Safety Science 121 (2020) 100–110

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

A framework for corridor-level traffic safety network screening and its implementation using Business Intelligence

Naveen Veeramisti[a], Alexander Paz[b,*], Justin Baker[c]

[a] CA Group, Inc., 2785 S Rainbow Boulevard, Las Vegas, NV 89146, United States
[b] Civil Engineering and Built Environment, Science and Engineering, Queensland University of Technology, 2 George St, Brisbane, QLD 4000, Australia
[c] Las Vegas Water Valley District, United States

ARTICLE INFO

Keywords:
Traffic safety management
Network screening
Corridor screening
Crash frequency
Business intelligence

ABSTRACT

Currently, there are both methodological and practical barriers that together preclude the use of theoretically sound approaches for network screening as part of a traffic safety management process. Methodological barriers include, among others, lack of a comprehensive framework for corridor-level network screening. Existing corridor screening methodologies use observed crash frequency as a performance measure. In practice, corridor-level screening is extremely important because traffic safety engineers prefer to deploy countermeasures and provide homogenous conditions throughout corridors to meet drivers' expectations and avoid confusion. On the other hand, practical barriers that limit the use of sound approaches for traffic safety include (1) significant data integration requirements, (2) a particular data schema is needed to enable analysis using specialized software, (3) time-consuming and intensive processes are involved, (4) substantial technical knowledge is needed, (5) visualization capabilities are limited, and (6) coordination across various data owners is required. This research proposes a systematic methodology for corridor-level network screening. The solution algorithm is implemented within a Business Intelligence (BI) platform to address, to the extent possible, the practical barriers listed above. BI provides methods and mechanisms to integrate and process data, generate advanced analytics, and visualize results by using intuitive and interactive web-based dashboards and maps. Experiments and results illustrate the advantage of using the proposed framework for corridor-level network screening implemented within a BI platform.

1. Introduction

Ensuring traffic safety is the focus of such federal legislation as the Transportation Equity Act for the 21st Century (TEA-21), the Safe Accountable Flexible and Efficient Transportation Equity Act - A Legacy for Users (SAFETEA-LU), and the Moving Ahead for Progress in the 21st Century (MAP-21). SAFETEA-LU and MAP-21 both require that states develop comprehensive Highway Safety Improvement Plans (HSIPs) (FHWA, 2013). One of the critical programs of HSIPs is the traffic safety management process, which involves annual reporting of the highway locations that exhibit the most severe traffic safety needs. By identifying the most hazardous roadway site locations, specific countermeasures can be implemented to improve safety conditions. In a traffic safety management process, identifying locations with the potential for safety improvements is known as network screening. Network screening is critical because a detailed engineering study for all network sites is expensive. The purpose of network screening is to review the entire roadway network, or portions, and identify and prioritize corridors and sites with potential for safety improvements. These identified corridors are recommended for further investigation and a detailed safety engineering study.

For network screening, despite the availability of approaches such as the Empirical Bayes (EB) method recommended by the Highway Safety Manual (HSM) (AASHTO, 2010) and those that deal with unobserved heterogeneity (Mannering et al., 2016; Mannering, 2018), practitioners continue using very simplistic methodologies which rely only on observed crash frequency or crash rates. For HSIP reports submitted to the Federal Highway Administration in fiscal year 2014, only four states used the EB for network screening as described in the HSM (FHWA, 2015).

Currently, there are both methodological and practical barriers that together preclude a substantial use of theoretically sound approaches for network screening as part of the traffic safety management process. Although the state-of-the-art provides great modeling tools, there are

* Corresponding author.
E-mail address: alexander.paz@qut.edu.au (A. Paz).

https://doi.org/10.1016/j.ssci.2019.08.042
Received 20 December 2018; Received in revised form 5 July 2019; Accepted 28 August 2019
Available online 11 September 2019

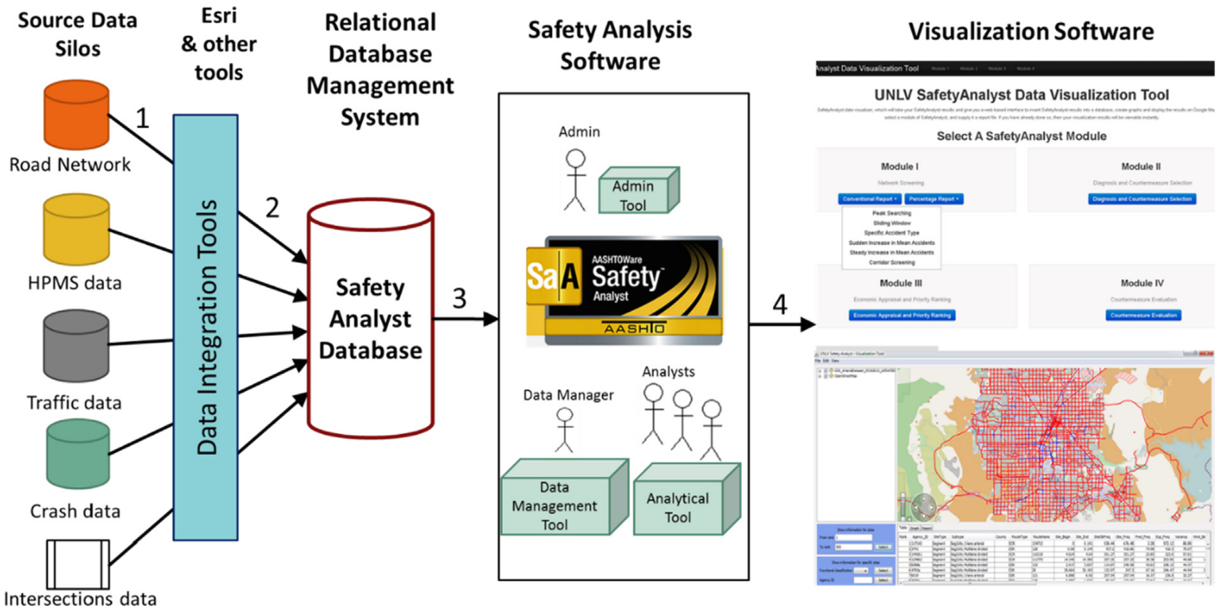

Fig. 1. An example of a traditional approach for traffic safety analysis.

still various important capabilities missing. A key methodological barrier is the lack of a theoretically framework for corridor-level network screening. In practice, screening and analysis/ranking at the corridor level for the entire network is required for various reasons including the need to provide as homogeneous as possible infrastructure conditions across the roadway network to meet drivers' expectations. Homogeneous conditions are associated with less driving distractions, surprises, or confusion. In order to provide homogenous infrastructure conditions, it is recommended to improve safety of long sections of roadways, corridors. These corridors could include multiple sites with the potential for safety improvements. Few agencies have corridor-wide safety programs. Some programs in place include Nevada's Kietzke Lane Safety Management Plan; the Safe Corridor Programs of both New Jersey and Wisconsin; and Integrated Corridor Management plans that develop safety plans/programs for cities and municipalities (Nemmers et al., 2008; Shimko and Walbaum, 2010).

Several studies have used observed crash frequency, crash rates, or a crash severity index for corridor screening (ITT, 2011; Qin et al., 2013). Using observed crash frequencies may result in a volume bias, while using crash rates may result in a segment length bias. In addition, using observed crashes may results in a regression-to-the-mean bias. For corridor screening, Hamidi et al. (2010) used crashes that occurred only on major roads at intersections. Ignoring interactions of major and minor road characteristics at intersections affects predicted crash frequency, leading to incorrect estimation of expected crash frequency. Other studies (ITT, 2011; Zhao et al., 2017) did not search corridors that had potential for safety improvements, but instead estimated the crash frequency on pre-aggregated sites. As an alternative for defining corridors for implementing safety improvements based on aggregated sites or lengths, determining them using a Sliding Window mechanism and characteristics including crash data provides a superior approach. A Sliding Window mechanism addresses crash location reporting errors by evaluating the same section of roadway multiple times, using overlapping windows.

The use of existing theoretically sound approaches is also limited by important practical barriers including (1) significant data integration requirements, (2) particular data schemas needed to enable analysis using specialized software, (3) time-consuming and intensive processes are involved, (4) substantial technical knowledge is needed, (5) visualization capabilities are limited, and (6) coordination across various data owners is required.

With regards to practical barriers, Alluri and Ogle (2012) documented the current safety-analysis practices related to engineering as used by various states. In addition, they described perspectives in adopting and implementing the methods provided by the HSM. Alluri and Ogle (2012) indicated that barriers faced by traffic safety engineers include requirements for comprehensive data sets, data integration, and management. Tarko et al. (2014) and Paz et al. (2015) discussed the complexities of data integration and management for network-level traffic safety analysis, specifically for the traffic safety management process.

GIS methodologies developed by ESRI® have been widely used for traffic safety data processing, analysis, reporting, and visualization (Pulugurtha et al., 2007; Wellner and Qin, 2011; Aylo, 2010). The Critical Analysis Reporting Environment (CARE), developed by the University of Alabama (CAPS, 2009), sorts, analyzes, and compares crash data using functions that allow statistical analyses with charts and graphical displays (Paz et al., 2014). For traffic safety management, Safety™ Analyst is an AASHTOWare™ software system that was developed using network screening methods from the HCM (ITT, 2011). Although Safety Analyst provides significant capabilities, the software has several limitations. For example, data integration and processing capabilities are lacking, the data needs to be in a specific schema, and it has marginal visualization capabilities. Tjandra (2014) developed a Business Intelligence (BI) system for traffic data integration by linking roadway, crash, and traffic flow data to improve traffic safety. The system provides descriptive performance indices for traffic safety.

At present, traffic safety analysis usually is conducted using a traditional approach that involves multiple individual steps and such components as customized data management and visualization tools. This makes the process complicated and time consuming. Fig. 1 illustrates an example of the traditional approach for traffic safety analysis.

This paper proposes a framework to address the above listed methodological and practical barriers to enable the broad adoption of sound methodologies for network screening. The framework was implemented using a Business Intelligence (BI) approach, which provides methods and mechanisms to integrate and process data, generate advanced analytics, and visualize results. The proposed framework facilitates traffic safety engineering and enhances the outcomes of a HSIP.

As yet, no single framework exists that provides capabilities for (1) data process, integration and management, (2) advanced analysis, and (3) visualization. Key characteristics of the BI framework proposed in

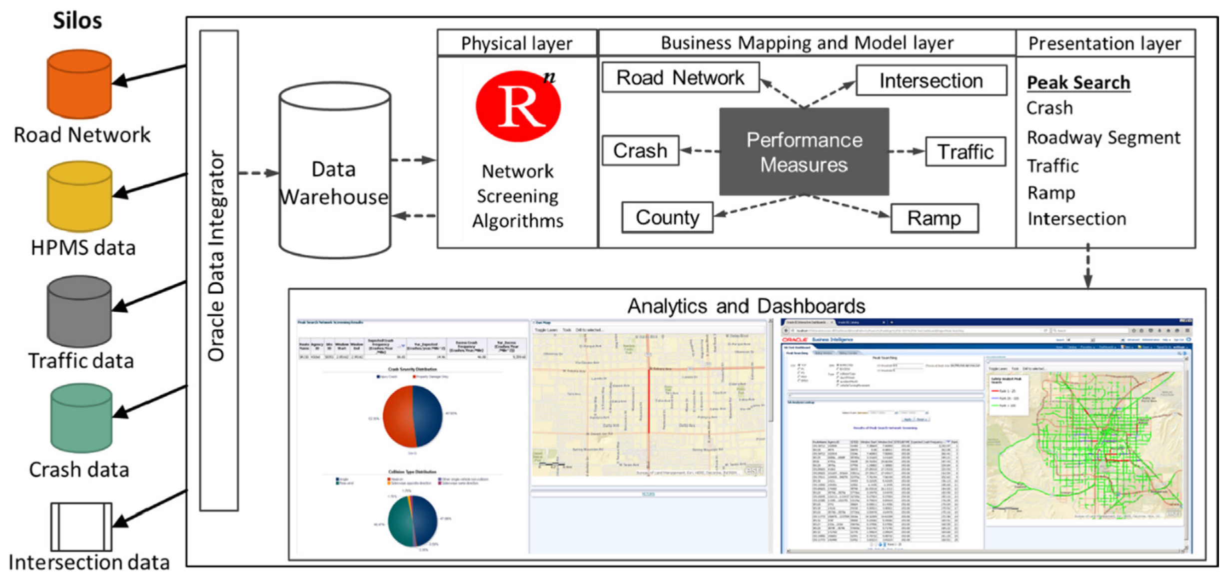

Fig. 2. A business intelligence approach for traffic safety analysis.

this paper for implementation include: (1) an extract-load-transform (ELT) process; (2) tools for integration of data from a wide variety of data sources; (3) algorithms for sound analysis; (4) a methodology for effective corridor-level network screening; and (5) visualization tools for network-wide site-specific and corridor-level analysis. These characteristics provide an effective platform for generating analysis and information for various types of decision makers. Furthermore, as new source data is provided to the proposed framework, all analyses, reports, and visuals are updated.

In addition to the proposed new methodology for corridor screening, network screening algorithms from the HSM were implemented within a BI framework to provide a single platform for data integration, management, analysis, and visualization. For illustration purposes, our implementation was performed using Oracle R Enterprise (Rittman, 2013) within the Oracle Business Intelligence Enterprise Edition (Dupupet et al., 2013). With this framework, as illustrated in Fig. 2, network screening can be performed on a web-based interface with a data warehouse directly connected to the source data. This proposed approach represents a paradigm shift by which algorithms required for analyses are made available to practitioners by means of a platform that addresses existing barriers that currently prevent the use of these methodologies. A similar implementation can be performed using alternative BI platforms.

The contributions of this research include a systematic methodology for corridor-level network screening and a comprehensive implementation using concepts from data warehousing and Business Intelligence to facilitate the adoption of the proposed approach. Required data sources include those commonly used by state agencies.

Critical data for analysis from various sources was integrated including crashes and associated characteristics, traffic flow information, and roadway geometry features. A theoretically sound approach for corridor screening was proposed and implemented to estimate expected crash frequencies, using the integrated data, and rank roadway corridors with potential for safety improvements. Business Intelligence dashboards were created to visualize the results of the analyses. Data visualization and exploration capabilities were enabled to rank corridors with potential for safety improvements as well as to drill down into sites to analyze crashes, traffic patterns, and associated characteristics for effective countermeasure selection.

2. Methodology

Network-wide corridor screening provides the capability to compare the safety performance across extended corridors rather than individual sites (AASHTO, 2010). A corridor may consist of multiple roadway segments, intersections, and/or ramps, which are aggregated together and analyzed as a single entity. In this study, two types of corridor screening algorithms are proposed: (1) Fixed Corridor screening, and (2) Corridor Searching. Fixed Corridor screening can be used for predefined corridors. When a user specifies predefined corridors, the expected crash frequency of these corridors is estimated by aggregating the expected crash frequencies of individual roadway elements. These predefined corridors are ranked from the highest to the lowest with regard to expected crash frequencies. This method is useful when engineers are evaluating known corridors in the network. Sites are assigned by the engineer/analyst to a specific corridor at the data management level. If sites assigned to the corridor need to be modified or more corridors need to be added, the analyst is required to perform this task.

Corridor Searching reviews a road network in a systematic manner to identify corridors, using a corridor length and an incremental length. The user selects a predefined length to estimate the expected crash frequency of the corridor and selects a predefined incremental length that slides the corridor to evaluate the following corridor. For each corridor, the expected crash frequency is estimated by aggregating the expected crash frequencies of individual roadway elements, such as roadway segments, intersections, and/or ramps. Then, corridors in the network are ranked from highest to lowest with regard to expected crash frequencies. Moving the corridor by a small incremental length is used to compensate for sites being falsely selected that had randomly high crash counts.

2.1. Notation and definitions

The following notation and definitions are used to describe the proposed solution algorithm for corridor screening.

PF_{yi}	Predicted crash frequency of site i in year y
$AADT_{yi}$	Annual Average Daily Traffic of site i in year y
α, β_1 and β_2	Estimated model parameters for crash frequency functions
c_{yi}	Estimation factor for site i in year y
C_{yi}	Yearly correction factor for year y relative to year 1 at site i
O_{yi}	Number of observed crashes for year y at site i

w_i Weight for site i calculated for the EB method. If the EB is not required because the model used to predict crash frequencies is able to handle unobserved heterogeneity (refer explanation in Section 2.2), w can be set equal to 1 to provide full weight to the predicted performance function.
b_i Over-dispersion parameter obtained from SPF regression for site i belonging to corresponding site subtype
L_i Length of site i
EF_i Expected crash frequency for site i
EFc Expected crash frequency of the considered corridor

2.2. Proposed solution algorithm

The solution algorithm proposed in the following steps can be applied to both Fixed Corridor and Corridor Search screening. Expected crash frequency for all sites in the corridor is estimated using steps 1–5. In the proposed framework, functions, PFs, used to estimate crash frequencies are inputs. For illustration purposes, expected crash frequencies are estimated using simple performance functions which can be substituted by superior expressions capable of dealing with critical aspects such as unobserved heterogeneity. PFs were estimated only using AADTs for segments and intersections. However, potential crash contributing factors include among others vehicle types and designs, impact angles, lighting, environment, and pavement conditions, weather, and other physiological and human factors (Mannering et al., 2016; Mannering, 2018). PFs estimated considering all these data can better address unobserved heterogeneity. Additional data and further research are required to develop such functions (Mannering et al., 2016; Mannering, 2018). Step 6 is used to aggregate and obtain the average expected crash frequency (crashes per mile per year). The corridors in the network are then compared and ranked in Step 7.

Step 1: Calculate the predicted crash frequency per mile for roadway segments and intersections and ramps in a corridor for each year, using Eqs. (1) and (2), respectively.

$$PF_{yi} = c_{yi} * e^{\alpha} * AADT_{yi}^{\beta} \tag{1}$$

$$PF_{yi} = c_{yi} * e^{\alpha} * AADT_{yi}^{\beta 1} * AADT_{yi}^{\beta 2} \tag{2}$$

Usually, the data contains various site subtypes of roadway elements. Hence, appropriate model parameters, α and β, for the PF for associated site subtypes need to be used. PFs for site subtypes that were estimated using local data are preferred over those available in the literature. PFs estimated using data from other regions need to be corrected using local information. A correction factor multiplies the PFs.

Step 2: Compute the yearly correction factors for number of years considered in the data, using Eq. (3).

$$C_{yi} = \frac{PF_{yi}}{PF_{1i}} \tag{3}$$

Step 3: Compute the weights, w, to provide weightage for observed and predicted crashes, using Eq. (4).

$$w_i = \frac{1}{1 + b_i \sum_{y=1}^{Y} PF_{yi} * L_i} \tag{4}$$

Step 4: Calculate the expected crash frequency for the first year of data for site i, using Eq. (5). The unit of expected crash frequency is crashes per mile per year. In the case of intersections and ramps, the length is considered as '1' and the units are crashes per year.

$$EF_{1i} = w_i PF_{yi} + \frac{(1 - w_i) \sum_{y=1}^{Y} O_{yi}}{L \sum_{y=1}^{Y} C_{yi}} \tag{5}$$

Step 5: Calculate the expected crash frequency for the final year of data for site i, using Eq. (6). The unit of expected crash frequency is crashes per mile per year. As in Eq. (4), in the case of intersections and ramps, the length is '1' and the units are crashes per year.

$$EF_{Yi} = EF_{1i} * C_{Yi} \tag{6}$$

Step 6: Calculate the expected crash frequency for the entire corridor, using Eq. (7).

$$EF_C = \sum_{i=1}^{I} EF_{Yi} \tag{7}$$

Step 7: Rank corridors in order according to *EFc*.

For Corridor Searching using the Sliding Window mechanism, the starting and ending sites in the corridor could be a fraction of a site. For these cases, the length of the site, L, is the length of a fraction of the site considered in the corridor. Similarly, for observed number of crashes, the number of crashes on the corresponding fraction of site should be used. Both Fixed Corridor and Corridor Searching algorithms, including ELT process and star schemas (ASHTO, 2010), were implemented using the proposed BI approach.

In addition to Network-wide Corridor Screening, Network Screening for individual sites was also designed and implemented using the BI approach. The three network screening algorithms implemented in this study were (1) Peak Search, (2) Sliding Window, and (3) Simple Ranking. In the experiments, the first two were used for roadway segments, and the third one was used for intersections and ramps. The HSM and other literature (AASHTO, 2010; Montella, 2010) identified Peak Search and Sliding Window as the two recommended algorithms for network screening for individual sites along roadway segments.

3. Implementation

Oracle Data Integrator (ODI) Edition (Dupupet et al., 2013) was used to build a safety data warehouse, which is accessed by OBIEE to facilitate the development of advanced analytics, dashboards, and maps. The connection to the database was created by the Repository Design Model (RPD), which contains physical models, business mapping models, and presentation models for use by OBIEE (Rittman, 2013). Oracle R Enterprise (McDermid and Taft, 2014) scripts were developed to implement network screening algorithms. These scripts were executed in the physical layer (Rittman, 2013) of the RPD. The output of the Oracle R Enterprise scripts was saved in datastores, which allowed other queries to access the results for network screening. These queries were used in the RPD to enable OBIEE to perform on-the-fly computation and retrieval of the network screening results in the dashboards. The JavaScript application program interface (API) (ArcGIS, 2015) for Esri® maps was used in the dashboards, along with analytics to display network screening results and associated site locations spatially.

3.1. Data warehouse design

Silos of source data from various sources were collected and integrated using ODI to create a safety data warehouse. Source data includes the physical characteristics of road network, traffic volumes, and the Highway Performance Management System (HPMS) as well as crash data and their associated characteristics. Table 1 provides silos of source data and their associated data elements. Additional details of data collection were provided in Paz et al. (2015).

The data warehouse was developed using an ELT process. ODI interfaces extract data from the source, and loads the information, using a Loading Knowledge Module (LKM), into the OBIEE target database (Rittman, 2013). In this study, the data was transformed into a star schema for use with OBIEE. Data across systems were integrated using a location reference system, County/Route/Milepost. Crashes and their characteristics were mapped to the segments, intersections, and ramps. Similarly, traffic stations were mapped to road segments in order to obtain the average annual daily traffic (AADT) on respective road segments.

Data from signalized intersections can be obtained from such sources as the Freeway and Arterial System of Transportation (FAST) (Xie and Hoeft, 2014) of the Regional Transportation Commission of

Table 1
Silos of source data and their associated data elements.

Road network	HPMS files	TDM model	Crash data	Intersection
Segment Identifier	Routes	Travel direction	Accident Identifier	Intersection Identifier
Ramp Identifier	Functional classification	Functional classification	Crash location	Intersection location
Length of Segment	Access control	Operation way	Cash date	Type of control
Begin milepost	Speed limit	Speed limit	Collision type	Number of legs
End milepost	Through lanes	Number of lanes	Severity	
Route ID	Left Lanes	County	Relationship to junction	
County	Right Lanes	Area code	Direction of involved vehicle	
Milepost direction	AADT	Ramp configuration	Maneuvers by involved vehicle	
	Urban	Ramp type	Weather conditions	
	County	Segment length	Environment conditions	
	Median Type		Pavement conditions	
	Median Width		Vehicle types	
			Drug involvement	
			Pedestrian/Bicycle involvement	

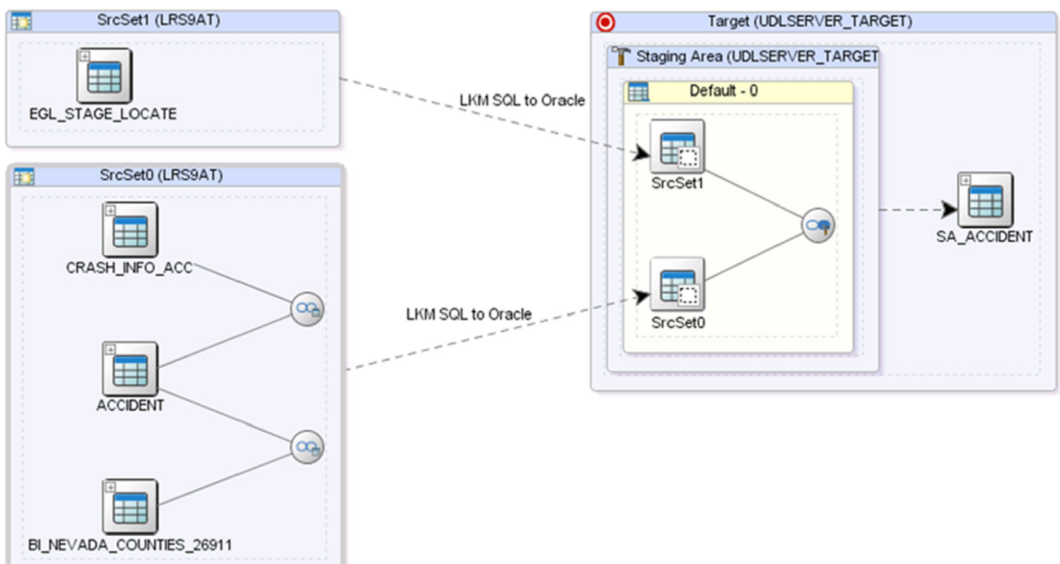

Fig. 3. ELT process of crash information to the SA_ACCIDENT target table.

Southern Nevada (RTC-SN). Data from stop-controlled intersections can be collected using Google Earth. ODI can be used to integrate the data from intersections with data from the road network as well as with crash data. Fig. 3 illustrates an ELT process for crash-related information from various tables of crash data to a target database table, SA_ACCIDENT. Similarly, three target database tables were created, SA_ROADWAYSEGMENT, SA_INTERSECTION, and SA_RAMP. Contiguous sites with similar physical characteristics needed to be aggregated in order to create homogeneous segments. Procedures were developed using Oracle Procedural Language/Structured Query Language (Pl/SQL), and were connected to a web-based interface for homogeneous segmentation. As a result, analysts and engineers could use a Choice List in the interface to choose parameters that can be used for homogeneous segmentation. These include such parameters as a district, county, or route; the number of through lanes in one direction or a combined direction; median type; median width; and the percentage of the AADT threshold. By using this interface, as shown in Fig. 4, the segments can be aggregated, and the new site list created and stored in the target database for further analysis.

Once the site list is created, the sites with characteristics for area code, functional class, number of lanes, access control, and median type can be used to group sites into site subtypes. This operation easily can be performed using a single SQL statement. Sites with the same site subtypes are used to estimate predicted crash frequency. Predicted crash frequency is estimated using a calibration factor multiplied with the safety performance function (SPF), as documented in the HSM (HSM, 2010). National default values for safety performance functions can be obtained from the HSM for all site subtypes. Correction factors can be calculated as the ratio of the sum of observed crash counts from the target database to the sum of the predicted crash counts from the safety performance function. All the procedures mentioned above were implemented in ODI, and tables were created to store results in the target database.

3.2. Screening using Oracle R in the RPD

Network-wide corridor screening algorithms and network screening for individual sites algorithms were developed using Oracle R scripts. OBIEE use Oracle R scripts to execute the screening algorithms. These R scripts were saved to the database by using Oracle R Enterprise libraries. They can be executed with the ‘rqTableEval’ stored procedure. An R script that has a final data frame to return will output a standard Oracle database table when executed.

Two sets of R functions are saved in the Oracle database. One set of R scripts responsible for performing the network screening, getting results, and saving the results as a data frame to a datastore, which is a table accessible with the Oracle R Enterprise libraries that allows R variables to be saved to the database. The second set of scripts is responsible for loading the data frame from the datastore and returning the data frame. An Oracle SQL select statement can be used to execute

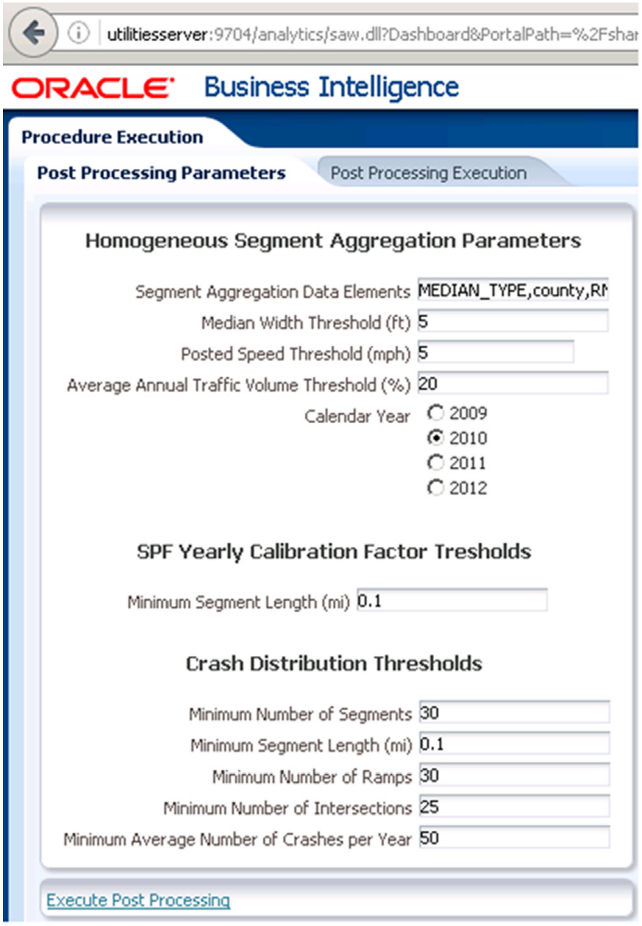

Fig. 4. Dashboard interface for post processing and calibration.

these R scripts. By saving the SQL select statement as a view and loading the view into the physical layer of the RPD, OBIEE is able to execute screening algorithms and load the results (Rittman, 2013).

The data required for network screening algorithms are accessed from the target database mentioned in Section 3.1. The view with the results is called a fact table and the target database tables are called dimension tables (Rittman, 2013). The star schema (Rittman, 2013) created in the physical layer of the RPD is illustrated in Fig. 5. These layers are brought into the business layer and the presentation layer for further analytics. The business layer performed joins, which helps mapping site locations in Fact table as well as crash, roadway, ramp and intersection characteristics in their respective dimension tables (Rittman, 2013).

4. Results

4.1. Data management

An interface was created using an OBIEE dashboard to execute the developed procedures for homogeneous segmentation of roadway segments. In Fig. 4 from Section 3.1, an input section for parameters used for homogeneous segment aggregation was shown, by which the user could enter aggregation data elements and threshold values for median width, posted speed, and AADT. Based on the parameters entered, the aggregation of roadway segments is performed. In addition, the user could perform calibration and crash distribution using the same post processing interface.

Minimum segment length for calibration and threshold inputs for crash distribution were provided in order to execute post processing, using the link, Execute Post Processing, as shown earlier in Fig. 4. Once the post processing was performed, the results were saved as database tables. Various post processed tables were created, including (1) homogeneously segmented datasets for roadway segments, (2) intersection and ramp dataset tables with associated site subtypes, (3) tables with calibrated factors for site subtypes, and (4) tables with crash distribution values for all crash types. Later, these tables were accessed by R scripts to perform network screening.

4.2. Corridor screening

A web interface was designed and implemented to run network-wide corridor screening on the fly, using OBIEE Presentation Services (AASHTO, 2010) by executing algorithm described in Section 2.2 using R-scripts. This web interface also included a dashboard prompt for parameter inputs, analytics for the presentation of performance measures and other related information as well as filters for specific values to activate dashboard prompts.

A dashboard with the dashboard prompt was created using the presentation variables for input parameters, as shown in Fig. 6. As a first step, a user has to select the screening algorithm; in this case network-wide corridor screening. Then, a section would be expanded with the dashboard prompt that has radio buttons to input crash severity variables (CSV), screening performance measures (SPM), type of screening (Type), and the limiting performance measure (XY threshold) for flagging sites for the selected algorithm. With this user input, the analyst can screen the network for various crash (collision) types or can select particular days of week or months. In addition, the name of the analysis can be provided, which enables multi-user analyses. With this functionality, various users can perform analyses and display results on the dashboard, based on the analysis name. Once the input parameters are entered, a user has to click ‘Apply’ to set a platform for the type of analysis. The ‘Run Network-wide Corridor Screening’ link enables running the analysis. The ‘View Network-wide Corridor Screening Results’ link provides a view of the results.

Analytics were created using columns from two tables, Fact and Dim-Roadway Segment. The columns used in this study to present the results were Route Name, Agency ID, Site ID, Window Begin, Window End, Expected Crash Frequency, Excess Crash Frequency, and Variance. Filters were created to filter ranks, and the name of the analysis. The dashboard was created using the various analysis objects, including tables, graphs, and maps. Drill down analysis for diagnostics of crash patterns can be performed on each of the corridors to the segments involved in the corridor. Both the fixed Corridor and Corridor Searching algorithms were implemented for network-wide corridor screening. Dashboards were prepared using the same concept to screen or search the corridors.

4.2.1. Results from Fixed corridor screening

For Fixed Corridor screening, corridors were predefined in the roadway segment dataset. For illustration purposes, approximately five miles of corridors were predefined and analyzed. An Esri map was created in the dashboard to display the spatial locations of fixed corridor results. Based on route, begin and end mile of predefined corridors, the geometry of segments within a corridor is displayed in the ESRI map. The results of top 10 fixed corridors that are displayed on the dashboard are shown Fig. 7. Results include the corridor ID, sites in the corridor, route, begin mile, end mile, the expected crash frequency, and the ranks of the corridor. In the results, the column ‘Sites in the Corridor’ includes sites that contain roadway segments, intersections and ramps in the corridor. Ranks of the fixed corridors are based on the expected crash frequency of the corridor.

The results obtained in this study were compared with the results using AASHTOWare Safety Analyst. The current literature for fixed-corridor screening uses crash frequency and rate methods, which also are used by this software. The literature states that when considering extended corridors for analyses, there is less variability or randomness

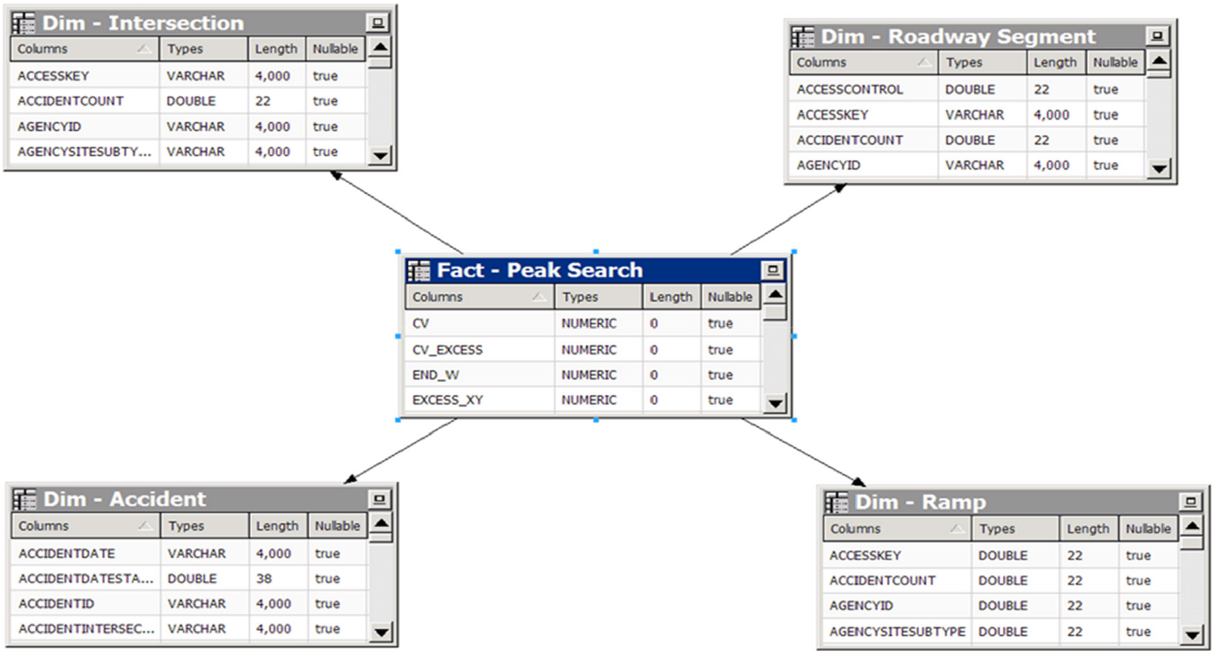

Fig. 5. Star schema for Peak Search network screening.

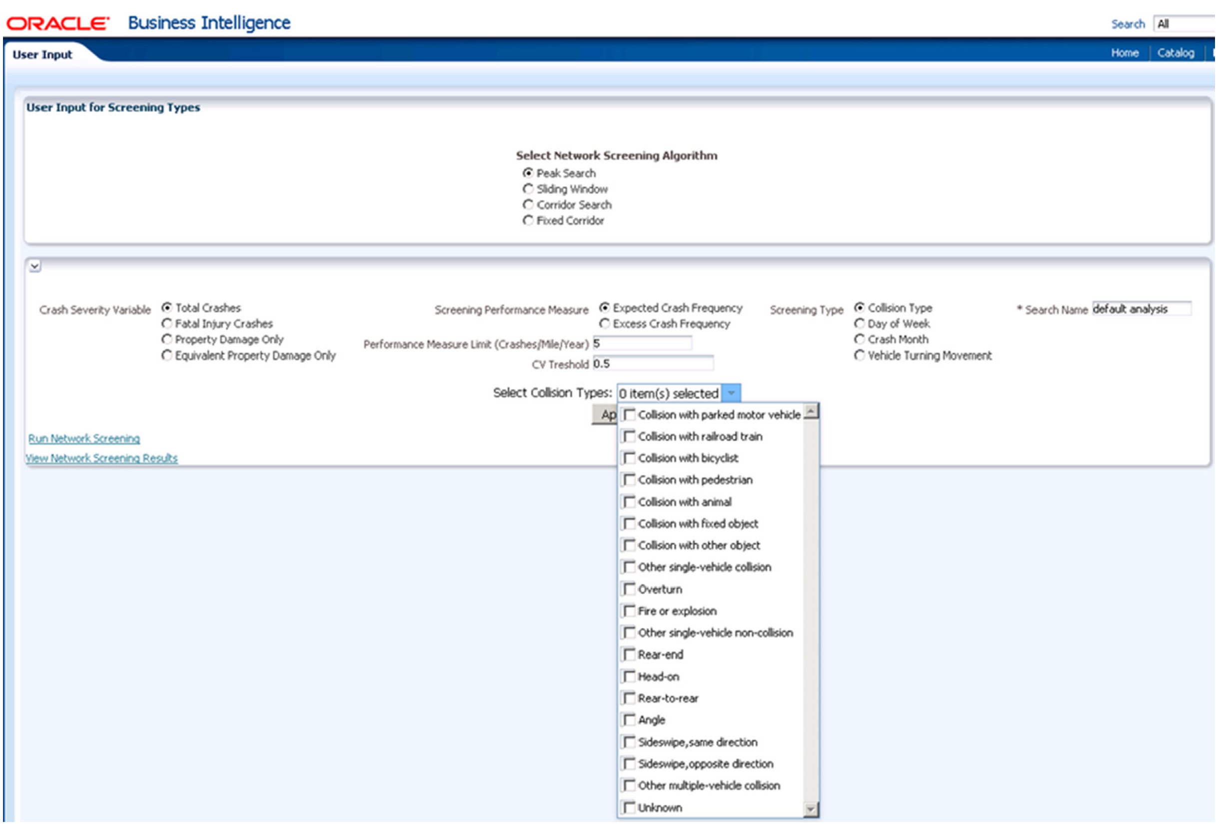

Fig. 6. Dashboard illustrating the user input interface for network screening.

in the crash data. Hence, frequency and rate methods provide accurate corridors for safety improvements. Fixed Corridor screening was analyzed using Safety Analyst for the same corridors used in this study.

Table 2 shows the top 15 ranked corridors obtained in this study, using the EB adjusted expected crash frequency, and the corresponding corridor ranks using the observed crash frequency and crash rate methods from Safety Analyst. The ranks obtained using all the three methods were different due to the use of safety performance functions and the advantages of EB-adjusted expected crash frequency that were incorporated in this approach. Table 2 shows that traditional methods, such as those using observed crash frequency and observed crash rates, gave ranks 21 and 4, respectively, to corridor route id 122385 from mile 7.715 to mile 12.73. However, with the improved EB expected crash frequency fixed corridor screening, the same corridor was ranked 1. Using observed crash frequencies results in a volume bias, while using crash rates may result in a segment length bias. In addition, using

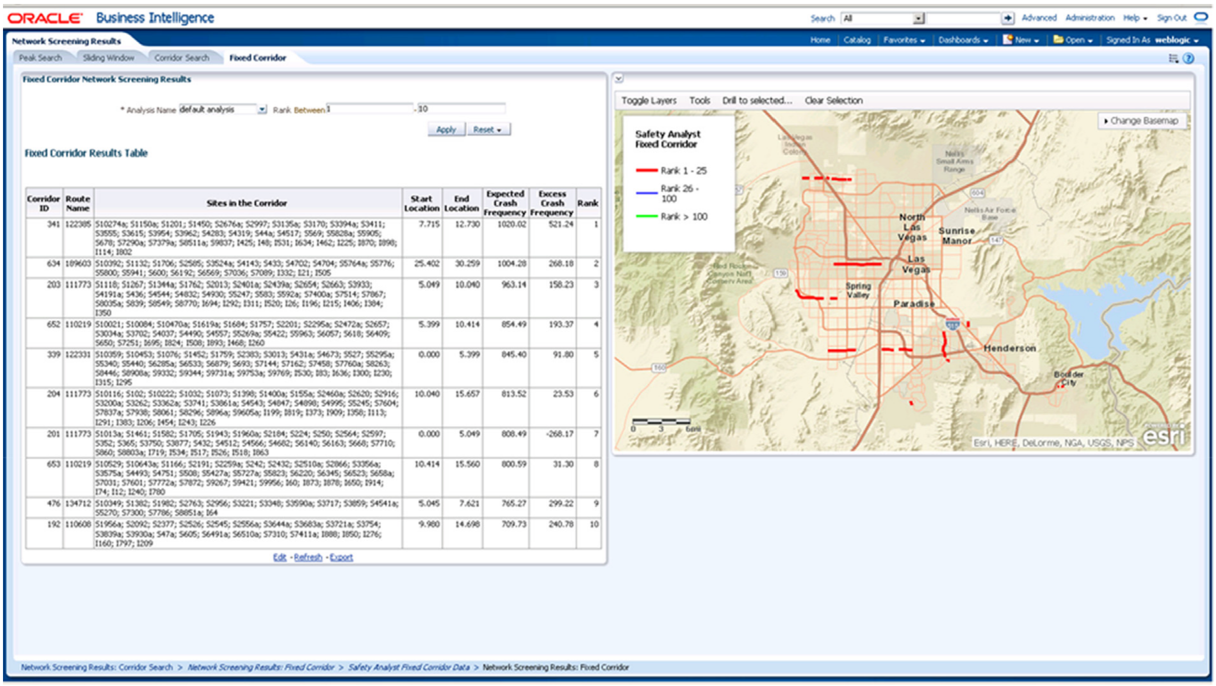

Fig. 7. Dashboard illustrating top 10 fixed corridor results.

Table 2
Ranking for the Top 15 Fixed Corridors using EB Expected Crash Frequency, Observed Crash Frequency, and Crash Rate Methods.

Corridor Route ID	Corridor Begin Mile	Corridor End Mile	Rank – EB Expected Crash Frequency	Rank – Observed Crash Frequency	Rank – Crash Rate
122385	7.715	12.73	1	14	4
189603	25.402	30.259	2	21	24
111773	5.049	10.04	3	10	13
110219	5.399	10.414	4	12	11
122331	0.000	5.399	5	19	18
111773	10.04	15.657	6	9	7
111773	0.000	5.049	7	16	22
110219	10.414	15.56	8	11	11
134712	5.045	7.621	9	8	14
110608	9.98	14.698	10	20	23
128	0.000	3.784	11	2	19
137611	5.067	7.924	12	4	8
129234	5.324	10.34	13	17	9
119961	0.472	5.569	14	15	3
113550	5.016	8.416	15	7	15

only observed crashes will result in a regression-to-the-mean (RTM) bias. EB expected crash frequency combines the observed crash frequency with the predicted crash frequency using the PF thus bringing the crash count towards the mean, accounting for the RTM bias. If limited budget is available, an engineer can consider a corridor for potential improvements where expected crash frequency is higher than predicted crash frequency rather than corridors ranking higher just based on crash frequency and crash rates.

4.2.2. Results from corridor search

For Corridor Searching, a 5-mi corridor length and 0.1-mi increment lengths to move windows were provided as input, along with other input parameters discussed in Section 3.2. Results are stored in the respective View tables, and analytics are presented using a dashboard. The results of corridor screening are presented in table and map format. Similar to the Peak Searching method, drill-down analysis can be performed on corridors to diagnose the crash patterns on the corridor. The Esri geometry for the Corridor Search algorithm requires Dynamic Segmentation (Cadkin, 2002) based on network screening results. Hence, a PL/SQL script was developed to segment the geometry dynamically. This script joined sections of roadway segments, based on routes. From the results of Corridor Search, the geometry of corridors were created using information regarding the route as well as the begin and end mile of corridors. For purposes of illustration, results from Corridor Search screening on a dashboard are shown in Fig. 8. Network screening using Peak Searching, Sliding Window, and Fixed Corridor algorithms do not require Dynamic Segmentation.

Table 3 shows the top 15 ranked corridors obtained in this study for the proposed EB-adjusted expected crash frequency and the corresponding corridor ranks, using the observed crash frequency and crash rate methods. As before, the ranks obtained using all the three methods were different due to the use of safety performance functions and the advantages of EB-adjusted expected crash frequency.

4.2.3. Results of network screening for Peak Search and Sliding Window

In the Peak Search algorithm, the roadway segment of interest is divided into windows of equivalent length that do not overlap; then, a performance measure of interest is calculated. A small window length of 0.1 mi is evaluated first and adjusted gradually for greater lengths. The coefficient of variation (CV), which is the ratio of the standard deviation to the mean of the expected value, is calculated for each segment. If the standard deviation is less than the mean of the expected or excess crash frequency (i.e., a small CV value), this indicates a high level of precision in the estimate. Thus, a smaller CV increases the user confidence level regarding the results, and vice versa (AASHTO, 2010).

In the Sliding Window algorithm, the user selects a pre-defined window length. The algorithm estimates the performance measure for the window, and then slides the window by incremental lengths to estimate the performance measures of the subsequent windows. All the windows are ranked with regard to the estimated performance measure.

In contrast to the Peak Search and Sliding Window methods, the Simple Ranking approach is used when considering roadway components, such as intersections or ramps, as a single entity. These components are ranked using the estimated performance measures. Details of the algorithms for all network-screening methods can be obtained from

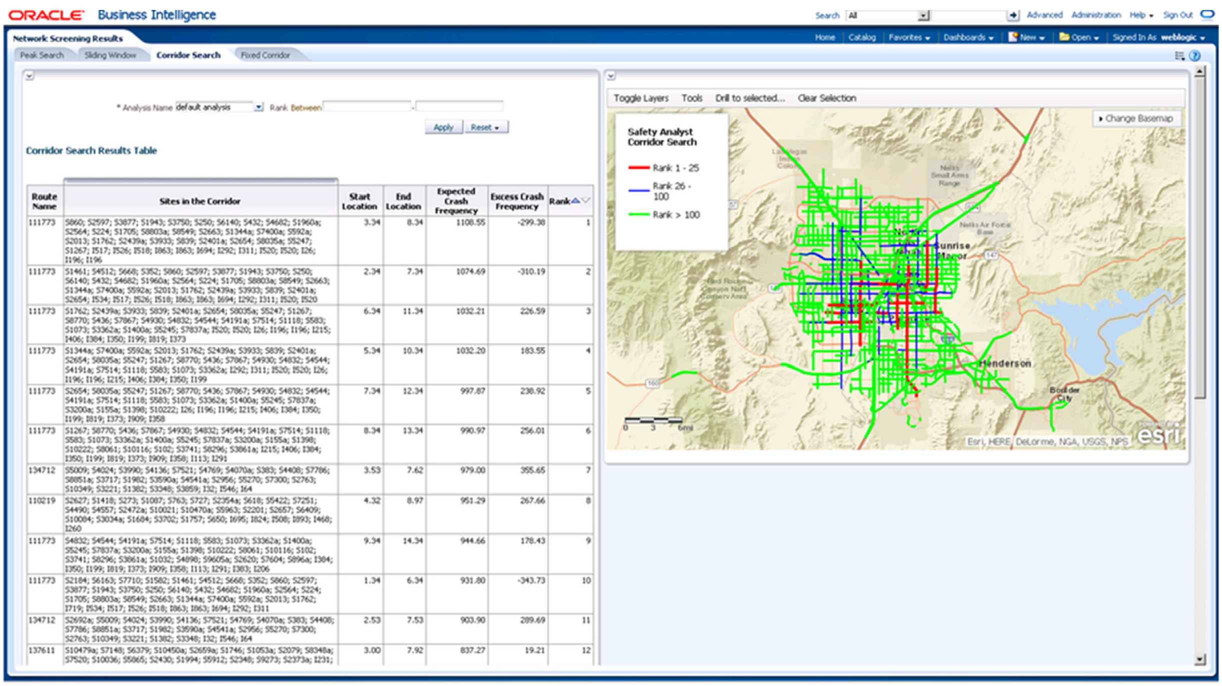

Fig. 8. Dashboard illustrating results for a Corridor Search.

Table 3
Ranking for the Top 15 Corridor Search using EB Expected Crash Frequency, Observed Crash Frequency and Crash Rate Methods.

Corridor Route ID	Corridor Begin Mile	Corridor End Mile	Rank – EB Expected Crash Frequency	Rank – Observed Crash Frequency	Rank – Crash Rate
111773	3.339	8.339	1	9	18
111773	2.339	7.339	2	10	19
111773	6.339	11.339	3	14	13
111773	5.339	10.339	4	17	12
111773	7.339	12.339	5	18	17
111773	8.339	13.339	6	19	16
134712	3.525	7.621	7	4	8
110219	4.322	8.974	8	11	23
111773	9.339	14.339	9	19	20
111773	1.339	6.339	10	13	21
134712	2.525	7.525	11	8	11
137611	3.000	7.924	12	5	4
110219	9.213	13.364	13	12	3
110608	8.976	13.016	14	20	9
122385	7.715	11.627	15	3	6

Part B of the HSM (AASHTO, 2010).

To run network screening for individual sites on the fly, a web interface similar to the one for network-wide corridor screening was designed and implemented. The only difference in this case, a user has to select – the network screening algorithms – Peak Search, Sliding Window or Simple ranking.

Analytics were created using columns from two tables, Fact-Peak Search and Dim-Roadway Segment for Peak Search. The columns used in this study to present the results were Route Name, Agency ID, Site ID, Window Begin, Window End, Expected Crash Frequency, Excess Crash Frequency, and Variance. Filters were created for ranks and the name of the analysis. A dashboard was created using the various analysis objects, including tables, graphs, and maps.

Users have an option of selecting an analysis name with a drop-down menu as well as ranks, using the analysis prompt in the analysis section. An Esri map was included in the dashboard to present the spatial location of roadway segments, which are color-coded based on their ranks. Selecting the ranks in the analysis would filter the segments in the map. Fig. 9 shows a snapshot of a Peak Search analysis using data from Nevada.

Users can drill down further on the segment to diagnose a high-crash location for detailed characteristics of crashes and roadway segments. These characteristics provide the crash pattern in the site location, such as a high number of night-time rear-end crashes. In addition, the user can turn on a Google Earth satellite image for further site information. This information may provide insights to the user to determine countermeasures that could mitigate crashes.

A snapshot of a drill-down analysis is shown in Fig. 10. The figure shows a description of crash severity and crash (collision) types for a top-ranked roadway segment from a screening analysis. Distributions for light conditions, crash time of day, day of week, number of vehicles, vehicle types involved, and weather condition can be created and displayed in the same drill-down analysis. These distributions provide a clear picture in order to select the type of countermeasure to mitigate future crashes. The user can export the analysis to a portable document format, Microsoft Excel, or PowerPoint by using the export or print tools inherent in OBIEE. This information can be disseminated to decision makers by means of email.

Similar to Peak Search, the Sliding Window dashboard was designed and implemented to run analyses and display results; the only difference is the input parameters. The results of the analysis are stored in the View, as discussed in Section 3.2. Analytics created using the table, Sliding Window View, display the results for expected and excess crash frequencies in the dashboard, using tables and maps. The drill-down analysis was created to diagnose characteristics for crashes, roadways, and traffic at high-crash locations.

5. Conclusions and recommendations

This research proposed a framework for corridor-level network-wide screening for traffic safety analysis. From the perspective of traffic safety engineers, network screening is of significant importance to meet the requirements of a HSIP. The proposed framework is motivated by the need to provide homogenous conditions, as much as possible, throughout the network to avoid confusion and meet drivers' expectations. The existing state-of-the-art traffic safety literature lacks a

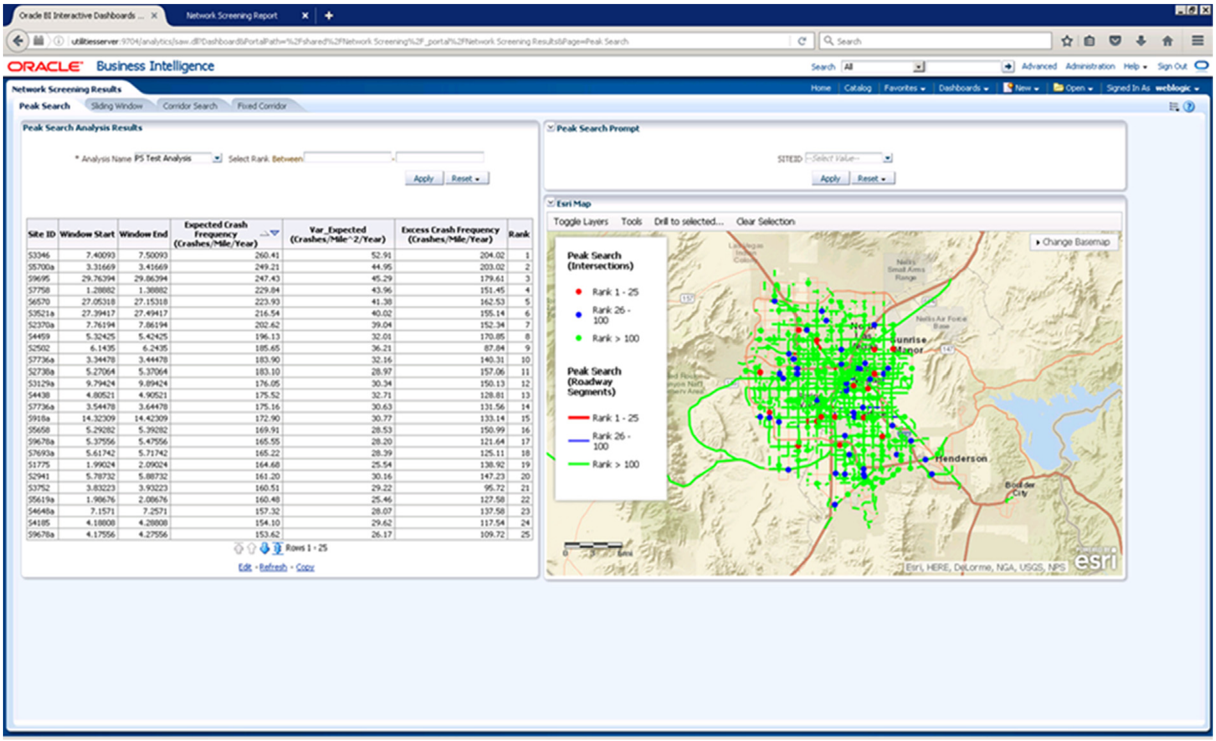

Fig. 9. Dashboard illustrating results and visualization from Peak Search network screening.

comprehensive framework that enables this type of corridor-level network screening. The propose framework includes algorithms based on Fixed and Corridor Search methods. In contrast to the existing literature, expected crash frequencies were used instead of observed crash frequencies or rates for selecting corridors with the potential for safety improvements. Top-ranked corridors obtained using the proposed methodology for corridor screening were compared with ranked corridors based on rate and frequency methods. The order of ranks of the corridors are completely different as a consequence of using a sound approach.

The implementation approach aimed to provide a framework to facilitate the broad adoption of adequate methodologies for network-wide corridor and network screening for traffic safety analysis. Traditionally, separate tools are used (1) to integrate, process, and manage the data; (2) for modeling analysis; and (3) to visualize the results. This traditional approach may result in data replication, and it requires substantial technical knowledge as well as being time consuming. Hence, analysts choose easy-to-implement legacy methodologies, which may lead to identifying incorrectly those sites with safety needs, thus resulting in inefficient roadway-safety management.

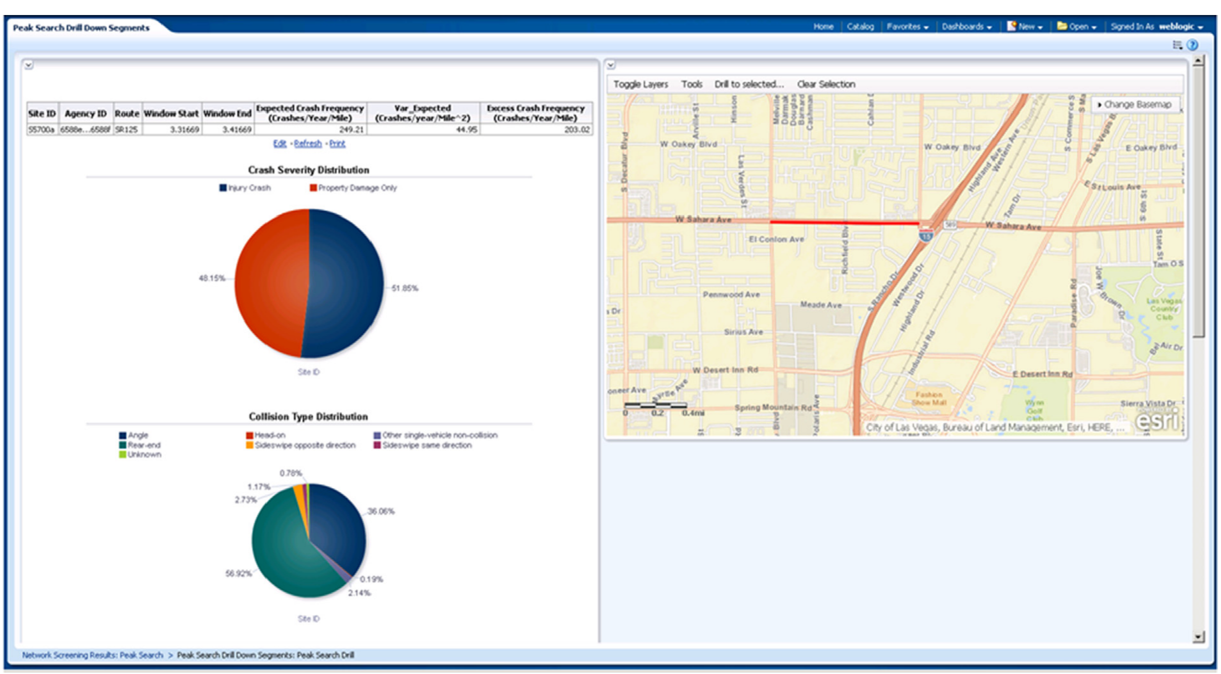

Fig. 10. Dashboard illustrating drill-down results of results from a roadway segment.

In this research, a BI approach is proposed to address barriers associated with data integration, management, and visualization for the implementation of sound methodologies similar to those in the HSM. The outcome is a single framework that accesses the data from the source, integrates and manages the data, processes analytical models, and provides the results by means of a web-based interface. To illustrate the advantages of the proposed framework, network screening algorithms from the HSM were implemented and expanded. Results were presented by using dashboards that included maps, filters, and drill downs. Results of network screening produced by this framework were verified by using Safety Analyst and from outcomes by Paz et al. (2015).

Advantages of using the proposed framework include the following benefits.

(1) It has the capability to perform corridor-level network screening using a sound approach.
(2) It provides data integration, analysis, and visualization capabilities.
(3) When new data is loaded into the source, it is automatically loaded into the warehouse, using an ELT process.
(4) It has better visualization capabilities than existing methods (Tarko, et al. 2014; Paz, et al. 2014; AASHTO, 2014).
(5) Development cost and time are minimized.
(6) Required training and maintenance are minimized.
(7) It uses a web-based approach for development and use.

Future work includes automation of dynamic geometry generation for Esri maps for corridor screening. The other three steps of a roadway safety management process including diagnosis and countermeasure selection, economic analysis and priority ranking, and countermeasure evaluation also need to be incorporated within the implementation framework.

Desirable additional capabilities within the proposed framework include methods and tools to (1) estimate PFs using local data and techniques to address unobserved heterogeneity, temporal and spatial instability, and self-selectivity issues (Mannering et al., 2016); (2) analyze for diagnosis, countermeasure selection, economic analysis and priority ranking, and countermeasure evaluation to complete safety management process; and (3) perform regional-level forecasting of crash trends. The proposed framework relies on the availability of PFs. In addition to the outcomes from a standard safety management process, decision makers are required to provide system-wide forecasts and associated targets for long-term planning.

Acknowledgments

Many thanks to the Nevada Department of Transportation for providing the data that was used for this research and to our Technical Writer at UNLV's Howard R. Hughes College of Engineering, Mrs. Julie Longo, for her help reviewing this manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssci.2019.08.042.

References

AASHTOWare Safety Analyst v4.3.3. (2014), Safety Analyst Newsletter, July, http://www.aashtoware.org/Safety/Pages/Safety-Analyst-Newsletter.aspx.
American Association of State Highway and Transportation Officials, "Highway Safety Manual," American Association of State Highway and Transportation Officials, 2010.
Alluri, P., Ogle, J., 2012. Road safety analysis, in the United States. Transport. Res. Record: J. Transport. Res. Board 2318, 7–15.
ArcGIS Web API. (2015), ArcGIS API for Javascript. Retrieved from https://developers.arcgis.com/javascript/latest/api-reference/index.html.
Aylo, R. 2010, Wave propagation in negative index materials, Electronic Thesis or Dissertation, University of Dayton.
Cadkin, J. (2002). Understanding dynamic segmentation-working with events in ArcGIS 8.2, ESRI. ArcUser October-December. Retrieved from http://www.esri.com/news/arcuser/1002/files/dynseg_2.pdf.
Center for Advanced Public Safety (CAPS), CARE – Critical Analysis Reporting Environment, 2009, http://caps.ua.edu/online_analysis.aspx.
Dupupet, C., Gray, D., Boyd-Bowman, P.C., Testut, J., 2013. Oracle Data Integrator 11g Cookbook. Packt Publishing, UK.
Federal Highway Administration, Safety Provisions in Moving Ahead for Progress in the 21st Century (MAP-21), 2013, http://www.fhwa.dot.gov/map21/safety_overview.cfm.
Federal Highway Administration, (2015), Highway safety improvement program reports, U.S. Department of Transportation.
Hamidi, A., Fontaine, M.D., Demetsky, M.J. (2010), A planning-level methodology for identifying high-crash sections of Virginia's primary system, Virginia Transportation Research Council – Research Report, no. FHWA/VTRC11-R4.
ITT Corporation, 2011. Safety Analyst User's Manual. Springs CO., Colorado.
Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Anal. Methods Accident Res. 17, 1–13.
Mannering, F., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accident Res. 11, 1–16.
McDermid, D., Taft, M. (2014), Oracle R Enterprise User's Guide, Release 1.4. Oracle.
Montella, A., 2010. A comparative analysis of hotspot identification methods. Accid. Anal. Prev. 42, 571–581.
Nemmers, C.J., Vap, D., McDonald, T.J. (2008), Effectiveness of safety corridor programs, report on tasks 1–3, Midwest Transportation Consortium, no. MTC Project 2007-08.
Paz, A., Khanal, I., Veeramisti, N., Baker, J., Belmonte, L., 2014. Development of a visualization system for Safety Analyst. Transport. Res. Record: J. Transport. Res. Board 2460, 176–185.
Paz, A., Veeramisti, N., Khanal, I., Baker, J., de la Fuente-Mella, 2015. Development of a comprehensive database system for Safety Analyst. Sci. World J.
Pulugurtha, S.S., Krishnakumar, V.K., Nambisan, S.S., 2007. New methods to identify and rank high pedestrian crash zones: an illustration. J. Accid. Anal. Prevent. 39 (4), 800–811.
Qin, X., Sultana, M., Chitturi, M., Noyce, D., 2013. Developing a truck corridor crash severity index. Transport. Res. Record: J. Transport. Res. Board 2386, 103–111.
Rittman, M., 2013. Oracle Business Intelligence 11g Developer's Guide. The McGraw-Hill Companies, USA.
Shimko, G., Walbaum, C., 2010. City of Edmonton traffic safety strategy. