Improving audit trail in the use of qualitative data in developing conceptual models

Oluwasegun Aluko

Abstract— **Developing a system dynamics (SD) model usually involves extracting the mental model of stakeholders related to the problem being tackled. A popular method for undertaking this is Group Model Building (GMB) method. However, it is possible that stakeholders cannot be gathered together in a GMB session or some of the stakeholders cannot be available for multiple sessions for economic reasons. In this paper, a method for exploring the creation of causal loop diagrams (CLDs) from qualitative data is described. It is shown that it is possible to develop a CLD with robust audit trail from qualitative interview data. This method helps to further build confidence in the resulting model.**

Index Terms— **qualitative data, causal loop diagram, causal network, system dynamics.**

I. Introduction

System dynamics (SD) models can be conceptual models or formal models. Conceptual models help to address modelling processes such as problem articulation, boundary selection, and variables identification [1]. An example is a causal loop diagram (CLD). CLDs are network diagrams. They also show causal relationships between variables in a manner that feedback loops and time delay characteristics of the network can be identified. Formal models, on the other hand, are quantitative models that can be used to test hypotheses and proposed policies. In formal models, behavior relationships are more explicitly specified and the numerical values of parameters and initial conditions are carefully estimated. Conceptual models (CLDs to be specific), however, form the subject of this paper.

A large number of CLDs are still developed without recourse to any standard method of formulating them. Some of the methods in use are based on the modeler's best judgment. This is partly due to the diversity in the nature and sources of data used for SD models. As noted by [2] there are three basic knowledge sources for SD models: mental, written, and numerical data. Numerical data which is most easy to deal with (in analysis and interpretation) usually offers the least amount of contextual information for SD model development. Mental and written sources are however more informative but more difficult in analysis and interpretation. This difficulty with the analysis and interpretation (and extraction when required) not only creates an avenue for variety of methods but also for less standardized ones. Moreover, while there are some standardized methods generally accepted in the SD community, these methods do not always respond to all modelling needs. In this paper, a case for stakeholders whose characteristics do not fit well for some standardized methods for the development of a CLDs is presented. First, some of the methods currently in place for developing CLDs are discussed. This is followed by the introduction of the method this paper seeks to present.

Manuscript received March 18, 2015.

Oluwasegun Aluko, Civil Engineering Department, Ekiti State University, Ado Ekiti, Nigeria, +2349028629730,

II. METHODS FOR DEVELOPING CLDS

The need to standardize the method for developing SD models is well acknowledged. There have therefore been various suggestions on recommended scripts for SD model development. Reference [3] established one of the first recognized methods for developing an SD model and named it Group Model Building (GMB). The GMB is a method that requires the participating stakeholders to be physically present during the modelling process and build the model together with the modeler(s). This method has been shown to be good for organizational studies where stakeholders have an existing relationship with one another and/ or where stakeholders share common interest [4], [5], and [6]. It might however, be difficult where the social status of stakeholders is widely varied or where they have varied interests.

In the same vein, [1] introduced a similarly rigorous method which does not necessarily require the physical presence of participating stakeholders. It is similar to Delphi method of data collection but more rigorous and exerting in using stakeholders to both specify the relationship between variables and their quantities. The rigor and discipline required for this process mean that stakeholders must be really interested in the process to commit so much to it. This is however not always the case.

There are yet other effort addressing this methodological issue on the identification of system components and causal links for conceptual models. Reference [7] describe in detail a method for eliciting expert knowledge for conceptual model development. This method too, like the previous ones, rely on the participation of stakeholders to fully obtain the model.

Following a rather different approach, [8] suggest the adoption of data coding (in qualitative analysis) using computer aided qualitative data analysis software (CAQDAS) tools to develop a CLD. The method developed by [8] involves a coding system. The process could also be further developed for automation. It involves the analysis of qualitative data using qualitative analysis methods (such as tree analysis) and binary matrices to identify how words used in the data are linked to one another. This method however might be more appropriate for problem fields with more uniform technical terms to allow the analysis capture optimally important themes in the qualitative data. In addition, the adoption of word length as a censor in this method can lead to the loss of important concepts.

