THE STRUCTURE OF STAGES IN THE EVALUATION CYCLE: AN EVENT SEQUENCE ANALYSIS

Clarke Wilson
Canada Mortgage and Housing Corporation
Ottawa, Ontario

Abstract: From time to time, research on evaluation issues deals with a sequence of events or episodes which an analyst would prefer to examine in its entirety rather than event by event. However, quantitative methods for comparing sequences of events or activities and analyzing their similarity among populations and samples have rarely been applied in social research. This article describes the tasks undertaken during the cycle of evaluation of housing programs by CMHC prior to 1997 and analyses the major evaluation products in terms of the similarity of constituent tasks. A brief introduction to combinatorial methods for measuring similarities among whole character sequences is presented and new research applications are proposed. The analysis uses a new program for sequence alignment analysis called ClustalG, which is a modification of the ClustalX program used in molecular biology.

Résumé: Il arrive à l'occasion qu'une recherche sur des questions d'évaluation se penche sur une suite d'événements ou d'épisodes qu'un analyste préfère examiner dans son ensemble plutôt que séparément. Cependant, les méthodes quantitatives qui existent pour comparer des séquences d'événements ou d'activités et analyser leur similarité parmi des populations ou échantillons sont rarement utilisées en sciences sociales. Cet article décrit les différentes tâches entreprises durant le cycle d'évaluation des programmes d'habitation par la Société canadienne d'hypothèques et de logement (SCHL) avant 1997 et analyse les principaux produits des évaluations de la Société selon leur similarité en ce qui a trait aux tâches constituantes. De plus, l'article résume brièvement les méthodes combinatoires pour mesurer les similarités parmi des séquences de caractères et suggère de nouvelles applications de recherche. On utilise un nouveau programme d'analyse séquentielle, nommé «ClustalG», qui est une modification du programme «ClustalX» employé dans la biologie moleculaire.

████████   Many social processes are best examined as sequences rather than as discrete events. For example, it may be of interest to know whether certain sequences of occupations, say at the beginning of a working career, are more likely to lead to high income, infrequent unemployment, or some other result. Another example would be to determine how efficient children are doing homework. Holding homework time constant, are some patterns of homework episodes more likely to result in completed homework and good scholastic results than other patterns? To answer such questions, some measure of the similarity of sequences of events is necessary.

Methods for measuring similarity and differences among sequences of qualities, states, or events were not available until the 1970s and are still almost unknown in social science. Social research on sequential data such as life or employment histories, business processes, or time-use patterns has relied almost exclusively on narratives or content analysis. These provide only weak measures of similarity of the processes as a whole, or of parts of processes. Development of sequence alignment, or optimal matching algorithms as they are also called, provides quantitative measures of similarity or difference between character sequences.

This article briefly introduces the techniques of sequence alignment and illustrates their application using a recent exercise undertaken in the evaluation division at Canada Mortgage and Housing Corporation that identified the steps that typically were followed during the cycle of program evaluation. It then compares the structure of the stages of the evaluation cycle using the measures of sequence similarity produced by the ClustalG software package.

SIMILARITY MEASUREMENT USING ALIGNMENT ALGORITHMS

Many social processes consist of sequences of events, and the tasks or events which comprise the process can be represented by characters or symbols. The similarity between two sequences can be seen by writing the characters of the sequences as the columns and rows of a table and marking matches between column and row elements with an asterisk. Two such tables are illustrated in Table 1.

Intuitively, we can see that the character sequences [although] and [though] are identical if we ignore the [al] characters. When identical sequences are written as a dot matrix as above, they display a strong diagonal pattern, with occasional off-diagonal elements where

letters are repeated. More distantly related sequences show weaker diagonal patterning.

The sequences [wherever] and [everywhere] are composed of nearly identical syllables but in a different order. The similarity of the syllables is illustrated by short downward sloping patterns in the top right and bottom left corners of the table. The reversal of the order of the syllables is illustrated by the displacement of the two lines.

If the characters represent tasks within a process, for example the drafting of an evaluation framework, one can see that it is possible to analyse processes in a variety of ways. Different processes can be compared. Processes that are in principle the same, such as preparation of evaluation frameworks for two different programs, may be examined to measure how similar they are. Three or more processes can be compared to categorize them according to relative similarity.

