



OPEN ACCESS

ORIGINAL ARTICLE

Does this treatment work for me? The patient's role in assessing medical care

Duncan V Neuhauser,¹ Jennifer Chu²

¹Department of Epidemiology and Biostatistic, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA

²Department of Physical Medicine and Rehabilitation, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence to
Dr Jennifer Chu, Department of Physical Medicine and Rehabilitation, Perelman School of Medicine, University of Pennsylvania. Mail to Jennifer Chu, 233 East Lancaster Avenue, Suite 101, Ardmore, PA 19003-2321, USA; jchu@etoims.com

Received 16 November 2016

Revised 24 April 2017

Accepted 17 August 2017

ABSTRACT

Randomised clinical trials are designed to determine whether a particular treatment is appropriate to make a significant difference to the health of a defined population and to aid its approval for use. For an accurate, cheap and simple assessment to see if a treatment benefits an individual person, all that is needed is a pen, paper, simple pocket calculator and daily recording of a few variables. It requires the ability to read and write and to understand addition and division. Factorial design of experiments is used to show the impact of several variables and their interaction on the person's health status. An example of a 75-year-old man with an enlarged prostate is used here to illustrate this approach. This person was able to understand and reduce side effects, lower the costs of medication by 83% and improve measured health status by 28%. A multivariate model for this person was then created with about 450 person-days of data.

INTRODUCTION

The method to improve care described here is a synthesis of known ideas including personal quality improvement,¹ measured costs, risks (side effects), benefit of treatment and factorial design of experiments, self-care, the patient's perspective, N of one trials, utility scores, knowledge of clinical medicine, randomised clinical trials, pharmacokinetics and statistical reasoning. To our knowledge, the first use of factorial design of experiments for an additional patient was reported by Olsson *et al.*^{2,3} The experience of one patient is used here as an example.

PATIENT PRESENTATION

A 75-year-old man has enlargement of the prostate (benign prostatic hyperplasia, BPH) which results in a frequent need to get up at night to urinate

(nocturia). This is a common condition in elderly men. Getting up as many as five times a night is inconvenient but is not a life-threatening condition. The patient's primary care physician prescribed the drug tamsulosin hydrochloride (Brand name Flowmax) to be taken one pill (capsule 0.4 mg) a day.

The patient was taking no other medications. The local pharmacist gave him a standardised fact sheet for patients when the prescription is filled. According to this fact sheet, two expected benefits are to reduce the need to urinate often and a stronger urine stream. Two side effects were dizziness and a runny nose.

The patient experienced symptoms of dizziness which were apparent. With the primary care physician's permission, he started to take the pill every other day and to measure the end-results by the number of times he had to get up at night. Each day he kept data on the date, whether or not he took a pill, the number of times he got up at night, the level of exercise on the day before and notes on special cause variation. From 15 July to 11 September 2014 he had collected 55 days of data for his trial. For 28 of these days, after taking the pill he got up 2.22 times on average. Without the pill he got up 2.393 times in 27 days. The overall rate of 'ups' for these 55 days was 2.309.

Global health problem analysis

For this patient with BPH there are four outcomes of potential concern: two results and two side effects. This medication results in the need to trade off the benefits against the side effects. The two benefits and two side effects that are described here were given a utility score that probably differs from patient to patient. The medicine resulted in an improvement in



CrossMark

To cite: Neuhauser DV, Chu J. *BMJ Innov* 2017;3:157–162.
BMJ Innov 2017;3:157–162. doi:10.1136/bmjinnov-2016-000177



this person's health. The usual and customary cost of this drug is US\$119.99 for a 30-day supply. With insurance coverage this drug costs the patient US\$12 for 30 pills.⁴ On request, the technical package insert from the pharmaceutical firm Sandoz⁵ was provided by the pharmacist.

Package insert

The clinical evidence in support of this drug is of very high quality. Two studies are described and the statistical analysis is commendable. Controlling for the placebo effect, the evidence for efficacy is strong. That said, there is some missing information that could help the patient to answer the question: Does this treatment work for me? Two outcome measures are used in the two reported trials: urine flow and the American Urological Association (AUA) symptom score. A literature reference to this score is not given. There are no literature references in the package insert. This individual patient would like to see a distribution of the outcome scores. More people benefit from the drug more than from the placebo effect. In how many patients is there no or minimal benefit? These people could avoid the costs and risks by not taking a pill that does not benefit them.

