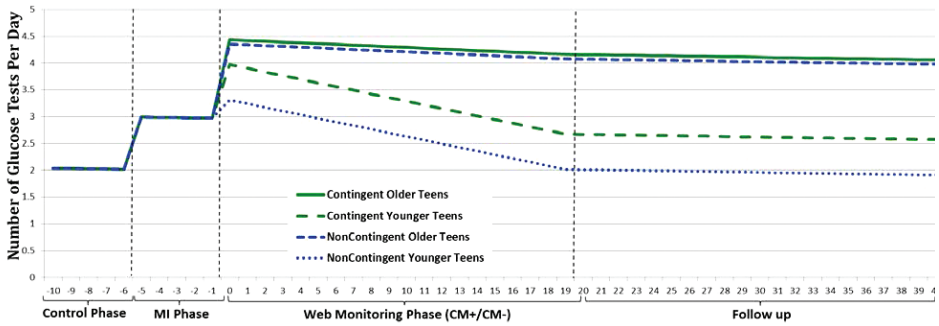
Comparative efficacy encompasses how outcomes differ between treatments or among subpopulations in response to a treatment. For example, if a large clinical trial demonstrates null or small efficacy, homogeneous subgroups may nevertheless respond well to the treatment. A subgroup also may respond poorly to a medication, thereby reducing the apparent overall efficacy. Traditional RCTs are frequently not designed to detail subgroup differences, >

especially if subgroups are not identified a priori. If during an RCT, insight is gained regarding a subgroup that responds differentially to a medication, an ICT could test the hypothesis. Illustration 2 presents results from a behavioral intervention to demonstrate ICTs' potential utility to address differential efficacy.

Over 200,000 US citizens with diabetes are younger than 20, most of whom have type 1 diabetes (T1D) [15]. Adolescent management of blood glucose is especially important, given diabetes' chronicity and cumulative health problems. However, as diabetes management shifts from parent to adolescent, glycemic control usually declines and is poor on average in adolescents [16]. A fundamental step in T1D management is taking four or more glucose tests daily; each additional daily blood glucose test is associated with 0.4% decreased glycated hemoglobin (A1C) [17] and in turn 10.5% decreased risk of diabetes-related complications [18,19].

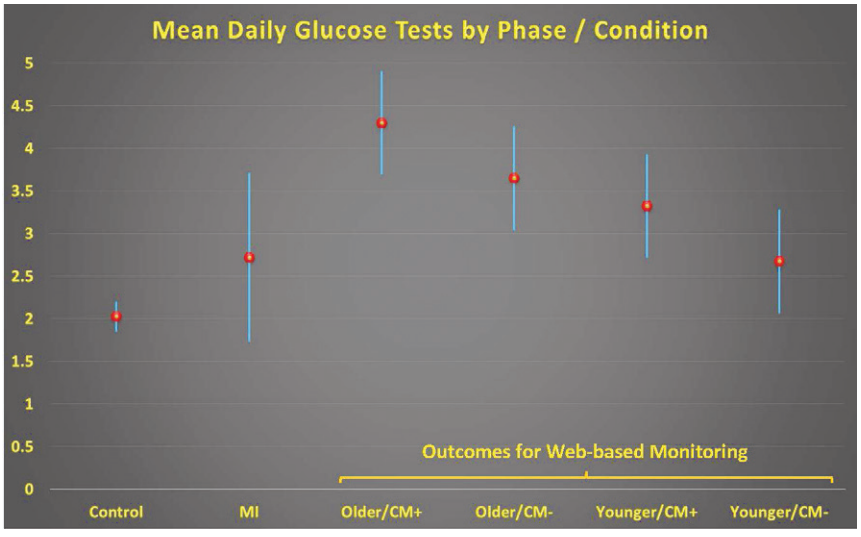
A program recently designed to increase daily blood glucose tests involved adolescents recording their glycemic tests over streaming video and entering test results at the study monitoring website (validated by glucometer readings) [20]. One study objective was to demonstrate the web program's clinical utility over and above two interventions with previously documented efficacy (motivational interviewing, or MI, and contingency management, or CM). In addition, its differential efficacy was compared between ages 13 to 15 versus 16 to 18.

Figure 3: Four-subgroup Trajectories of Tests Completed per Study Day



Note: The corresponding MMTA equation is Frequency of Daily Tests = 1.9885 - 0.00501 (per study day) + 0.9805 (for motivational interviewing) + 1.3240 (during treatment phase) - 0.06317 (per day of treatment phase) + 1.0430 (older teens during treatment phase) + 0.6598 (while receiving CS) - 0.05378 (per day of treatment phase for younger teens). Each model parameter reached P<.01 statistical significance.