It is easy to imagine that the simple tabular device of illustrating matching sequence elements with an asterisk can be formalized to produce more powerful quantitative techniques. The first account of alignment methods including an explicit scoring system for measuring sequence similarity was published by Needleman and Wunsch

Table 1.
Dot Matrix Analysis of Character Sequences

	a	l	t	h	o	u	g	h
t			*					
h				*				*
o					*			
u						*		
g							*	
h				*				*

	e	v	e	r	y	w	h	e	r	e
w						*				
h							*			
e	*		*					*		*
r				*					*	
e	*		*					*		*
v		*								
e	*		*					*		*
r				*					*	

(1970). This is a classic source and is quite comprehensible to non-biologists. A number of combinatorial theorems were published during the 1970s that established the mathematical basis for different approaches to sequence alignment, and the methods have become standard in molecular biology since the 1980s. The seminal text covering applications to many fields is Sankoff and Kruskal (1983). A more recent text on biological applications is Waterman (1995). A good introduction to the methods and applications in social sciences is Abbott (1995). Wilson (1998a, 1998b) applied sequence alignment methods to time-use data gathered through diary surveys.

Sequence alignment is a procedure by which one sequence, called a source, is transformed into another sequence, called a target, using a set of eligible operations. The optimal alignment(s) is the one that equates the two sequences with the least cost or the largest similarity score, depending on how operations on sequence elements are weighted. The basic operations are matches, substitutions (or mismatches), and insertions and deletions. In the top of Table 1 the [al] characters may be thought of as being deleted from [although] or inserted in [though]. Insertions and deletions are thus always used in pairs and are called *indels* in some texts. An insertion in one sequence creates a gap in the other sequence and optimal alignments will usually show a pattern of matched, mismatched, and gapped elements. Matches will tend to dominate the alignment of similar sequences. Alignment analysis produces two kinds of output: the graphic alignment of sequence characters and the similarity (or distance) measure between two sequences.

Alignment of two sequences is called pairwise alignment. The basic procedure can be generalized to any number of sequences using multiple alignment algorithms. However, the computational demands of multiple alignment increase as the average length of the sequences is raised to the power of the number of sequences to be analysed. Multiple alignments therefore rapidly become uncomputable and are carried out in practice using some variant of clustering procedures based on a matrix of difference indices derived from pairwise alignments. Algorithms that compare whole sequences are called global alignments. Others are designed to find similar subsequences and are called local alignments.

TRANSITION AND REORGANIZATION AT CMHC

In 1996, the Government of Canada redefined the responsibilities of Canada Mortgage and Housing Corporation in the areas of hous-

ing finance, assisted housing, export promotion, and research and information transfer. The transition to a new corporate organization to carry out the new responsibilities entailed a review of the systems and services of all sectors of CMHC and of the field-office structure. The ensuing reorganization involved retention of certain functions within the Corporation and the purchase of others from external suppliers.

As part of this exercise, the program evaluation division undertook a review of its products and procedures as they had been delivered up to 1996 to identify those that could be purchased more cost-effectively.

DEFINITION OF THE EVALUATION CYCLE

Up to 1996, the program evaluation cycle at CMHC consisted of an *evaluation framework* prepared at the time that a program was implemented or substantially changed, an *assessment report* prepared when a program was being considered for evaluation, *the evaluation study* itself, and a *publication and follow-up* process for the evaluation study report.

In preparation for the review and decision regarding contracting out some or all of the evaluation stages, the divisional staff had to define the steps that comprised each stage. Staff had all participated in some or all of the stages of the cycle, and individually wrote out the sequence of tasks in each stage. The first drafts identified a few dozen tasks, but by the end of the analysis we had identified nearly 250 separate tasks within the four stages. Each major product or stage contains a number of components, each consisting of a number of tasks. Table 2 shows the four evaluation products and their major components. The detailed tasks for the framework stage are shown in Appendix A. The author will supply a detailed list of tasks to interested readers on request.

A CLASSIFICATION OF EVALUATION TASKS

Once we had identified the sequence of tasks and time comprising evaluation frameworks, assessment reports, and studies, we considered how the large number of individual tasks might be classified into somewhat more general groups of activities. These could be used to report and analyse estimates of time consumption, among

other things. However, such an analysis of the internal structure of evaluation products turned out not to be necessary for subsequent stages of the reorganization. Only the initial thoughts of the team were ever recorded, and we produced no draft of an activity classification system.

The classification of the several hundred tasks from the task list shown in the appendix is my own development of our initial efforts and represents no consensus among any two evaluators as to what constitutes an adequate definition of evaluation tasks. I developed it mainly for the purpose of illustrating sequence analysis methods in the context of program evaluation. The system consists of 12 activity categories.

P - Planning

This function includes defining objectives, outlining options, and estimating the resource needs of future tasks.

Table 2
Components of Evaluation Products at CMHC

Evaluation Framework (41 tasks)
1. Program profile (5 tasks)
2. Framework (6)
3. Internal review (9)
4. Approval (15)
5. External distribution (6)

Assessment Report (76 tasks)
1. Update profile (7)
2. Identify evaluation issues (8)
3. Determine evaluation approach (5)
4. Costing and resources (4)
5. Options (6)
6. Internal review (9)
7. Draft AR approval (13)
8. External review (6)
9. Final AR approval (12)
10. External distribution (6)

Evaluation Study (83 tasks)
1. Partnerships (1)
2. Terms of reference (12)
3. RFP and contractor selection (5)
4. Contract negotiation (12)
5. Questionnaire design (7)
6. Develop sample (8)
7. Data collection (10)
8. Contract finalization and payout (6)
9. Research and analysis (13)
10. Internal review (9)

Evaluation Reporting (45 tasks)
1. Draft study approval (13)
2. External review (6)
3. Final study approval (12)
4. Communications plan (4)
5. External distribution (6)
6. Follow up (4)

R - Collecting and Reading Published Sources

> These tasks include the initial orientation of the evaluation team or single evaluator to the background of the program and to the associated academic, professional, and public literature related to a program. Tasks include time spent thinking about or discussing material that was read.

A - Analysis

> This activity involves breaking data down to examine differences or similarities among subgroups within the data, for example among types of program clients. It may consist of tabular or multivariate analysis of quantitative data. For qualitative data, it may involve simple review of interview notes or more formal content analysis. In the future, sequence analysis may become a useful method.

S - Synthesis

> This class of tasks involves constructing hypotheses about how programs work or how tests can be constructed to demonstrate or falsify theories, impressions, myths, or reports about the performance of programs. It includes developing conclusions and recommendations.

C - Oral Communication

> These activities comprise meetings with program managers and presentations of draft frameworks, assessments, studies, or components of evaluation products.

W - Written Communication

> These are all the writing tasks including drafting and revising reports and composing submissions to management and other correspondence.

E - Review and Evaluation

> Reviews consist of commentary by evaluators outside the evaluation team or by persons outside the evaluation division, including

policy and program staff within CMHC, senior management, or external interest groups.

M - Distribution

This group of tasks comprises the mailing or circulation of drafts, results or final evaluation products to affected offices within CMHC, to other federal government agencies, and to external interest groups.

V- Decisions and Approvals

These are milestones in the sense that some group, usually CMHC management, authorizes the evaluation team to proceed to a next task. They rarely involve any significant use of evaluation resources or time but are significant events in the evaluation process.

H - Administration

This involves the major tasks in the administration of contracts with consultants and suppliers and includes contract preparation, review, and payment.

D - Data Collection

These are the primary data collection tasks of an evaluation, as opposed to literature reviews. They include surveys, interviews, and focus groups. These tasks had usually been contracted out.

T - Translation

Translation is also a milestone that does not consume evaluation resources.

SIMILARITY OF STAGES OF THE EVALUATION CYCLE

To begin the analysis, we represent the tasks of the components of the four evaluation stages using the classification and write them as character strings (see below). The *greater than* sign indicates a label line in the data file. The label line is followed by task sequences. Illustrating the steps of the process as characters shows the simi-

larity of certain sections very clearly. I have selected some by eye which are shown bold and underlined. An analyst has no way of judging if these portray a critical feature of the processes or if they are more similar than other similar subsequences. Several contain repeats of subsequences of varying lengths but again it is very difficult to assess their importance.

```
> Framework - 5 components, 41 tasks
rcsswrcsswpeweewevtmweewevwemwecvwvwewvmm
> Assessment - 11 components, 78 tasks
rcasccrcrswewewrcsswrcpwpwewsceweewevtmweewevwemwecv
wewvmewtmweewevwemvwvwewvmm
> Evaluation study - 10 components, 84 tasks
cwwwwapaeewvmammavweeevmvmchvvsweewmamdaamehvvdaaaadmehvv
mmahvvaaaaaaaaaaswsssweweewevtm
> Follow up - 7 components, 45 tasks
weewevwemwecvwewvmewtmweewevwemvwvwevvwewvmmmeas
```

This is where the application of sequencing algorithms can help. The analysis was done using the ClustalG software package which is a rewrite of the well-known ClustalX package (Thompson, Gibson, Plewniak, Jeanmougin, & Higgins, 1997). ClustalG removes the 26-letter constraint on the alphabet used to represent amino acids that is usual in software written for applications in molecular biology. Events may now be coded in words of up to six letters allowing a reasonably rich representation of event categories, although this capacity is not used in this example. ClustalG also eliminates the explicitly biochemical features of the software and also offers improved mechanisms for transferring diagnostic data to standard statistical analysis software. An application package including the source code, documentation, and ClustalG itself is freely available to researchers at the web site of the International Association for Time Use Research, <www.stmarys.ca/partners/iatur/clustalG>.