In this paper, a method is presented that deals with the analysis of written data. Such written data includes extracted information from the mental knowledge of stakeholders as well as other written information available and relevant to the problem to be tackled. The method is also demonstrated with an example case study.

III. METHOD OF DEVELOPING A CLD FROM WRITTEN DATA SOURCE

The proposed method is a six stage method. These six stages are sequentially described below.

A. *Obtain qualitative data:*

The method addresses the use of written qualitative data for developing a causal loop diagram. The first step therefore is to obtain available (or the required, as the case may be) qualitative data about the problem being addressed. Qualitative data comes from various sources including written sources such as documents, and oral sources such as interviews. Once this qualitative data is obtained, it is necessary to make it available in a written form for further analysis.

B. *Code data:*

Once the qualitative data is available in a written form it can then be analysed using qualitative methods. One way of undertaking the analysis is by coding the data to identify the themes (usually called codes) that are represented in the data. Coding is defined as the use of a word or short phrase to describe the basic topic of a passage of qualitative data [9]. By adopting what [10] call *causation coding*, it is possible to extract attribution codes that are suitable for further development to causal networks[1] and CLDs. Causation codes are codes that indicate attribution and show that a cause leads to an effect. It is obtained by reading through the data time and again, labelling chunks of data each time. The use of CAQDAS such as NVivo makes this process easier and provides an audit trail of what is being done. Reference [11] suggests several repetition of this process until there is no new emerging theme or code.

C. *Generate causal patterns:*

The completion of coding process only prepared the analysis process for "sense-making" exercise. Both [9] and [10] suggest the use of graphical representations for the outcome of a coding process to support this sense-making exercise. With the adoption of causation coding, an appropriate graphical representation is the use of what [10] calls causal networks. Causal networks are graphical illustrations of cause and effect as they emerge from the data. They are drawn with the use of arrows and labels. Arrows link the labels to one another and indicate how one thing leads to (or is related to) the other.

D. *Generate network narrative:*

It is usual to provide a worded description of all the links present in the causal network. This description helps to provide a story-like account of how and, often, why one cause leads/ relates to its effect. This description is called a "narrative". A major advantage of the narrative is that it provides a succinct account of the data in a manner that everything important (to the problem topic) in the data is included. But it does this in few words when compared to the original qualitative data. In a word, a narrative provides a complete description of a system's causality relationship as found in the data without including illustrations, examples, and other needless information that make the original data bulky.

E. *Summarize narratives to generate dynamic hypothesis:*

The phase of the analysis that precedes development of a CLD is the generation of summary statements from the narrative. This summary is different from the narrative in that while the narrative is a story-like description of all the links identified in the data, the summary is a list of bullet points/ statements of the content of the story. The summary identifies processes/ events in the story and why they happen the way they do. More specifically, for the purpose of the development of a CLD, these summary statements describe processes and their feedback loops in a manner that they form a dynamic hypothesis for the problem structure in the system being analysed.

While the process described thus far is a typical qualitative analysis method, the possibility at this stage to obtain summaries that can form dynamic hypothesis makes the method suitable for adoption in developing conceptual models such as a CLD.

F. *Sketch the CLD:*

Once the statements of dynamic hypothesis have been generated, it is possible to develop them into a graphical representation in the form of a CLD. The emerging CLD is different from causal network in many sense. Particularly, the statements of dynamic hypothesis that result from the narrative previously generated reflect a dynamic sense which is not obvious in a causal network. More specifically, a CLD always shows features such as reinforcing loops, balancing loops, and time delays which cannot be represented in a causal network. In this way, a CLD eventually emerges from the qualitative analysis process.

IV. CASE STUDY EXAMPLE DEMONSTRATING THIS METHOD

An illustrative example of this process is presented below. Before describing this example, a brief background information is presented about the problem structure being treated.

The problem treated in this example is the safety challenge of commercial motorcycle operation in Nigeria. Commercial motorcycle service is the use of motorcycles for carrying passengers for a fare. It is a common mode of transport in Nigerian towns and cities. But motorcycle transport account for one in five road traffic accident victims in Nigeria and for as much as 35% fatality and commercial motorcyclist are usually blamed for this problem. A number of attempts have been made to combat the safety problem. Unfortunately, most of these attempts have not been successful. Worse still is the fact that the state is confused about how to tackle the problem. While there are various studies on the nature of this safety problem, there has not been any that review the problem from a systems perspective. This absence of systems review of commercial motorcycles' safety problem was the basis for this illustrative case study.

To conduct the systems analysis of the safety problem, it was necessary to consult with stakeholders in the operation of this transport. These stakeholders included the commercial motorcycle drivers, the road traffic police (including the

[1] This is described later

Federal Road Safety Corps – a special police unit dedicated to road safety), the Vehicle Inspection officers, staff of Accident and Emergency unit of hospitals, and academics and researchers working on transport. One important characteristic of these stakeholder groups is the lack of trust within them (which was expressed during data collection). This made it impossible to bring the groups together in a GMB. As a result, it was decided to conduct interviews separately for different stakeholder groups. However, a major constraint was that most respondents were not available for follow-up interviews particularly due to financial and time constraints. The rest of the process is described below.

A. Obtain qualitative data:

As earlier noted, a case is presented where respondents could not be available for a GMB session or a repeat consultation. What was done was to meet as many stakeholders as were available for semi-structured interviews. In all 25 respondents participated in 13 interview sessions as shown in the table below. Most of these interviews were audio recorded while others that could not be recorded were documented by hand-written notes.

Table 1: Stakeholder groups and participation in data collection

Group	Number of respondents	Number of interviews with group
Drivers	15	3
Enforcement agencies (3 agencies)	6	5
Vehicle Inspection Officers	1	2
Medical workers	1	1
Academic researchers	2	2

Following the completion of the interviews, the interview data was transcribed to ensure the entire data is in the same format and to enable analysis.

In addition to the interview data, other written documents such as newspapers, reports, and literature on commercial motorcycle safety were included for analysis.

B. Code data:

The data analysis process started with coding. Once the data was completely prepared in worded form, the coding process was initiated. The coding methods adopted was based on attribution and is called causation coding. The process involved reading through the transcript several times and making a note of all meanings that could be made out of the data. Usually, it is done by using a word or short phrase to describe the meaning of a chunk of data. During this process, several codes (themes) emerged. These codes were reviewed by reading through to ensure there were no duplications of codes or repetition. In addition, because the codes that emerged were really many, there was the need to group related codes together. This grouping together is also known as clustering. A sample tabular representation of clustering process is shown in table 2. In the table three causation codes are listed for each cluster title. There are however more than this in the actual study.

Table 2: Sample clustering

Cluster title	Cause	Effect
Training	Available spare time	Willingness to give time for training
	Availability of training opportunities	Participation in training
	Ignorance (of driving rule)	Risky and dangerous driving
Enforcement and regulation	Deterrence	Violation
	Method of arrest	Dodging arrest
	Enforcement coverage	Probability of arrest
Violations	Loss from accident	Violations
	Violations	Enforcement coverage
	Violations	Accident

C. Generate causal patterns:

Clustering was done for ease of evaluation in the preceding stage. This clustering makes it easy to develop graphical causal patterns. In the case study, the clusters obtained were developed into small causal graphs. An example of this is shown below.

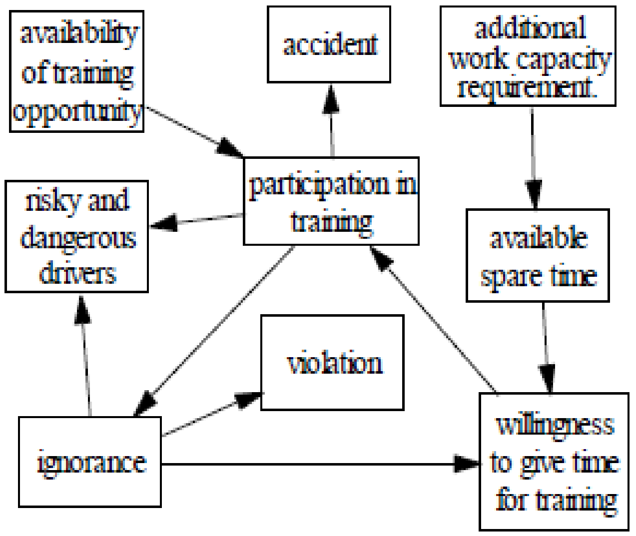

Fig. 1: Causal network for a cluster title [12]

In this case study, five of these clusters of causal relationships were obtained. Thus, once these cluster representations were completed, it became necessary to combine them into a single representation to obtain a whole, unified picture of the system. The result of the combined cluster gave rise to the figure below.

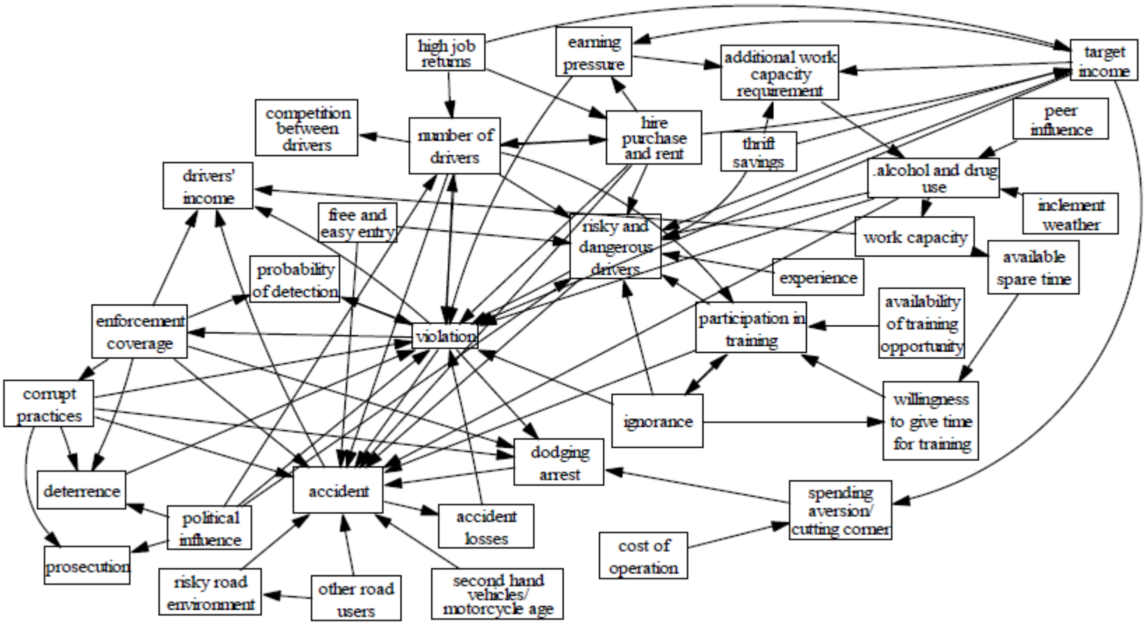

Fig. 2: Combined causal network for all clusters [12]

However, figure 2 has a number of redundancies which were removed. The final causal network that emerged is figure 3. A table showing the redundancies identified was prepared and included in the analysis (but not shown here).

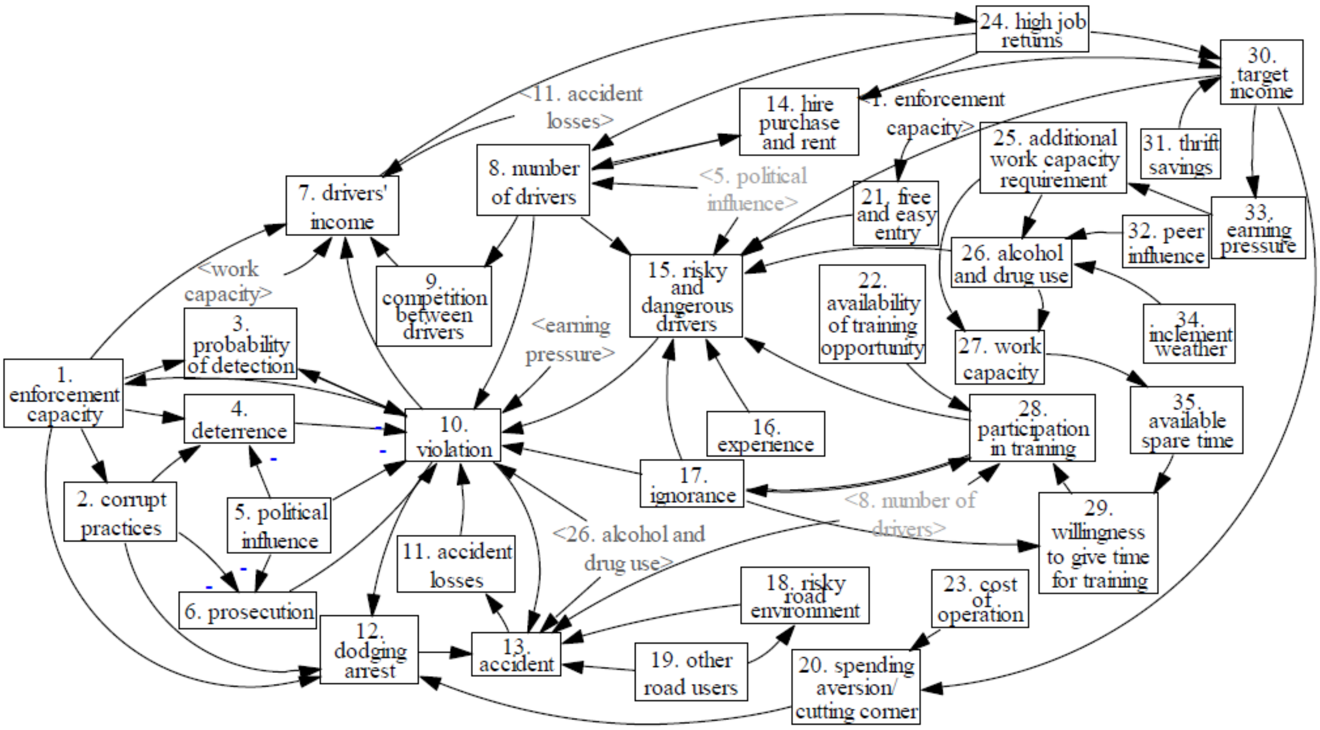

Fig. 3: Revised combined causal network [12]

However, figure 2 in has a number of redundancies which were removed. The final causal network that emerged is figure 3. A table showing the redundancies identified was prepared and included in the analysis (but not shown here).

D. Generate network narrative:

The development of the combined causal network paved way for the generation of a narrative that describes all the links in the causal network. There are no rules about the starting and end points of the narrative. It is however important that all the links and codes in the causal network are included in the description. A short portion of the narrative is presented below for illustration. This illustration shows how a narrative is written.

"Enforcement capacity (1) which represents the combined ability of traffic enforcement agencies in the study location affected several other items. It was found that whenever there were more officers on patrol, fewer drivers worked due to increased probability of detection (3) of a violation by enforcement officers. This was more so as more

monitoring by officers meant more spending on fines and bribery for the drivers. Thus, more violations (10) led to more enforcement capacity (1) which led to reduced drivers' income (7). Notwithstanding, there were times an increase was noted in violations (10). This was because violations (10) offered some financial benefits too (increased drivers' income (7)). Whenever violations increased, more officers were drafted to increase enforcement capacity (1) and match the problem. This obviously would result in increase in the probability of detection (4) and violation would go down. It was also noted that some drivers were naturally deterred (deterrence (4)) from violating laws due to increased likelihood of being arrested. In this way, increasing enforcement capacity (1) could reduce the total number of violation (10)." [12]

Writing out the content of the causal network in this manner makes the network more comprehensive and easier for analysis when compared to the causal network as a graph. This is more so as the generation of summary points from the causal network is essential for the emergence of dynamic hypothesis required for building a CLD.

E. Summarize narratives to generate dynamic hypothesis:

Following from the generation of a narrative, a summary of the narrative was made. This summary identifies the important processes that are represented in the narrative. For example, the following can be deduced from the content of the narrative excerpt presented above.

Officers could enforce laws by detecting and arresting violators. This way they deter drivers from engaging in violations and reduce the total number of violations. In a sense, if violations increased, officers increased and vice versa.

F. Sketch the CLD:

This is the final stage of the method discussed in this paper. It involves converting the summary points obtained in the previous stage into a CLD. It is important to emphasise that these summary points are developed to form a dynamic hypothesis which can easily be converted into a CLD. For example, the summary point shown above is an example of a balancing loop. It can therefore be drawn out to form a CLD. This is illustrated below.

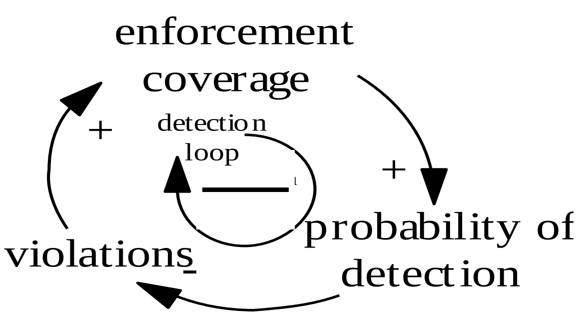

Fig. 4: CLD outcome of summary statement [12]

V. CONCLUSIONS

In this paper, a case has been presented for the development of a causal loop diagram from qualitative data that is sourced from both mental and written knowledge sources. It has been shown that such data does not fit the use of Group Model Building method or some other standardized methods for developing causal loop diagrams. While qualitative data coding has been previously adopted in building causal loop diagrams, the method presented in this paper is shown to minimize the possibility of the loss of important concepts in the analysis while at the same time providing a robust audit trail to support analysis outcome. In addition, the paper has shown how a typical qualitative data analysis method can be adopted for building a causal loop diagram in a systematic manner.

The process involved in this method has been described and illustrated. It can be useful to compare the outcome of this method with other methods to test for how well it covers important concepts in a typical problem context. This will be a future research direction for the authors.

VI. REFERENCES

[1] Ford, D. and Sterman, J. 1998. Expert Knowledge Elicitation for Improving Mental and Formal Models. *System Dynamics Review*. 14(4), pp.309 - 340

[2] Forrester, J.W. 1992. Policies, decisions and information sources for modeling. *European Journal of Operational Research*. 59, pp42-63

[3] Andersen, D.F. and Richardson, G.P. 1997. Scripts for group model building. *System Dynamics Review.* 13(2), pp.107-129.

[4] Irene, N., Badinelli, R., Polese, F., Nauta, P.D., Löbler, H., and Halliday, S. 2012. S-D logic research directions and opportunities: The perspective of systems, complexity and engineering. *Marketing Theory.* 12(2), pp.213-217

[5] Laurenti, R., Lazarevic, D., Poulikidou, S., Montrucchio, V., Bistagnino, L., and Frostell, B. 2014. Group Model Building to identify potential sources of environmental impact outside the scope of LCA studies. *Journal of Cleaner Production* 72, pp.96 - 109

[6] Shin, D. and Konrad, A.M. 2014. Causality Between High-Performance Work Systems and Organizational Performance. *Journal of Management.*

[7] Morecroft J.D.W. and van den Heijden, J.M. 1994. Modelling the oil producers: Capturing oil industry knowledge in a behavioural simulation model. In *Modeling for Learning Organisations*, ed J.Morecroft and J. Sterman. Portland, OR: Productivity Press

[8] Yearworth, M. and White, L. 2013. The uses of qualitative data in multimethodology: Developing causal loop diagrams during the coding process. *European Journal of Operational Research.* 231(1), pp.151-161.

[9] Saldana, J. 2013. *The coding manual for qualitative researchers*. London: Sage.

[10] Miles, M.B., Huberman, A. M, and Saldana, J. 2014. *Qualitative data Analysis: A Methods Sourcebook.* California: SAGE Publications, Inc.

[11] Burnard, P. 1991. A method of analysing interview transcripts in qualitative research. *Nurse Education Today.* 11(6), pp.461-466.

[12] Aluko, O.O. 2014. Understanding the safety performance of commercial motorcycles in urban transport using a system dynamics approach based on qualitative data. Unpublished PhD thesis, University of Leeds, UK