Figure 4: Mean Daily Glucose Tests by Phase / Condition



A hybrid ICT-with-randomization design was used. Following a control phase with no intervention, MI was provided to all 41 participants to account for its effects. Next, participants received the web program with randomized CM+ (monetary rewards were contingent upon completing glucose tests; n=23) or CM- (the same monetary amounts were provided, but randomly; n=18). Gender and age group stratified CM randomization.

Figure 3 presents results as trajectories among four subgroups. During the control phase, all subgroups averaged two glucose tests conducted per day. Following MI, all subgroups increased about one glucose test per day and maintained this improvement. When the web monitoring program (plus either CM+ or CM-) began, differential efficacy occurred as an interaction between randomization and age group. Older adolescents and CM+ were associated with more daily glucose tests compared to the alternative subgroups. Younger participants gradually lost the benefit of web monitoring, whereas older participants largely maintained their benefit. After withdrawing all interventions, each subgroup's outcomes slightly decreased. Box-and-whiskers visualization of results more closely resembles traditional efficacy estimates (Figure 4) by depicting the signal-to-noise ratio similar to RCTs (e.g., for later phase clinical trials).

CONCLUSIONS

This introduction demonstrated how early phase ICTs might inform efficacy of an intervention at lower cost and faster than RCTs. When assessing whether a new asset has a sufficient effect and further development is justified, ICTs may be a more efficient and less-expensive alternative to RCTs. Software, methods for data collection, and analytics for multi-episode data are readily available, so these methods can be implemented now. As noted earlier, ICTs cannot replace RCTs. However, given the rigor of ICTs, the amount of data collected per patient, and the ability to learn the effect of alternative interventions/doses (i.e., treatment-/dose-switching becomes a time-varying covariate), ICTs can be a rich source of data for evaluating treatment effects as well as patient and disease progression.

REFERENCES

1. Kwok OM, West SG, Green SB. The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. Multivar Behav Res 2007;42:557-592.

2. Sivo SA, Fan X, Witta EL, Willse JT. The search for" optimal" cutoff properties: Fit index criteria in structural equation modeling. J Experim Educ 2006;74:267-288.

3. Franklin RD, Gorman BS, Beasley TM, Allison DB. Graphical display and visual analysis. In R.D. Franklin, D.B. Allison, & R.S. Gorman (Eds.), Design and Analysis of Single-Case Research (pp. 119-158). Mahwah, NJ: Lawrence Erlbaum. 1997.

4. Smith, JD. Single-case experimental designs: A systematic review of published research and current standards. Psycholog Methods 2012;17:510-550.

5. Ridenour TA, Wittenborn AK, Raiff BR, et al. Illustrating idiographic methods for translation research: Moderation effects, natural clinical experiments, and complex treatment-by-subgroup interactions. Translat Behav Med 2016;6:125-134.

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 Science 2013;14:267-278.

7. Ridenour TA, Chen SHK, Liu HY, et al. The clinical trials mosaic: Toward a range of clinical trials designs to optimize evidence-based treatment. J Person-Oriented Res in press.

8. Raudenbush SW, Bryk AS. Hierarchical linear models: Applications and data analysis methods. Sage; 2002.

9. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. New York: Oxford University Press, 2003.

10. Hedeker D, Gibbons RD. Longitudinal Data Analysis. Hoboken, NJ:Wiley & Sons, 2006.

11. Benedict N, Felbinger M, Ridenour TA, et al. Correlation of patient reported outcomes of sedation and sedation assessment scores in critically ill patients. J Crit Care 2014;29: 1132.e5-1132.e9.

12. Momper JD, Ridenour TA, Schonder KS, et al. The impact of conversion from Prograf to generic tacrolimus in liver and kidney transplant recipients with stable graft function. Am J Transplant 2011;11:1861-1867.

13. Potter WZ. Optimizing early Go/No Go decisions in CNS drug development. Expert Rev Clinical Pharmac 2015;8:155-157.

14. Beltz AM, Wright AG, Sprague BN, Molenaar PC. Bridging the Nomothetic and Idiographic Approaches to the Analysis of Clinical Data. Assess in press.

15. Centers for Disease Control and Prevention. National Diabetes Fact Sheet - Publications - Diabetes DDT. Available from: www.cdc.gov/diabetes/pubs/factsheet11.htm. [Accessed February 6, 2012].

16. Crowley R, Wolfe I, Lock K, McKee M. Improving the transition between paediatric and adult healthcare: a systematic review. Arch Dis Child 2011 archdischild202473.

17. 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.

18. Haller MJ, Stalvey MS, Silverstein JH. Predictors of control of diabetes: monitoring may be the key. J Pediatr 2004;144:660-661.

19. 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.

20. 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 1 Diabetes Mellitus: A randomized, controlled trial. Translat Behav Med 2016;6:179-188.

Additional information:

The preceding article is based on a workshop given at the ISPOR 19th Annual European Congress.
To view the authors' presentations, go to: https://www.ispor.org/Event/ReleasedPresentations/2016Vienna#workshop presentations