The global alignment of the tasks of the four evaluation products is illustrated below, printed in two blocks. The block width is 60 activities, so the full alignment needs two blocks to be printed. It shows the best possible arrangement of a set of sequences that can be achieved by opening gaps of varying lengths in one sequence to create matches with characters in other sequences. It illustrates the arrangement of tasks based on the pairwise similarity scores calculated from the weights used to reward matches and penalize mismatches in this particular trial. In this example matches scored ten,

mismatches zero, and gaps were given a penalty of one. The alignment clearly highlights similar segments of the framework, assessment, and reporting stages comprised of writing (w), reviews (e), approvals (v), and distribution that occur roughly from positions 40 to 80. The variability of lengths of the input sequences forces the algorithm to create gaps at the ends of the alignment that have no substantive importance. When sequence lengths vary substantially, it is probable that local rather than global comparisons are appropriate.

```
Frame   -------------------RCSSWRCSSWPEWEEWEVTMWEEWEVWEMWECVWV-WEWV
Asses   RCASCCRCRSWEWEWRCSSWRCPWPWEWSC-EWEEWEVTMWEEWEVWEMWECV---WEWV
Follow  ------------------------------------WEEWEVWEMWECV---WEWV
Evalu   ----------------------CWWWWAPAEEWVMAMMAVWEEEVMVMCHVVSWEEW


Frame   MM-----------------------------------------------
Asses   MEWTMWEEWEVWEMVWVWEWVMM---------------------------
Follow  MEWTMWEEWEVWEMVWVWEVVWEWVMMMEAS-------------------
Evalu   MAMDAAMEHVVDAAAADMEHVVMMAHVVAAAAAAAAASWSSWEWEEWEVTM
```

Table 3 shows the percent identity of the pairwise alignments of the four evaluation stages. The matrix shows that the framework is very similar to the assessment and that the assessment is somewhat similar to the reporting stage. The evaluation stage is unique with respect to the other three. This is due to the data collection and analytical activity of the research phase.

One may ask why the cycle as it existed contained so much redundancy? Two factors are at play. Firstly, the framework and the assessment covered much of the same ground. The frameworks were usually written when new programs were launched or when old programs were amended. The assessment usually occurred several years later at the time a program was being considered for evaluation. In

Table 3
Pairwise Similarity Scores for Four Evaluation Products

Framework	100	78	29	56
Assessment		100	20	81
Evaluation			100	33
Reporting				100

Source: ClustalG 1.0 output

the intervening time, the environment would likely have changed and new issues may have arisen, but essentially the same procedures would be followed.

Secondly, the internal review and approval processes for external distribution had become standardized and were common to the framework, assessment, and reporting stages.

CURRENT STATUS OF THE EVALUATION PROCESS

The evaluation function at CMHC was amalgamated with the internal audit function in 1997, and the combined Audit and Evaluation Service was reorganized early in 1998. With respect to the program evaluation side of AES, evaluation studies are expected to be contracted out in the future. Frameworks will be drafted in-house, but the structure of the process has not been completely determined. In particular, it is not clear how the preparation and maintenance of program frameworks will relate to the production and use of assessments. To the extent that evaluations have recently tended to have been undertaken fairly soon after programs are launched, it is possible that the use of assessments will decline. This still leaves the reporting responsibility with CMHC staff. AES has not finalized the structure of the reporting and approval components of frameworks or of evaluation studies. The definition of the steps previously followed and the identification of repetitive phases will be useful in determining how the process may be effectively shortened and redesigned.

CONCLUSION

This article faced a dilemma at its inception. One cannot write about applications of an unfamiliar methodology in a discipline, such as program evaluation, when there really aren't any, but there will never be any applications of new methods until practitioners have some familiarity with them. Thus a rather strained example was pressed into service to introduce the topic. While a number of social scientists are becoming interested in problems that involve the complete sequences of events as opposed to examination of factors influencing particular events, sequence alignment is comparatively new and is all but unknown outside of a small number of scientific disciplines.

Initially, the first applications of sequence alignment methods in social sciences that come to mind are taxonomic. Such applications allow us to examine a number of questions relating to event sequences:

- How similar are processes of two subsamples of a population?
- What is the relative similarity of three or more subsamples?
- Has a process changed between two observations?

When multiple alignments are carried out on many sequences, graphic patterns of event matches and gaps emerge that are generated both by the sequences that are input and by the analyst's selection of mismatching and gapping penalties. One of the tasks that will be necessary in the development of useful social science applications will be to establish procedures to guide the weighting of similar events. The graphic output of large sequence analyses helps researchers see where behaviour is generally the same and where it is generally different. How this contributes to social research and to the evaluation of public spending programs remains to be seen.

The process of constructing a large alignment may use any one of a large variety of cluster analysis algorithms. This in turn creates a classification of individuals or observations based on the similarity of event sequences. In other words, the method generates a nominal variable that identifies classes of processes or behavioural groups. Applications of such variables are somewhat easier to foresee.

A variable that classifies sample observations can be used as input to certain multivariate methods, say regression, as either a dependent or independent variable. Group membership can be used as a dependent variable in a logit regression in which the independent variables are the characteristics that are thought to influence group membership. As an independent variable, group membership can be used to explain some other condition. To return to the example in the introduction, one might use the style of doing homework as a dummy variable in a regression explaining performance in school.

In program evaluation, research on event sequences would seem to be relevant to both formative and summative studies. In formative evaluation, evaluators are interested in program implementation and may concentrate on administrative processes. One might apply alignment methods to the sequence of events during the implemen-

tation of a program to examine how much a new procedure differs from an old procedure or to examine consistency of implementation among field offices.

In summative evaluation, alignment methods are more likely to be applied to the substance of social processes rather than to the evaluation tasks themselves. In a recent evaluation conducted by CMHC, an issue related to the role of on-the-job training of young workers. Had resources been available for more extensive research, it would have been possible to gather information on the daily activities of workers, including formal and informal training sessions. If recognizable patterns of training or supervision were related to particularly successful or unsuccessful working internships, this would have provided evidence of the effectiveness of internships under different training regimes. Currently, research at CMHC and elsewhere is examining the patterns of transitions of households from location to location. These housing trajectories are being analyzed using alignment methods to group respondents into typical trajectory patterns. Group membership is then used to identify migration patterns.

As researchers become more familiar with considering the similarity and variability of social processes as wholes, deeper questions about the origins of patterns, of similarity, and of direction of changes may be investigated.

ACKNOWLEDGMENTS

REFERENCES

Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology, 21*, 93–113.

Needleman, S., & Wunsch, C. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology, 48*, 443–453.

Sankoff, D., & Kruskal, J. (1983). *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison*. Reading, PA: Addison-Wesley.

Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F., & Higgins, D. (1997). The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research, 25*(24), 4876–4882.

Waterman, M. (1995). *Introduction to computational biology*. New York: Chapman Hall.

Wilson, C. (1998a). Activity pattern analysis by means of sequence alignment methods. *Environment and Planning A, 30*(6), 1017–1038.

Wilson, C. (1998b). Analysis of travel behaviour using sequence alignment methods. *Transportation Research Record, 1645*, 52–59.

Appendix A
Detailed Evaluation Framework Tasks

Task	Code	Task	Code
Program Profile (5 tasks)		*Framework Approval (15 tasks)*	
Collect and read documentation on program design	r	Draft submission to management	w
Consult program managers and others re: program design	c	Director review of submission (1)	e
Develop understanding of program activities, outputs, and impacts	s	Review of submission by program sectors, Finance and Legal	e
Develop understanding of program logic	s	Revise submission	w
Write up program profile and logic chart	w	Director review of submission (2)	e
		Review by Vice-president, Evaluation	v
Framework (6 tasks)		Revise submission	w
Collect and read literature re: program rationale/impacts/alternative	r	Director review of submission (3)	e
Consult program managers and others re: issues	c	Submission to MC (through Boards and Committees office)	m
Identify evaluation issues	s	Prepare supporting documents (summary, briefing notes)	w
Identify tentative evaluation indicators	s	Director review of supporting documents	e
Describe information and data requirements	w	Presentation to MC	e
Tentative evaluation plan (timing and general resource requirements)	p	Receipt of RD	v
		Report to Board of Directors	w
Internal Review (9 tasks)		Presentation to Board of Directors	v
Review by PED Staff and Director	e		
Revisions	w	*External Distribution (6 tasks)*	
Review by Director	e	Draft covering letters	w
Review by program managers and field staff	e	Review letters	e
Revisions	w	Revise letters	w
Final review and approval by Director	e	Sign letters	v
Final review and approval by VP Evaluation	v	Send copies of framework to central agencies, vested interest groups, academic community	m
Translation	t	Distribution of framework through CHIC	m
Printing	m		