Measurement issues

The dependent variable is driven by patient preferences and may be unique to a specific person. Measurement can change over time. In this example, getting up at night turns out not to be so easy to measure. No daytime measures were made because the symptoms were not bothersome during the day. Eventually, this person's dependent variable was measured, for example, as follows: sleep/up/sleep/up/sleep/up for the day. For an 'up' to count, it had to be bracketed by 'sleep' before and after. This example yields a score of 2. It took several months to decide that this is the appropriate measure for this man. Changes in measuring the variables over time make comparisons possible but risky.

Placebo effect

Blinding is used to control for the placebo effect when the question is: Does this treatment provide benefit by itself independently of the placebo effect? In our example, the placebo effect is part of the benefit of treatment and it is purposely not controlled for.

Exercise

This factor was chosen as a control variable because exercise seemed to lead to better sleep. Some people are involved in organised exercise programmes and could randomly exercise or not to see if this has an impact on the dependent variable. In this person's case, the decision to exercise was weather-dependent such as subzero temperatures, snow shovelling, yard work, log splitting, rain and walking. The measurement of

exercise evolved into a 4-point scale for this person: baseline, every day exercise (E) included climbing 10 flights of stairs and routine walking E(0); an additional 15 min of heavy exercise E(1); half an hour of exercise including being out of breath and perspiring E(2); and over an hour of such exercise E(3). This level of exercise could be too light for a 25-year-old man.

Side effects

Of the two side effects, having a runny nose was not considered important and was ignored. On the other hand, dizziness was important for this person. The desire to reduce dizzy spells was one reason the alternative day schedule for taking the drug was started. It turned out that this side effect was almost entirely explained by one behaviour pattern. It occurred after he sat at his computer for half an hour or more, and then got up and promptly climbed a flight of stairs. This could be both predicted and controlled for.

Special cause variation

This is a concept that comes from quality improvement thinking. If there is a clear known plausible explanation for an extreme outlier, there is a case for dropping it from the analysis. This can narrow the distribution for the series of outcomes and provide a clue to understanding variance. An example of special cause variation here was when the neighbour's house totally burned down in the middle of the night. The fire was fought by crews of seven fire trucks. This was counted as a special cause and excluded from the analysis.

There are many reasons for sleeping well or not, too many to measure. The large family cat jumping on the bed at 04:00 hours is a cause of waking up. These variables can be understood even if only a few can easily be measured. One measure of successful care is that the patient has an accurate multivariate model of his condition.

The factorial design of experiments, can address the following questions "Does this pill and exercise improve my health? Does the combination do good more than one alone?"^{6 7} the usual randomised controlled trial (RCT) design compares two groups of patients. One group gets the drug and the other is the control group. Factorial design refers to the use of more variables at once. This can show both first-order and interaction effects between the individual variables. This approach was used in this person to analyse three variables and their interaction, thereby gaining more information than one would get from a one-variable trial. These variables were taking the pill, level of exercise and number of times up at night and the interaction effect of exercise and pill combined (see table 1).

Note that in table 1 there is an interaction effect between exercise and pill taking. Without exercise, taking the pill reduced the frequency of getting up at night from 3.00 times to 2.65 times. With low, medium and high exercise taking, the pill made

Table 1 Clinical variables measured to assess the need for medication for nocturia at different levels of exercise (E)

Exercise		Number of days observed	Average times up at night
	Taken pill		
Minimal E(0)	Yes	20	2.65
	No	15	3
Low level E(1)	Yes	12	2
	No	16	1.88
Medium E(2)	Yes	10	2.9
	No	11	2.09
High E(3)	Yes	20	2.2
	No	24	2.13

his health status worse. This interaction effect was surprising and led to a new treatment strategy, to take the pill only when there has been no exercise E(0) and to try to increase the number of days with exercise. This was done for 12 days from 24 April to 5 May 2015. During this time period all but 2 days were exercise days, so two pills were taken in 12 days. The average 'up' score during this time was 1.67. If the two pill days were excluded, the average 'ups' would be 1.3 times per night. To summarise, pill use reduced from one a day to one in 6 days. This is a cost reduction of 83%. Health status as measured by the number of 'ups' went from 2.309 in 2014 to 1.67 in 2015. This is an improvement of 28% or 0.64 'ups' per night, which the patient decided was a clinically significant difference for him. The data on exercise suggest that there is room for more improvement by increasing the number of high exercise days.

Building a multivariate individualised model of a chronic condition

This person continued to record on a daily basis five variables related to his chronic condition for 16 months from 1 April 2015 to 1 June 2016, resulting in about 450 person-days of data. This recording took very little time and became a habit. Variables tracked included the following:

1. Number of 'ups' per night: the dependent variable.
2. Taking the prescribed medicine; this went to zero use.
3. Level of exercise measured on a 4-point scale.
4. Number of flights of stairs climbed per day: In the two-storey house used in the winter, there was little variation, from 15 to 20 flights per day. This measure was independent of the exercise measure. Spread out over the day rather than all at once meant that this climbing had little cardiovascular impact. The summer house has no stairs and counting this variable was discontinued during the summer months.
5. Number of hours in bed per night: the longer in bed the more hours of exposure to 'ups'.
6. Asleep before 21:00 hours: this is related to longer hours in bed and light sleep.
7. Special cause variations. Travel: he took seven short trips in a year; these days were excluded. Weather:

snow shovelling led to more exercise, rain led to less exercise. Week days versus weekends: did not make much difference to this retired person.

8. Common cause variation. There are unexplained apparently random fluctuations in the dependent variable. Using behaviour data for the previous day, the dependent variable was predicted. If the predictions are accurate, then there is little common cause or random variation. These predictions were often inaccurate. This variation can be seen as analogous to the error term in a regression model.
9. Measurement error. Forgetfulness and bias did occur. To track stairs climbed, a dish and pennies were used. After each flight was climbed, a penny was added to the dish and these were added up at the end of each day. This simple method substantially reduced errors in this measure.
10. Long-term variation month by month for about 2 years: this is now a stable process.
11. Interaction effects combining more than one variable would lengthen this list. With a sample size of 450 person-days, there are enough observations to consider the effect of a lot of independent variables.

How to encourage people to do this kind of individual analysis

What is in it for the patient? The quantified self-movement shows that some people are interested in doing this. The idea of the certified patient is that patients who have accurate multivariate models of their chronic conditions should be allowed to write their own refill prescriptions. Perhaps such patients should have a red carpet entrance and valet parking at their clinic. Participant's insurance could pay for expensive non-generic drugs which are shown to help them. Perhaps non-participants get coverage for only generic drugs or pay a premium for more expensive alternatives.

DISCUSSION

The example presented here shows how measurement and factorial design can be used to understand a person's health condition, lower costs and improve measured quality. The patient's physician recommended the medicine taken by our patient; it is possible that there are newer better medicines available. The patient chose the outcome measures and measured them based on his unique preferences. The patient's preferences were based on his opinions. This person's specific results are not being used to generalise to anyone else.

The quality and side effects measures are also unique to this individual. It is proposed that this approach to 'the quantified self'⁸ could be a companion and not a substitute to other sources of clinical evidence such as clinical trials, clinician knowledge and the pharmacist's expertise. Randomised trials require an identical outcome measure across all subjects. Here the 'utility score' chosen was unique for this patient.

With regard to statistical versus clinical significance, because there is no desire to compare this unique individual with another population, statistical significance is not an issue. Clinical significance is important. Is the reduction of one 'up' in 5 days worth the effort? If not, how big does the difference have to be to become worth the costs and side effects for this person? For this person, the difference between 3 and 2.65 'ups' shown in [table 1](#) is clinically significant. It would be interesting but not necessary to have baseline data before the medication is started. Delaying treatment for a month or two would often not be acceptable.

If this method costs next to nothing to do, saves money and improves measured health status, why is it not used ubiquitously? What are the barriers? A drug or treatment may lose its effect over time. Conversely, the long-term effect may be beneficial. How do physicians find this out for individual patients and adjust therapy accordingly? Physicians do not have the time. There is no reimbursement for this. Reducing pill use is not in the drug company's short-term economic interest but, in the long run, having drugs that really benefit is in their economic interest. There are programmes online which can be used to record individual data, but these companies may own the data and resell them for their profit.

Lots of people do not like numbers. What if everyone who has a prescription costing more than \$50 a day would only be reimbursed if they joined a support group and routinely collected their own data? Patients who reduced their costs and improved their health status would get half the savings.

Human subject review is not designed to deal with personal improvement efforts like this. A person should be able to carefully examine their own health as a matter of individual liberty and freedom and not need anyone's permission to do so. Therefore, permission was not sought. These data were not entered into an online site which would then own the data.

There are lots of fancy statistical methods that could be applied including control charts and regression models,⁸ utility scores and decision analysis. Levels of exercise could be randomly or purposely assigned. To do so would miss the point of this example, which was directed to show that an average person could do this analysis at minimal cost. It is proposed that persons can bring new essential data about their own health into the technology assessment process.

Measurement errors are certainly possible. Regular daily recording of the measured variables reduces but does not eliminate this problem, particularly with short-term memory loss.

In the process of daily recording, it is possible to develop a multivariable model for this person's condition. There are many variables that influence sleep and its duration. The variables can be measured differently and priorities change over time.

This method is not appropriate for constant monitoring and is less useful when variables are measured less often than daily.⁹ In between these extremes there is a large medical domain including exercise, hypertension, stress, blood sugar,⁸ smoking, substance abuse, asthma,¹⁰ diet, study habits, persistent pain¹¹ and other behaviours.^{12 13}

The results of a clinical trial could be as follows. There are 100 patients in both control and experimental groups. In the control group, 20 patients benefit and 80 do not. In the experimental group, 40 patients benefit and 60 did not. This is a statistically significant difference, the drug company is delighted and their stock soars. If we could find the 60 patients who did not benefit, there are substantial healthcare savings to be achieved. Subanalyses of the trial results can be done. Perhaps such an analysis shows, for example, that short patients benefit more than tall patients. The trial is too small to answer such questions. It would be better to do an analysis for each unique patient. Thus, there are potential savings and benefits for not taking a drug that does not work for some patients.

The best evidence to support or reject the usefulness of the methods described here is in replication. Try this method yourself for your own condition and decide for yourself. The consequences of this individualised approach to improved care could massively change healthcare as we now know it. We propose to generalise this method. This patient's preferences are not generalisable and are unique to that person. Ethical application is that patients take their own medicine or not, based on personal freedom. Our methods could allow them to do this more systematically if they wish.

Statistical implementation

In order to answer whether this individual approach adds to knowledge obtained from RCTs, which use many patients and a standardised outcome measure, there are some important differences.

1. The unit of analysis: by changing the unit of analysis from one patient to one patient-day, it allows for the collection of a lot more information quickly and cheaply. Generalisability: there is no goal to generalise the specific results beyond this individual. This avoids the need for a lot of statistical analysis to show differences between groups due to chance. This patient is not a sample, but the statistical universe.
2. The dependent variable: RCTs require the use of the same dependent variable for all participants measured in the same way. But patient outcome preferences differ and need to be measured differently to obtain an optimal outcome for that person. Similarly, for the measurement of side effects, there are four consequences of interest to this patient. Of these, two got utility scores of 1 (times up and dizziness) and two outcomes which had no value for this patient: urine stream and runny nose (utility scores: zero). Other patients could have different utility scores and possibly a different optimal treatment. These utility scores could change with time. When one outcome is

managed, another outcome may start having relevance and a positive utility score. The AUA score used in the trials is the same metric for all patients. An individual would want to know the distribution of these scores in order to design their own optimal treatment. The patient might wish to know how the symptoms were actually measured and the distribution of these scores in the trials in order to design their own optimal treatment.

3. Randomised versus alternative assignment: randomised assignment of patients in a clinical trial is expected. In this example of a factorial design, an alternate day approach was used. A rule followed by the main author of this manuscript was that, when there was a close call choice between simplicity and statistical rigour, simplicity was chosen. For example, alternate days of treatment or no treatment was chosen rather than assignment by use of random numbers. Simple measures were chosen over regression models. The placebo effect was not excluded or controlled for because it may be one of the benefits of treatment. The cost of one patient doing this is vanishingly small. By contrast, a randomised trial can cost millions of dollars. We argue that both our method and RCTs are complementary and could beneficially go together.
4. Alternating treatment and non-treatment for this person seems unlikely to be biased by unknown variables. Daily alteration makes sense from the pharmacokinetics of this drug related to the rate of absorption. Alternating treatment and non-treatment days combined with measuring treatment on day 1 and outcome on day 2 create the conditions that can show causation and control for confounding variables. The unit of analysis as the patient-day rather than the patient assumes independence between days. This way, in 2 months there can be 30 treatment and 30 non-treatment daily observations rather than one single patient.
5. Costs of the study: the standard RCT can cost millions of dollars and take years to do. This individualised factorial approach costs little or nothing, depending on your perspective. In this example, it requires a pen and paper to record at least four data points each day and a simple pocket calculator for addition and division.
6. There is no large incentive for people to do this method of analysis. There is no product for which a company can make money. An incentive might be to require this analysis if health insurance is to pay for expensive drugs. For patients who want to avoid unpleasant side effects, this method might guide their decision.
7. Systematic review showed that the effectiveness of tamsulosin was similar to other alpha-blocker drugs. Future studies should focus on long-term effectiveness and whether use of tamsulosin alone or in combination with other drugs that shrink the prostate can prevent urinary retention and/or the need for surgical intervention.¹⁴ Tamsulosin is no longer under patent protection. An April 2017 search of the Cochrane Reviews found that the review for this drug was withdrawn to await updating.

CONCLUSION

Our goal was to make this method of self-evaluation as inexpensive and easy to use as possible. This method does what many patients do on their own.

'I forgot to take my medicine last night and today I feel worse.' As presented, this method is not useful for continuous monitoring such as the continuous measurement of blood pressure. Monitoring devices may not present data in a form that could guide patient decision-making.

This method is proposed as a complement to RCTs and not a replacement. Our experience is that readers have lots of questions, perhaps due to the several concepts that underlie this work. Instead of reviewing these concepts here we recommend that interested readers should try this method for themselves. Replication is the gold standard of scientific evidence.

Correction notice Confirmed that all corrections done and completed and all statements are correct

Twitter @stopmusclspain

Contributors DVN is the main author. JC is the co-author and did the editing, proofing, reformatting, contributed to the references and is also the corresponding author.

Competing interests None declared.

Patient consent Obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data concerned with this manuscript are available from DVN.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- 1 Alemi F, Neuhauser D, Ardito S, et al. Continuous self-improvement: systems thinking in a personal context. *Jt Comm Qual Improv* 2000;26:74–86.
- 2 Olsson J, Terris D, Elg M, et al. The one-person randomized controlled trial. *Qual Manag Health Care* 2005;14:206–16.
- 3 Neuhauser D. Why design of experiments just may transform health care. *Qual Manag Health Care* 2005;14:217–8.
- 4 Eri LM, Tveter KJ. Alpha-blockade in the treatment of symptomatic benign prostatic hyperplasia. *J Urol* 1995;154:923–34.
- 5 Narayan P, Lepor H. Long-term, open-label, phase III multicenter study of tamsulosin in benign prostatic hyperplasia. *Urology* 2001;57:466–70.
- 6 Montgomery D. *Design and analysis of experiments*. 5th edn. New York: John Wiley and Sons, 2001.
- 7 Grant EL, Leavenworth RS. *Statistical quality control*. 7th edn. New York: McGraw Hill, 1996.
- 8 Solodky C, Hegang C, Jones P, et al. Patients as partners. *Medical Care Supplement* 1998;36:AS 13–20.
- 9 Topol EJ. *The patient will see you now*. New York: Basic Books, 2015.
- 10 Alemi F, Neuhauser D. Control charts for time between asthma attacks. *Jt Comm J Qual Saf* 2004;30:95–102.

- 11 Chu J, Bruyninckx F, Neuhauser DV. Chronic refractory myofascial pain and denervation supersensitivity as global public health disease. *BMJ Case Rep* 2016;2016:bcr2015211816.
- 12 Neuhauser D, Provost L, Bergman B. The meaning of variation to healthcare managers, clinical and health-services researchers, and individual patients. *BMJ Qual Saf* 2011;20 (Suppl 1):i36–i40.
- 13 Duan N, Eslick I, Gabler NB, et al. *Design and implementation of N-of-1 trials: a user's guide*. AHRQ publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality, 2014. <http://www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm>
- 14 Wilt TJ, MacDonald R, Rutks I, et al. Tamsulosin for benign prostatic hyperplasia. *Cochrane Database of Systematic Reviews*, 2011.