Quality Technology & Quantitative Management
Vol. 9, No. 2, pp. 137-152, 2012

QTQM

Ishikawa Cause and Effect Diagrams Using Capture Recapture Techniques

R. Ufuk Bilsel[1] and Dennis K.J. Lin[2,3]

[1]The Boston Consulting Group, Kanyon Ofis kat 13 Buyukdere Cad. Istanbul, Turkey
[2]Department of Statistics, Pennsylvania State University, PA 16802-3603, USA
[3]School of Statistics, Renmin University of China, Beijing, China
(*Received* February 2010, *accepted* February 2011)

Abstract: When a problem occurs in a system, its causes should be identified for the problem to be fixed. Ishikawa Cause and Effect (CE) diagrams are popular tools to investigate and identify numerous different causes of a problem. A CE diagram can be used as a guideline to allocate resources and make necessary investments to fix the problem. Although important decisions are based on CE diagrams, there is a scarcity of analytical methodology that supports the construction of these diagrams. We propose a methodology based on capture-recapture analysis to analytically estimate the causes of a problem and build CE diagrams. An estimate of the number of causes can be used to determine whether the CE study should be terminated or additional iterations are required. It is shown that integration of Capture-Recapture analysis concepts into CE diagrams enables the users to evaluate the progress of CE sessions.

Keywords: Capture-Recapture analysis, CE diagram construction methodology, Ishikawa CE diagrams.

1. Introduction

Problems and disruptions are common concerns in all systems. Effort should be spent towards fixing problems once they occur to ensure sustainability of a system. In order to fix a problem, one needs to know about causes that lead to it. An *Ishikawa Cause and Effect* (CE) *Diagram* (also called fishbone or simply cause and effect diagram) is a simple but effective tool that is used to identify different causes of a problem. A CE diagram consists of a main "bone" to which main causes of the problem are connected. Each main cause may have several sub causes that lead to the main cause. Similarly each sub cause may have third level causes leading to them and so on. That structure is presented in a CE diagram to provide system analysts and managers valuable information about the roots of the problem and where to start at for fixing it. Therefore, the way an organization is going to spend its resources may very well depend on a CE diagram.

In practice, CE diagrams are the end products of an organized or ad-hoc brainstorming session. This paper proposes a complete analytical framework to follow in building CE diagrams. The method we develop makes use of the capture recapture analysis to gather expert knowledge and evaluate the completeness of the diagram. Any improvement made in the construction phase of CE diagrams is potentially a positive step towards optimizing the way an organization allocates its resources in solving problems. Obviously, an improvement in resource allocation would reduce costs and increase profitability. Towards improving CE diagrams, we will introduce some concepts from the fields of Capture-Recapture (CR) studies and propose a new methodology to be followed in applications.

To our knowledge, this paper is the first to propose a complete analytical approach to CE

diagram construction. Two related work approach the CE problem in different ways: [22] proposes several mathematical programming formulations to improve the cause and effect analysis. This work is more general and considers investment decisions to minimize for instance the time to find the true cause or to maximize the chance of finding the true cause. The probability of a cause being the true cause is modeled by a known discrete probability distribution. Reference [16] proposes using CR tools to improve CE diagrams. The author provides a step by step methodology similar to ours but he does not discuss the theoretical motivation of his methodology and does not present any application.

This paper is organized as follows. The next section outlines the basics of CE diagrams, reviews the recent relevant literature. The methodology to improve Ishikawa CE diagrams is then proposed. The next section lays out the details of the proposed methodology and discusses some statistical details followed by an implementation of the new methodology. The following section briefly overviews Capture-Recapture models. The penultimate section concludes the paper and comments on future work to be conducted.

2. State of the Art in CE Diagrams

2.1. Basics of CE Diagrams

CE diagrams were introduced by Kaoru Ishikawa in early 1940s. He first used these diagrams to help increase productivity in Kawasaki Steel Works in Japan [1]. A CE diagram is used to identify many possible causes of an effect or a problem and can be used to summarize a brainstorming session. It immediately sorts ideas into useful categories and offers an easy-to-understand display of the problem at hand [3]. A CE diagram consist of a main horizontal line, called bones, from which emanates a number of diagonal lines. These are used to list the general causes that may lead to the studied problem. For convenience, the problem description is usually written at one end of the main horizontal line. Several smaller bones can stem from each bone to further detail the secondary causes that may affect a main cause. The detail level depends on the user's perception of the problem. A generic CE diagram is given in Figure 1. CE diagrams are considered as one of the seven basic tools of quality management (with checklists, Pareto charts, histograms, scatter plots, control charts and flowcharts).

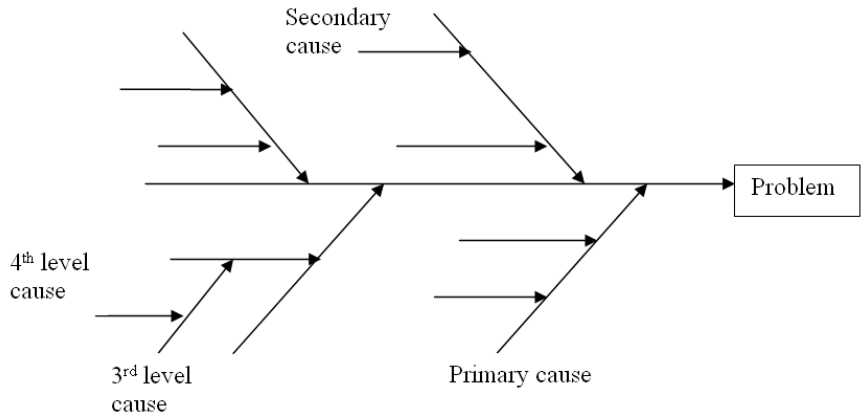

Figure 1. A generic CE diagram.

2.2. Past Work and Current Trends in CE Diagrams

Since its introduction, CE diagrams found a large domain of application for solving problems in many different fields. As previously mentioned, the first application of CE diagrams was in steel the industry in Japan. Recent research on CE diagrams including applications can be summarized as follows.

Reference [10] uses CE diagrams to determine dynamic error characteristics of touch trigger probes used to coordinate measuring machines. They report that the study was successful and several suggestions for improving available tools are made. An application of CE diagrams to identify risk factors affecting the design phase of automotive electronics is presented in [21]. Reference [17] presents an application of CE diagrams to evaluate thermodynamic data from UV-Vis absorption in spectroscopic analysis. There, contributors to measurement uncertainty are identified using a CE diagram. In reference [14] CE diagrams are utilized to discover causes of delamination defects in stacked die applications. Reference [18] studies a reliability problem for electronic components. The author argues that obtaining reliability data for electronic components is both costly and time consuming; therefore, alternative techniques such as accelerated life testing should be used. To correct the faults on parts, the author proposes correcting factors and links these correcting factors to problems via a CE diagram. Reference [12] uses a CE diagram in a chemical experiment to determine causes affecting the protonation constants in a chemical reaction. Reference [23] reports an application of CE diagrams in the aerospace industry. CE diagrams are used there to identify problems with satellite link performance such as loss of lock or high bit error rate. Apart from engineering and science applications, Reference [9] presents a framework of using CE diagrams to diagnose service processes in service industry.

As discussed above, most of the recent literature focuses on the application of CE diagrams rather than a methodology for constructing them. Reference [20] presents a new tool called the *process matrix* to deal with several limitations of CE diagrams. Their proposed tool borrows its structure from the Quality Function Deployment matrix and is used as a representation that unifies multiple CE diagrams into one matrix. More recently, [16] proposes a framework to link CE diagrams to capture recapture analysis. His work discusses a methodology but does not provide any verification.

2.3. Pros and Cons of CE Diagrams and Demanded Research

CE diagrams provide the following usage advantages [1]:

- Help to identify all possible causes of a problem,
- Help to determine these causes in a structured way,
- Use domain knowledge of all participants,
- Help focusing on the causes of the problem and omit complaints and irrelevant discussions,
- Provide an easy to understand graphical format,
- Help leveling the knowledge of all participants by allowing everyone to share their expertise,
- Identify areas where there is lack of data and future study is needed.

However, despite their strengths, CE diagrams have some weaknesses as well. Several references, including [1] and [20], report some of these weaknesses as the following:

- When analyzing multiple problems, an individual CE diagram has to be drawn for each problem, which makes the analysis time consuming. Also, it is possible to miss interrelations among different problems and causes;
- It is not possible to differentiate the strength of the cause and effect relations;
- CE diagrams are difficult and inefficient when used in an electronic form, and can become disorderly.

Reference [20] addresses the above weaknesses by proposing a new tool called the process matrix. Some relevant work, including [1] and [23], suggest using brainstorming sessions or organized meetings. Otherwise, the literature on the methods to construct a CE diagram is scarce.

3. Improved Cause and Effect (CE) Diagrams

A CE diagram is a powerful and intuitive tool of systems diagnosis to determine causes of problems. This section addresses the construction of CE diagrams through an analytical methodology that can be straightforwardly applied to real life cases.

In a CE diagram, a main problem and causes contributing to the problem are sketched. Identification of the problem might be achieved by error analysis or expert studies. Identification of the causes is potentially a more challenging undertaking. Methods such as Delphi or Nominal Group Technique sessions are advised for cause identification. However, these processes are more of organized brainstorming techniques rather than analytical methods. One analytical approach for CE diagram construction is to make use of the *capture-recapture* (CR) models that are widely applied in wildlife sciences. CR models are used to estimate general properties (such as size) of wildlife populations based on a sequence of experiments called capture occasions. In most cases, only a portion the whole population is captured in each experiment. The results of these experiments are then extended to the whole population using different statistical methods discussed in the following sections (See [2] for a broader overview of CR models).

The analogy between CE diagrams and CR analysis can be established by noting that the objective in a CE diagram is to discover as many causes as possible. Hence, the causes are in fact "animals" of a traditional CR study. Trapping sessions (or capture occasions) are the experts meetings in a CE diagram study. Table 1 presents an analogy between CR analysis and CE diagrams in terms of concepts.

Table 1. CR analysis and CE diagrams analogy.

CR terminology	CE diagram counterpart
Animal	Cause
Capture occasion	CE session
Capture	Discovery of a cause

There is a multitude of CR models available for analyzing unknown populations and choosing the appropriate model may not be straightforward. Following a literature study on CR models, discrete time closed population CR models are deemed as suitable for CE analysis. More on these CR models and justifications of this selection is provided below.

Discrete-time closed population CR models are suitable for experiments where the studied population is sampled at discrete occasions and the size of the population does not

vary over time. In a CE study, this would imply a number of distinct CE sessions and a constant number of factors affecting the problem. Discrete-time closed population models are suitable when the system studied is shut down for analysis or when the system itself is not dynamic or might be assumed stationary over the period of study. In each session of a discrete-time closed population study animals that get captured are marked and then released to the population. The capture recapture history of the animals is recorded in a summary matrix (similar to the checklist presented in Table 2). In a CE study, causes identified in each session are recorded in the checklist. This process is analogous to marking captured animals in a CR experiment. In CR experiments, captured animals are released to the population after being marked. This process is quite similar to closing a CE session. However, if the same team of experts runs all CE sessions, then it is safe to assume that they could discover all the causes they did in previous sessions. This is similar to assuming that all animals marked in previous sessions are most likely going to be captured again in the later sessions[a]. If, on the other hand, different teams run each session, we can no longer assume that they will identify all the causes discovered in previous session. This case is more similar to a CR study. Moreover, in a CE study causes arguably have different probabilities of being discovered. Some causes are more obvious than others; any participant might be able to identify them. Other causes, on the other hand, might require domain expertise and might be discovered by more experienced participants.

Finally, if the same panel of experts runs consecutive CE sessions, then causes identified in previous sessions can be copied over to the current sessions and causes identified for the first time in the current sessions should be added to the checklist. If the CE sessions are distributed among separate teams, then each team should record on a different checklist. Companies may prefer distributed sessions to leverage expertise of different groups of experts on the same problem. These checklists can be merged at the end of the CE study. The methodology presented here is flexible enough to be used in localized or distributed CE studies.

As discussed above, this paper proposes using closed population CR models, which assume no change in the number of individuals of the studied population during the study. That is, the underlying assumption with those models is that there is no birth, death or migration over the period of study. The CE counterpart of this assumption is that the number of unknown causes of a problem is fixed over time. A closed population model is called *heterogeneous* if individuals or group of individuals have different probabilities of being captured. The CE counterpart of this assumption is having causes (or groups of causes) with different probabilities of being discovered. *Heterogeneous closed population CR models* are suitable for improving CE diagrams.

It is important to note that some closed population CR models include "animal behavior" and "temporal effects" (e.g. age of the animal) into the model. Indeed, the most general variant of discrete time closed population models is the M_{thb} model where capture probabilities can vary over time (represented by subscript t), different behavioral aspects of animals are considered (b) and animals (or groups of animals) may have different capture probabilities (h). Having different capture probabilities is analogous to having causes with different probabilities of being discovered. Temporal and or behavioral effects are harder to tie to CE studies, however those may be relevant if the actual cause is a human being; which

[a] Note that this case can actually be modeled with CR models where animals are assumed to be *trap happy*, that is they enjoy being captured. The analogy to CE studies would be the causes being known once being discovered.

may occur in human systems. In this paper an M_h discrete time closed population model is used to illustrate how CR models and CE practices can be bridged together.

The methodology proposed here relies on one intuitive but key observation: more causes can be discovered as more CE sessions are held. To implement this observation, let N be the total number of unknown causes to the studied problem, i be the session index and n_i the number of causes discovered during the i-th session and p be the probability of discovering a cause of the problem. We assume for the moment that p is the same for all causes. This assumption will be relaxed in the simulation study. With these settings, the probability of discovering n_i causes in a session, noted as $p^{(i)}$, follows a Binomial (n_i, p, N) distribution. Using this notation, the CE diagram construction process can be illustrated as in Figure 2. Note that Figure 2 displays the case where n_i causes are discovered or no cause is discovered. Intermediate cases, where the number of identified causes is between 0 and n_i are excluded from the figure to keep the presentation as lean as possible.

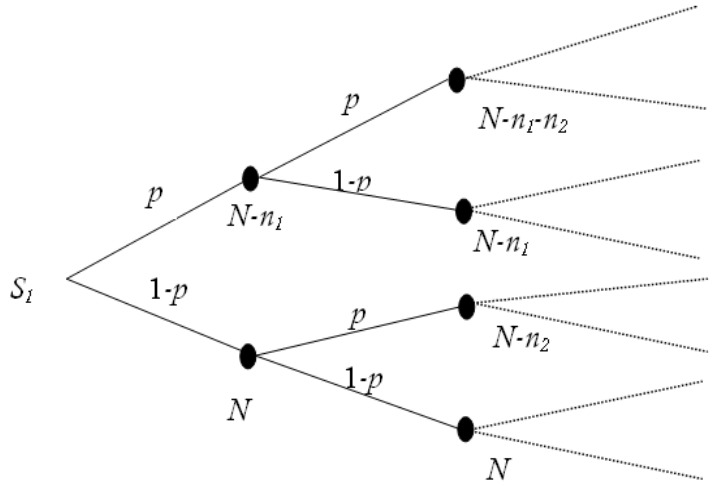

Figure 2. Progress of a sequence of CE sessions and their outcomes.

Each node in Figure 2 represents a CE session. At the beginning of the first session S_1 there are N undiscovered causes. Under the assumption of equal probability of discovery for all causes, S_1 can conclude with one of two outcomes: either n_1 causes will be discovered with probability $p^{(1)}$ hence there will remain $N - n_1$ undiscovered causes, or no cause will be discovered with probability $(1 - p^{(1)})$ and there will remain N undiscovered causes. Each outcome of S_1 can have two new outcomes in the second session and so on. Our observation can be stated in a theorem given in the appendix with its proof.

The assumption of uniform probability of discovery for all causes can be easily relaxed without violating the Theorem. Note that in the general case every cause might be discovered with a different probability and a proof would account for different causes and their related probabilities. The essence of the Theorem, on the other hand, would still hold.

The Theorem motivates conducting successive CE diagram sessions to discover more causes. Decision makers, on the other hand, need another tool to estimate the total number of causes to assess the effectiveness of the CE sessions and eventually terminate the study. CR

analysis provides methods to estimate the total number of causes given the results of several CE sessions. The following five-step methodology is proposed to apply CR analysis for improving CE diagram construction. The statistical properties of the methodology will be discussed in the next section.

- Step 0: Identify the problem to be analyzed and form an initial group of participants;
- Step 1: Run the first CE session, discover several causes and list them;
- Step 2: Update the participants if necessary, run additional CE sessions;
- Step 3: Run step 2 until a stopping criterion is verified;
- Step 4: Build up the final CE diagram based on the results obtained over the CE sessions.

Note that the methodology above requires a stopping criterion. In practice, the stopping criterion can be related to an available budget and the study can be finished when the budget is depleted. On the other hand, some parameters readily available in CR models can be used as stopping criteria as well. We develop a stopping criterion in the next section.

4. Details of the Step-by-Step CE Diagram Construction Methodology

Step 0. Identify the problem:

The first step in the CE analysis is the problem identification. In large organizations, management usually faces a myriad of problems to solve; therefore, a prioritization of these problems might be needed. Several of the seven Ishikawa quality tools, such as Pareto charts or histograms can be used to rank and prioritize the problems.

Step 1. First run of the CE diagram study:

Once the problem to focus on has been determined, a first group of experts conduct the first session of CE diagram building. Identified causes may be listed on a check sheet as shown in Table 2. The first column of the checklist presented in Table 2 is to be filled. Each discovered cause is assigned to a number and marked on the checklist. The checklist provides an easy way of bookkeeping of outcomes of the CE sessions which will be used later in the statistical analysis.

Steps 2 and 3. Additional CE sessions:

Assume that the number of causes does not change over time. This characteristic allows using closed population CR models. Let each cause i have an individual probability p_i of being identified. Then, the M_h CR model can be used to estimate the total number of unknown causes denoted by N. The challenge with the M_h model comes with the multitude of p_i probabilities and the necessity to attribute a distribution to these probabilities.

Table 2. CE sessions checklist.

	CE *Session*				
Cause	1	2	3	…	*S*
1					
2					
…					
N					

Reference [4] overcomes this difficulty by proposing a model where the capture frequencies replace the p_i probability values.

In steps 2 and 3, the analysts should execute several additional CE sessions to discover additional causes. The exact number of additional sessions depends on the expected number of total causes and can be determined using the stopping criterion proposed below. The number of causes is estimated using the following Jackknife formulas.

$$\hat{N}_{J1} = M_{k+1} + \left(\frac{k+1}{k}\right) f_1, \text{ and} \tag{1}$$

$$\hat{N}_{J2} = M_{k+1} + \left(\frac{2k-3}{k}\right) f_1 - \left(\frac{(k-2)^2}{k(k-1)}\right) f_2, \tag{2}$$

where k is the CE session index; M_{k+1} denotes the number of distinct causes discovered before the $k+1^{st}$ CE session, f_j is the number of causes captured j times during the entire CE study; $\hat{N}_{J1}$ is the first order jackknife estimator of the total number of unknown causes and $\hat{N}_{J2}$ is the second order jackknife estimator. Reference [4] presents approximate variance estimates for the population size estimators calculated using the jackknife method as well as higher level jackknife population size estimators. The jackknife method provides closed form estimators, but does not give a measure of the degree of heterogeneity of the problem causes.

An alternative way to estimate the total number of unknown causes relies on the *sample coverage approach* discussed in [6] and [13]. Sample coverage is the proportion of total individual capture probabilities to those of the captured animals and is proposed as a measure of completeness of a sample. The sample coverage method assumes that the heterogeneity is captured using the coefficient of variation of the discovery probabilities of causes-the $p_i's$. Our method proposes two estimators for the sample coverage,

$$\hat{C} = 1 - \frac{f_1}{\sum_{t=1}^{S} t f_t}, \text{ and} \tag{3}$$

$$\tilde{C} = 1 - \frac{f_1 - 2f_2/(S-1)}{\sum_{t=1}^{S} t f_t}, \tag{4}$$

where S denotes the number of CE sessions and f_i is the number of causes discovered i times over these S sessions. The population size can be estimated as in (5) or in (6).

$$\hat{N} = \frac{M_{k+1}}{\hat{C}} + \frac{f_1}{\hat{C}} \hat{\gamma}^2, \text{ and} \tag{5}$$

$$\tilde{N} = \frac{M_{k+1}}{\tilde{C}} + \frac{f_1}{\tilde{C}} \tilde{\gamma}^2, \tag{6}$$

where $\hat{\gamma}^2$ and $\tilde{\gamma}^2$ are estimators for the coefficient of variation of the capture probabilities and are calculated as in Equations (7) and (8).

$$\hat{\gamma}^2 = \max\left\{\frac{kM_{k+1}\sum_{i=1}^{S} i(i-1)f_i}{\hat{C}(k-1)(\sum_{i=1}^{S} if_i)^2} - 1, 0\right\}, \text{ and} \tag{7}$$

$$\tilde{\gamma}^2 = \max\left\{\frac{kM_{k+1}\sum_{i=1}^{S} i(i-1)f_i}{\tilde{C}(k-1)(\sum_{i=1}^{S} if_i)^2} - 1, 0\right\}. \tag{8}$$

Variation estimators for population size estimates are presented in [6].

The sample coverage approach presented above is argued to have the following advantages [7]. It offers closed form of estimates that provides a measure of the degree of heterogeneity and the estimators approach the true (population) parameters when the discovery probabilities p_i are Gamma distributed. On the other hand, the sample coverage approach is appropriate for studies where at least five consecutive CE diagram sessions are held.

Step 4. Finalization:

Apply a stopping criterion and finalize the CE diagram study. A good criterion can be built based the ratio of discovered causes to the expected number of total causes calculated using the jackknife estimators or the sample coverage approach. The CE diagram construction study might be terminated once the ratio is above an admissible level determined by the practitioners. The last task to accomplish is to create the final CE diagram (or matrix) to display the results.

An Illustrative Example

This section covers a CE diagram construction study. The example we present is a simulation. Simulation experiments are often utilized as a viable option to conceptually validate theoretical models. In this example, Jackknife estimators are used to estimate the population size. As discussed in Step 4 of the previous section, an estimate of the population size can be utilized to create a stopping criterion.

Problem Settings

Consider a problem of 100 causes. Assume that these causes are unknown and a group of experts is formed to discover those. Moreover, assume that the experts do not know the exact number of causes, but based on the difficulty of identifying the causes, they can discover those with different probabilities as in Table 3.

Table 3. Cause identification probabilities.

Group #	Causes	Probability	Description
1	1-20	0.75	Easy to identify
2	21-40	0.50	Moderately easy
3	41-60	0.25	Challenging
4	61-80	0.10	Hard
5	81-100	0.05	Expertise required

To estimate the number of causes, we fit an M_h CR model to the data presented in Table 3. The M_h model allows different individual capture probabilities (see column 3 of Table 3) for all causes, but assumes that these probabilities remain constant over time. The reliability of

this assumption is discussed in the Improved Cause and Effect (CE) Diagrams Section. Moreover, the capture probability of each cause is assumed to be independent from capture probabilities of others.

Next, we use the Jackknife estimators presented in Equations (1) and (2). Our aim is to simulate a number of example runs and test the quality of population size estimates. In real life applications the number of sessions to hold is limited due to time and resource constraints. We provide estimates for the total number of causes based on the number of sessions held, causes discovered and different probabilities of discovery. The example can be used as a guideline in determining the number of sessions to hold.

Simulation Study

Suppose the expert group can only hold up to 5 sessions due to resource and time restrictions; and replicates the simulation for 1,000 times. We use three jackknife size estimates: N_1, the level-1 estimate, N_2, the level-2 estimate and N_3, the level-3 estimate.

The steps described above are implemented in Matlab R2009 and run on a 3GHz dual core Windows PC with 4 GB of RAM. The run time of the algorithm for 1,000 iterations was about one minute.

Figure 3 plots these estimates over 1,000 simulation runs and Table 4 presents statistics on the estimates based on 1,000 runs. The horizontal axis on Figure 3 is the simulation run; the vertical axis is the estimate of the number of total causes.

On Table 4, N_1 is the estimate calculated by the first order Jackknife estimator, N_2 is calculated using the second and N_3 is calculated using the third level estimators. We can see that N_3 is a better estimate because it gives, on average, the closest estimate to the real population size (100). However, it is important to note that to calculate N_3 however, experts need to hold at least 3 CE sessions, which may be impractical in real life situations. Moreover, the standard deviation of the population size estimators over 1,000 runs increases with the estimator level. Improving the performance of the estimators is left as a future research direction.

Table 4. Population size estimate statistics (based on 1000 runs).

Estimate	Average	Std Dev
N_1	84.79	6.39
N_2	91.09	8.84
N_3	93.37	10.93

Table 5 and 6 provide additional information on the simulation experiment. Table 5 shows the average, maximum and minimum number of causes of discovery probability p_i (refer to Table 3) that has been discovered over the simulation study. As expected, the number of causes identified decreases when the degree of difficulty of identifying the cause increases. Table 6 gives a sample of f_i values, the number of causes identified i times. As expected, only a few causes have been identified a few times.

It is important to note that the simulation study is for verification of the methodology proposed in the paper and the result presented are valid only for the parameter values given in this section.

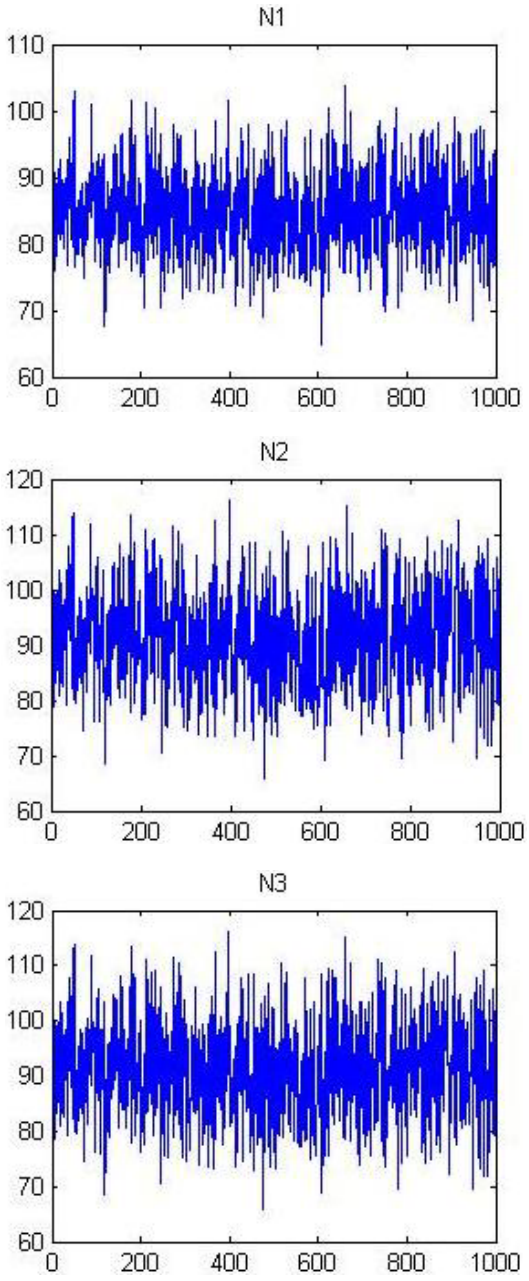

Figure 3. Population size estimates in 1000 runs.

Table 5. Statistics on the number of causes identified by probability of discovery.

p_i	Average	Maximum	Minimum
0.75	16	18	12
0.50	10	13	7
0.25	4	7	2
0.10	3	4	1
0.05	1	3	0

Table 6. Capture frequency results.

i	1	2	3
f_i	28	13	11

5. Conclusions and Future Work

In this paper, we propose a 5-step analytical methodology to improve the construction of CE diagrams. Specifically, a closed population Capture-Recapture (CR) method is used to improve CE diagram, which is typically built upon the brainstorming process. Experts hold successive CE sessions to discover the causes of a problem. The CR model is used to estimate the unknown number of causes of the problem and the proposed methodology is illustrated with a case example.

Several points of our work are open for further research. We assume here that once a cause is discovered, it is never removed from the checklist. In real applications, some causes may be evaluated as irrelevant and removed from the CE diagram. In such cases the number of the causes changes and open population CR models would be more appropriate than closed population models. Secondly, some more sophisticated CR methods might be investigated to improve the estimation process. Most of the recent models use simulation applications that may incorporate more features related to the analyzed problem. Thirdly, the simulation experiment we presented in the paper is for validation purposes. Our next application direction will be to implement our methodology through a real case study. Another future research direction is to explore methods other than the Sample Coverage approach when estimating the total number of causes. The Sample Coverage, as a rule of thumb, performs better when at least five runs are made, whereas this requirement may not always be reasonable in real life situations. Finally, Table 4 illustrates a feature with the method as standard deviations are getting larger as number of samples increase. Future work will include efforts to address this point.

Acknowledgements

The authors acknowledge the constructive comments of the Associate Editor and anonymous referees for improving the quality and content of the manuscript.

References

1. 12 Manage website, last accessed 05.26. (2008). http://www.12manage.com/methods_ishikawa_cause_effect_diagram.html
2. Amstrup, S. C., McDonald, T. L. and Manly, B. F. J. (2005). *Handbook of Capture-Recapture Analysis*, 1st edition. Princeton University Press, Princeton.
3. ASQ website, last accessed 05.26. (2008). http://www.asq.org/learn-about-quality/cause-analysis-tools/overview/fishbone.html
4. Burnham, K. P. and Overton, W. S. (1978). Estimating the size of a closed population when capture probabilities vary among animals. *Biometrika*, 65, 625-633.
5. Chao, A. (2001). An overview of closed capture recapture models. *Journal of Agricultural, Biological and Environmental Statistics*, 6(2), 158-175.

6. Chao, A., Lee, S. M. and Jeng, C. L. (1992). Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics*, 48, 201-216.

7. Chao, A. and Huggins, R. M. (2005). Classical Closed Population Capture-Recapture Models. In *Handbook of Capture-Recapture Analysis* (Edited by Amstrup, S. C., McDonald, T. L. and Manly, B. F. J.), 22-35. Princeton University Press.

8. Cormack, R. M. (1964). Estimates of survival from the sightings of marked animals. *Biometrika*, 51, 429-438.

9. Hermens, M. (1997). A new use for Ishikawa diagrams. *Quality Progress,* 30(6), 81-83.

10. Johnson, R. P., Yang, Q. and Butler, C. (1998). Dynamic error characteristics of touch trigger probes fitted to coordinate measuring machines. *IEEE Transactions on Instrumentation and Measurement*, 47(5), 1168-1172.

11. Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration: a stochastic model. *Biometrika*, 52, 225-247.

12. Kufelnicki, A., Lis, S. and Meinrath, G. (2005). Application of cause-and-effect analysis to potentiometric titration. *Analytical and Bioanalytical Chemistry*, 382(7), 1652-1661.

13. Lee, S. M. and Chao, A. (1994). Estimating population size for closed capture-recapture models via sample coverage. *Biometrics*, 50, 88-97.

14. Lin, T. Y., Xiong, Z. P., Yao, Y. F., Tok, L., Yu, Z. Y., Njoman, B., Chua, K. H. and Ma, Y. Y. (2003). Failure analysis of full delamination on the stacked die leaded packages. *Transactions of the ASME Journal of Electronic Packaging*, 125(3), 392-399.

15. Manly, B. F. J., McDonald, T. L. and Amstrup, S. C. (2005). Introduction to the Handbook of Capture-Recapture Analysis. In *Handbook of Capture-Recapture Analysis* (Edited by Amstrup, S. C., McDonald, T. L. and Manly, B. F. J.), 1-21. Princeton University Press.

16. Matejcik, F. J. (2007). Sampling procedures for extending the use of the fishbone diagram. South Dakota School of Mines and Technology. *Working paper*.

17. Meinrath, G. and Lis, S. (2002). Application of cause-and-effect diagrams to the interpretation of UV-Vis spectroscopic data. *Analytical and Bioanalytical Chemistry*, 372, 333-340.

18. Muncy, J. (2004). Modeling reliability of flip chip on board assemblies implementing a correction function approach comparing analytical and finite element techniques. *Proceedings of IEEE / SEMI International Electronics Manufacturing Technology Symposium*.

19. Pollock, K. H. and Alpizar-Jarra, R. (2005). Classical Open-Population Models. In *Handbook of Capture-Recapture Analysis* (Edited by Amstrup, S. C., McDonald, T. L. and Manly, B. F. J.), 36-57. Princeton University Press.

20. Schippers, W. A. J. (1999). The process matrix, a simple tool to analyze and describe production processes. *Quality and Reliability Engineering International*, 15, 469-473.

21. Seliger, W., Wolfgang, E., Lefranc, G., Berg, H. and Licht, T. (2002). Reliable power electronics for automotive applications. *Microelectronics Reliability*, 42, 1597-1604.

22. Silver, E. A. and Rohleder, T. R. (1998). Some simple mathematical aids for cause-and-effect analysis. *Journal of Quality Technology*, 30(1), 85-92.

23. Skjei, S. (2005). Using a fishbone diagram to troubleshoot a satellite link. *The Orbiter*, April/May.

Appendix A: The Theorem and Its Proof

Theorem: The number of undiscovered causes decreases as the number of CE sessions increases.

Proof: The number of undiscovered causes by the end of the k-th session can be estimated using $U_k = N - \sum_{i=1}^{k} pn_i$. We remove the superscript on $p^{(i)}$ for ease of presentation. This can be proven by recursion as follows. Using Figure 2 as a guideline, we can calculate the expected number of undiscovered causes at the end of the first CE session to be

$$p(N - n_1) + (1 - p)N = N - pn_1 = U_1.$$

Similarly, the expected number of undiscovered causes at the end of the second CE session is,

$$p[p(N - n_1 - n_2) + (1 - p)(N - n_1)] + (1 - p)[p(N - n_2) + (1 - p)N] = N - pn_1 - pn_2 = U_2,$$

where the left hand side of the equality can be calculated via tedious, but straightforward algebra. Assume at the end of session k the expected number of discovered causes is U_k. Using our notation above U_k is expressed as

$$U_k = N - pn_1 - pn_2 - \ldots - pn_k = U_{k-1} - pn_k.$$

Over the $k+1^{st}$ session, the expert team is expected to discover pn_{k+1} causes. Therefore, at the end of session $k+1$ the expected total number of undiscovered causes is

$$U_{k+1} = N - pn_1 - pn_2 - \ldots - pn_k - pn_{k+1} = U_k - pn_{k+1}.$$

Hence, $U_k = N - \sum_{i=1}^{k} pn_i$. Now note that p is the probability of discovery; thus, $p \in [0, 1]$. Moreover, n_i is the number of causes; thus, $n \geq 0$ and therefore, $U_k - U_{k-1} = pn_k \geq 0$; hence the proof.

Appendix B: More on Capture-Recapture (CR) Analysis

CR sampling techniques and models are widely used in wildlife sciences to estimate parameters of animal populations [2]. The basic idea behind CR analysis is to sample from a population of unknown size at different occasions and use the samples to estimate parameters of the population. One of the first real applications of CR techniques dates back to 1802 where the famous French mathematician Pierre Laplace used a sampling method to estimate the population of France [15]. He started with an initial "guess" on the yearly number of newborns, acquired live birth and census data from three municipalities in France and used the ratio of live birth to total population of these municipalities for estimating the country's population. There is no report on the accuracy of his projection, but Laplace's study is accepted as one of the earliest CR analysis ever conducted. As in Laplace's experiment, current CR models are mostly interested in estimating the size of the population in question. Depending on the assumptions on population dynamics, CR models are divided in two main categories as *closed population CR models* and *open population CR models*.

In the subsequent subsections, we will discuss about the properties of closed population and open population CR models and briefly present the most frequently referred statistical models in each category. Note that we will follow the jargon used in the CR literature. We denote individual experiment subjects as animals, the animals marked in any capture

occasion are referred to as "tagged" or "marked" and the capturing device or method is called a "trap". A capture study consists of several capture occasions. Animals are captured, marked and released during a capture occasion. A number of capture occasions are executed to complete a CR study.

Closed Population CR Models

Closed CR models assume no change in the number of individuals of the studied population during the study period [7]; that is, it is assumed that there is no birth, death or migration during the period of study [5]. Closed population models are deemed to be the first analytically studied CR models [7]. In closed population CR models analysts collect data on the capture history of animals. For large experiments the individual capture dataset may become very large; therefore, researchers usually represent the data by a tally of frequencies of each capture history [7]. For instance, in a large study analysts keep track of the number of animals captured for some specific times; for instance the number of animals captured k times during the study. Additional to the population being closed, the following assumptions are made in closed population CR models:

- Animals do not lose their tags,
- All tags are correctly recorded,
- Animals act independently.

Open Population CR Models

Closed population CR models assume that the population size would remain fixed during the period of a CR study. However, wildlife systems are very dynamic in nature, animals emigrate from or immigrate into the population, births and deaths occur over time. Hence, in wildlife sciences the closed population assumption introduced for modeling ease is usually not verified. Open population models remove this assumption and analyze the wildlife population as a dynamic system.

The most frequently referred open population CR models are the Jolly-Seber (JS) model in [11] and the Cormack-Jolly-Seber (CJS) model in [8]. Here we discuss the CJS model which is also classified as a *conditional model* because it yields probabilities of occurrence of a given capture history conditional on the number of animals released at a sampling occasion (details of the JS model can be found in [19]). The CJS model can be used to estimate for instance the conditional probability P(01010|release in period 2) where the capture history (01010) is used to represent no-catch at the first, third and fifth periods, first catch occurring at the second period and the only recatch at the fourth period. The probability quantity is computed conditional on the animal being released at the second period; that is, right after being caught. The model is general enough to allow releases happening in periods after the first catch period to include two possible cases: the animal might die at the trap or might be kept for several periods.

Authors' Biographies:

R. Ufuk Bilsel received his B.S. and M.S. degrees, both in Industrial Engineering, from Galatasaray University in Istanbul, Turkey, in 2003 and 2005 respectively. He received his M.Eng degree in Industrial Engineering and Ph.D. in Industrial Engineering and Operations Research from The Pennsylvania State University in 2008 and 2009 respectively. He is currently with The Boston Consulting Group. His areas of interest include supply chain optimization, decision making under uncertainty, risk analysis and statistical analysis of large datasets. He is a member of Alpha Pi Mu, Industrial Engineering Honor Society.

Dennis K. J. Lin Dr. Dennis Lin is a University Distinguished Professor of Statistics and Supply Chain Management at Penn State University. His research interests are quality engineering, industrial statistics, data mining and response surface. He has published over 150 papers in a wide variety of journals. Dr. Lin is an elected fellow of ASA and ASQ, an elected member of ISI, a lifetime member of ICSA, and a fellow of RSS. He is an honorary chair professor for various universities, including a Chang-Jiang Scholar of China at Renmin University, National Chengchi University (Taiwan), Fudan University, Jinan University, and XiAn Statistical Institute (China). Dr. Lin presents several distinguished lectures, including the 2010 Youden Address (FTC) and the 2011 Loutit Address (SSC). He is also the recipient of the 2004 Faculty Scholar Medal Award at Penn State University.