2011

Empirical Pilot Study Variable Selection Using Value Tree Analysis

Monifa Vaughn-Cooke
Gül E. Kremer, *Iowa State University*

Proceedings of the 2011 Industrial Engineering Research Conference
T. Doolen and E. Van Aken, eds.

Empirical Pilot Study Variable Selection Using Value Tree Analysis

Monifa Vaughn-Cooke[1]
Gul Okudan Kremer[1,2]
[1]Department of Industrial and Manufacturing Engineering
[2]School of Engineering Design
The Pennsylvania State University
University Park, PA, 16802, USA

Abstract

The selection of variables to collect in empirical pilot studies is critical due to time, financial and sample size limitations. The collection of patient data in the healthcare system poses additional concerns due to patient privacy, incomplete data and clinical feasibility. The existence of these intrinsic pilot study and healthcare constraints requires a strategic selection of study variables to maximize the variability explained in the patient population and knowledge gained from data analysis. We propose a value tree analysis to select data collection strategies (combination of study variables) that produce the highest preference value for the researcher while also taking into account the interests of other study stakeholders (patient, provider, clinical staff, Institutional Review Board, etc.). In the first stage (problem structuring), the decision context is framed to define the pilot study environment and stakeholder values that may impact the study design. The study objectives, decision alternatives and attributes are then generated to consider tradeoffs among conflicting criteria and stakeholder interests. In the second stage (preference elicitation), decision attributes are weighted using Analytical Hierarchy Processing and a value function is used to elicit the decision maker's preference for each variable combination. Value tree analysis is a low cost and low computation approach for time-sensitive variable selection in empirical pilot healthcare studies. The proposed approach provides a structured and strategic decision-making process that accounts for pilot study and patient data constraints and maximizes value for study stakeholders.

Keywords
Multi-criteria decision-making, healthcare, analytical hierarchy processing, preference elicitation

1. Introduction

Data collection in healthcare is a vital process to insure current evidence is available to inform and support healthcare practitioners in clinical decision making. Clinical evidence includes: patient-centered research on the effectiveness of behavioral interventions; the accuracy and precision of diagnostic tests; the power of prognostic markers; and the efficacy and safety of therapeutic, rehabilitative and preventative regimens. Collecting accurate and statistically acceptable data in the healthcare system is a complex process due to resource and patient privacy constraints. Testing research on a small sample to insure data collection accuracy, statistical variability and significance before embarking on the main study is critical. Just as new clinical procedures are tested before being widely used, aspects of the research process should also be piloted. Thus, a 'pilot' or 'feasibility' study may involve conducting a small-scale version of the full study or trying out a particular research instrument such as a survey, interview technique or other data collection measure.

One of the advantages of performing a pilot study is that it may provide advance warning of potential failures in the main research project or whether proposed protocols, methods or instruments are ineffective [1,2]. In addition, a pilot study can be used to convince funding agencies that a research proposal for the main study is worth funding. Pilot studies can be based on quantitative and/or qualitative methods and large-scale studies may employ several pilot studies prior to starting the main study. Many researchers may start with a qualitative data collection and analysis on a relatively unexplored topic and use the results to design a subsequent quantitative phase of the study [3].

In many research papers, pilot studies are only reported as a means of justifying the methods. This justification might refer to the overall research design, or simply to the validity and reliability of the research tools. It is uncommon for reports of pilot studies to include practical problems faced by researchers in the data collection process. Pilot studies are relevant to provide guidance for the best practices in research, but their potential to aid and help inform the design of empirical studies for other researchers has largely been ignored. This paper aims to provide guidance on addressing some of the major issues facing healthcare system researchers involved in empirical pilot study design. These issues include the selection of variables to be measured when several data collection constraints and conflicting decision-making criteria are present.

Multi-Criteria Decision-Making (MCDM) can be used as an approach to evaluate several aspects of proposed pilot study design alternatives. MCDM is a discipline aimed at supporting decision makers faced with making numerous and sometimes conflicting evaluations [4]. The identification of study objectives and alternatives in a healthcare setting involves all stakeholders that may be impacted by the proposed pilot study. These individuals include healthcare providers, the study participants, clinical staff and administrators, individuals responsible for collecting and analyzing the data and funding agencies. In addition, the Institutional Review Board (IRB), responsible for approval and monitoring of studies involving human subjects at the researcher's organization, will have primary interests in protecting the study participant [5]. Stakeholders will likely have conflicting interests and decision making criteria, thus necessitating the use of a MCDM approach.

Value tree analysis is an MCDM tool that mathematically weights the decision-making attributes for each alternative to elicit a preference for the decision makers. Value functions are used to aggregate alternatives' attribute-specific values into an overall measure [6]. This process is useful in designing pilot studies to insure that the variables selected for measurement in the study include appropriate attribute tradeoffs for each study stakeholders. The value tree analysis steps and their relevance for the collection of healthcare data in pilot studies is discussed in subsequent sections.

2. Problem Structuring

Problem structuring is the first stage of value tree analysis and gives a better understanding of the values affecting the decision and a basis for further analysis. This involves identifying and generating the decision context, objectives, alternatives and attributes.

2.1 Defining the Decision Context

The decision context defines the nature of the problem and its environment. This is a critical component when considering how the decision maker's values may impact the decision outcome. The framing of decision context becomes more complex when a decision has associated consequences [7]. In safety critical decision environments such as healthcare, ethical considerations are particularly important [8]. Undesirable consequences resulting from study design and the selection of variables to measure, may compromise the health and well-being of study participants [9,10]. Consequences in complex situations with conflicting alternatives will increase the reliance on values as a guideline for structuring the decision context.

For example, the IRB and the study participants' primary interest is for the study to pose minimal risks (physical, emotional, financial, legal) for those involved [5]. However, the researcher may want to select variables for effective measurement of an outcome that conflict with these risk minimization goals. The concept of tradeoffs, as demonstrated in the aforementioned example, is central to value tree analysis. In addition to the undesirable consequences of risks borne by the subject, high risk studies will also have an impact on the researcher, including: 1) a reduction in the enrollment rate for eligible subjects; 2) a higher subject drop-out rate; and 3) a reduction in subject numbers and the statistical power for future predictive and descriptive analyses. Higher risk studies will also require a longer review process for IRB approval, which may conflict with the researcher's pilot study timeline.

It's important to include all stakeholders when evaluating the decision context, to insure that conflicting interests and their impact on the pilot study are considered. Several questions must be asked to frame the decision context (Table 1).

Table 1: Decision Context Questions

Question	Response
Who will make the decisions?	For the purposes of this analysis, the primary decision maker is the researcher or Principal Investigator.
What are the decision maker's values and interests for the study?	In a pilot study, the researcher may value time, money and other resources. Most, importantly, the researcher values demonstrating adequate feasibility of a concept or outcome for future empirical testing in the main study
What are the decision alternatives?	The decision alternatives are combinations of prognostic variables that may be significant contributors to understanding the outcome measure.
Which stakeholders are affected by the decision and how are they affected?	The decision stakeholders include all individuals directly or indirectly impacted by the study. The interests of each stakeholder are defined by their values applied to the relevant decision context.
What is the purpose of the analysis?	The fundamental objective for the value tree analysis is to strategically select variables to empirically measure in a pilot study. The selected variables should quantitatively demonstrate a significant relationship between independent and dependent variables or provide qualitative explanation of an outcome measure.

2.2 Identifying the Objectives

After framing the decision context, the pilot study objectives are defined. Requirements elicitation is the practice of obtaining the requirements of a system from users, customers and other stakeholders [11]. This practice can be used to perform a qualitative evaluation of the needs, requirements and limitations of stakeholders to generate pilot study objectives. A requirement is a "function or characteristic of a system that is necessary...the quantifiable and verifiable behaviors that a system must possess and constraints that a system must work within to satisfy an organization's objectives and solve a set of problems" [12].

Several tools can be used to perform requirements elicitation and generate objectives [6]. These include:

- Literature Review: Examination of relevant literature
- Analytic Study: Development of a conceptual framework to integrate literature
- Expert Interviews: Discussions with stakeholders and observations in the decision context

Objectives can be divided into two classes, fundamental and means objectives [13]. Fundamental objectives characterize an essential reason for interest in the decision context. Means objectives are of interest in a decision context because they are a means to achieving fundamental objectives. Typical objectives for a pilot study in healthcare are shown in Table 2.

Table 2: Fundamental and Means Objectives

Objective		Description
Show initial concept feasibility	Fundamental Objective	In pilot studies, several outcomes have to be achieved to show initial feasibility for a theoretical model or concept. These include: population variability, sample size, and measurement of the appropriate type and number of variables. In addition, more practically, the study must adhere to time and resource limitations.
Achieve a sample size	Means objective	Performing power analyses and sample size estimation is an important aspect of experimental design, to set recruitment/enrollment goals. If sample size is too low, the experiment will lack the precision to provide reliable answers to the questions it is investigating. If sample size is too large, time and resources will be wasted, often for minimal gain.
Measure independent and dependent variables	Means objective	To empirically build a theoretical model and answer research questions, the researcher must be able to measure a list of prognostic variables for some outcome or dependent variable. In pilot studies, the small sample size and time constraints typically limit the number of variables that can be measured and statistically analyzed, so these variables must be carefully selected. Requirements elicitation can be used to generate a preliminary list of variables.
Collect data within a specified time period	Means objective	One of the main constraints in pilot studies is the ability to collect the proposed data within a specified time period. These studies may be driven by funding proposal and resource availability deadlines.
Collect data using available funding	Means objective	Typically, studies designed to show initial feasibility are designed with the purpose of preparing funding proposal for larger scale studies in the future. Thus, pilot study funding is limited. Funding requirements may include subject compensation, equipment, travel and research assistant funding.

2.3 Identifying the Decision Alternatives

Decision alternatives are unique combination lists of the independent variables to be considered for measurement in the study. If the study objectives include several measured outcomes, then multiple dependent variables may be assessed as alternatives. These combinations will be limited by inclusion and exclusion pairing constraints. For example, practically and statistically, it is not advisable to waste resources by measuring two variables that are closely related and will most likely result in multicollinearity. On the other hand, the inclusion of one variable may necessitate the inclusion of another to fully understand the meaning of each variable independently. These potential relationships must be considered when forming decision alternatives.

2.4 Specifying the Attributes

An objective is measured in terms of an attribute. Based on requirements elicitation, decision-making attributes must be selected to comprehensively represent the interests of each stakeholder. This process is a critical step of value tree analysis and will help to identify the alternative with the highest perceived value for the researchers and the success-critical variables for the data collection process. There are several stakeholder concerns that arise during an empirical pilot study. Typical attributes for a healthcare pilot study are shown in Table 3.

Table 3: Decision Attributes

Attribute	Detail
Time taken to collect data	Several data collection tools can be used to collect healthcare data, which include newly created data and existing data. Newly created data includes psychometrics and subject interviews, which involve a more time intensive patient recruitment and data collection process. Existing data includes patient records and databases that can be accessed quickly with appropriate patient consent.
Time for subject to provide data	If existing data is used, the subject will not have to provide data, only consent to access the data. On the other hand, if new data is created, the time to collect that that data must be taken into account.
Time taken to input data	Although data input is a process performed after data collection, this may be a critical attribute. Since many healthcare providers still use paper-based records [14], manual input, particularly for longitudinal data, may be tedious and restricted by time and resources. In addition, if a survey scanner or voice transcriber is not available, manual input of psychometric responses and interviews may be limited. Data input time can be estimated with a trial run of the data collection tools under consideration or a rough estimation.
Funding required to collect data	The type of data being collected may have an impact on required funds. This includes payment for individuals being reimbursed (hourly, monthly, etc.) for assisting in data collection, travel costs that may increase due to a longer study duration and the need for additional equipment to collect a particular type of data.
Probability of missing data for the measurement	Missing data in empirical studies is a concern due to its impact on statistical analyses. Missing data may be driven by the type of patient recruited, the type of data collected, the length of the study or the time required to provide data. For example, if a return visit is required to collect data on a specific variable, a lower study completion rate can be expected. Preventative measures such as removal of certain variables or data collection tools, can be evaluated for this attribute.

3. Preference Elicitation

By mathematically evaluating the variable selection alternatives against each decision attribute, the researcher's preference and pilot study design can be calculated. A value function assigns a number V(x) to each consequence to indicate the relative desirability of the consequence and can be used to derive preferences for alternatives. The value function incorporates: 1) the attribute weights determined using Analytical Hierarchy Processing (AHP) and 2) the Direct Rating or natural scale value for each variable combination set.

3.1 Determining Preferences on Attributes: AHP

AHP is a pair wise preference elicitation technique, used for complex decisions [15]. For Value tree analysis, AHP is used to weight the decision attributes by comparing the rows (i) and columns (j) of an attribute matrix. The comparison value (a_{ij}) is based on the decision-maker's perceived level of importance for each attribute. The following scale values are used for matrix comparison (Table 4).

Table 4: AHP Matrix Comparison Scale Values

Value a_{ij}	Comparison Description
1	Objectives i and j are of equal importance
3	Objective i is weakly more important than j
5	Objective i is strongly more important than j
7	Objective i is very strongly more important than j
9	Objective i is absolutely more important than j

The ratings (R_i, R_j) for the comparison are then normalized as per Equation 1. The normalized weights (w_i) are averaged across the columns to calculate the final weights, which represent the decision-maker's perceived attribute importance.

$$w_i = \frac{R_i}{\sum_{j=1}^{n} R_j}, \quad \sum_{i=1}^{n} w_i = 1 \tag{1}$$

3.2 Attribute Value Scaling

After weighting decision attributes, values are assigned to each attribute across each decision alternative. Natural values are used in cases where attributes can be measured in natural scale (i.e., time, money, travel distance). All other variables should be rated directly on a scale to represent the effects of the decision alternative on an attribute. In Direct Rating, the worst and best consequences with respect to a certain attribute are associated with values of 0 and 100 respectively. The values of the intermediate (between 0 and 100) levels are determined only by the alternatives under consideration. This approach is most appropriate when natural scales are not applicable, there are no commonly agreed scales of measurement or there is no time or resources to undertake measurement.

3.3 Value Function

Several value function models can be assessed based on the relationship between attributes, which will be specific for each pilot study design. For a set of mutually independent attributes, the value for the set can be found by summing the value for each attribute (Equation 2). The weight w_i corresponds to the change in the strength of the preferences as the attribute x_i changes from the worst to the best level.

$$V(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} w_i v_i(x_i) \quad \text{Where, } \sum_{i=1}^{n} w_i = 1 \tag{2}$$

The additive model shown in Equation 2 describes the decision makers' preferences only if the attributes are mutually preferentially independent. Thus, there cannot be synergies between attributes. For example, if time to collect data on a specific variable is increased and patient surveys are used as a data collection tool, this may impact other attributes. The patient may skip questions in an effort to rush through a longer survey, thus increasing the probability of missing data for the measurement. The required funds may also be connected to time required to collect data. If preference independent cannot be proven, another model form, such as a multiplicative model can be used.

In a multiplicative model (Equation 3), a constant K is used as a nonzero solution to the equation and k_i represents the value of an outcome having the best level on the attribute x_i and the worst on others. In this model, single attribute value functions have a multiplicative effect on the total value.

$$V(x_1, x_2, \ldots, x_n) = \frac{1}{K} \prod_{i=1}^{n} (K k_i v_i(x_i) + 1) - 1 \quad \text{Where, } K + 1 = \prod_{i=1}^{n} (1 + K k_i) \tag{3}$$

4. Discussion

The proposed approach aims to support healthcare pilot study design and variable selection by evaluating potentially competing criteria such as time, resource constraints and data collection risks. Value tree analysis is a low cost and computation approach to assess the preference values of each study stakeholder and form a strong empirical foundation for a main study. The first stage of value tree analysis (problem structuring) requires the researcher to perform interviews, observations and literature reviews to comprehensively understand the decision context and generate study objectives. This activity allows the researcher to accurately understand the tradeoffs associated with conducting a pilot study and their resulting impact on enrollment, missing data and resources. The preference values found in the second stage (preference elicitation) provide the combination of study variables that have the highest perceived value for the researcher, while also considering the preference of other study stakeholders.

Since the decision-maker (researcher) is a stakeholder *and* the generator of each alternative, decision-maker bias could be present in the value tree analysis. The researcher is the most familiar with the objectives and their impact across each stakeholder and would be the most qualified to weight the impact of each objective. However, if attribute tradeoffs and their impact across all stakeholders are not adequately considered, this bias may negatively impact the outcome of the study and produce undesirable consequences. A more unbiased process would be to integrate the other decision stakeholders in each value tree step. Each decision-maker or stakeholder in this case could have more weight in the preference elicitation process, thus providing a more robust analysis.

5. Conclusion

The value tree analysis was used to determine stakeholder preference for the selection of experimental variables for pilot studies in healthcare. These studies have the potential to form a strong statistical foundation for future empirical studies to provide support for clinical decision making. The decision alternative objective and constraints are defined to better understand the requirements of each proposed alternative and their impact on study stakeholders. Attributes are developed to elicit decision requirements from the decision maker and to allow a comparison of objective combination sets against each attribute using value functions. The result of the value tree analysis is a selection of variables to measure in the pilot study that optimizes tradeoffs among the study attributes to adequately satisfy stakeholder needs.

References

1. Lancaster, G.A., Dodd, S., and Williamson, P.R., 2004, "Design and Analysis of Pilot Studies: Recommendations for Good Practice," Journal of Evaluation in Clinical Practice, 10(2), 307-12.
2. Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L.P., Robson, R., Thabane, M., Giangregorio, L, and Goldsmith, C.H., 2010, "A Tutorial on Pilot Studies: The What, Why and How," BMC Medical Research Methodology," 10(1).
3. Tashakkori, A., and Teddlie, C., 1998, Mixed Methodology: Combining Qualitative and Quantitative Approaches, Applied Social Research Methods, Sage, Thousand Oaks, CA.
4. Ravindran, A.R., 2009, Operations Research Methodologies, CRC Press, Boca Raton.
5. Gittner, L.S., Roach, M. J., Kikano, G., Grey, S., and Dawson, N. V., 2011, "Health Service Research: The Square Peg in Human Subjects Protection Regulations," Journal of Medical Ethics, 37(2), 118 - 122.
6. Keeney, R.L., and Raffia, H., 1976, Decisions with Multiple Objectives: Preferences and Value Trade-Offs, John Wiley & Sons, New York.
7. Hosack, B., 2007, "The Effect of System Feedback and Decision Context on Value-Based Decision Making Behavior," Decision Support Systems, 43(4), 1605–1614.
8. Chae, B., Paradice, D., Courtney, J.F., and Cagle, C.J., 2005, "Incorporating an Ethical Perspective into Problem Formulation: Implications for Decision Support Systems Design," Decision Support Systems, 40 (2), 197-212.
9. Steinbrook, R., 2002, "Improving Protection for Research Subjects," New England Journal of Medicine, 346,1425–1430.
10. Silverman, H., 2007, "Ethical Issues During the Conduct of Clinical Trials," Proceedings of the American Thoracic Society, 4(2), 180-4.
11. Mate, J. L., and Silva, A., 2005, Requirements Engineering for Sociotechnical Systems, Information Science Publishing, Hershey, PA.
12. Sommerville, I. and Sawyer, P., 1997, "Viewpoints: Principles, Problems and a Practical Approach to Requirements Engineering," Annals of Software Engineering, 3, 101-30.
13. Keeney, R., 1992, Value-Focused Thinking: A Path to Creative Decisionmaking, Harvard University Press, Cambridge, Massachusetts, London, England.
14. Jha, A.K., DesRoches, C.M., Campbell, E.G., Donelan, K., Sowmya, R.R., Ferris, T.G., Shields, A., Rosenbaum, S., and Blumenthal, D., 2009, "Use of Electronic Health Records in U.S. Hospitals," New England Journal of Medicine, 360(16), 1628-1638.
15. Saaty, T. L., 1980, The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, McGraw-Hill International Book Co., New York, London.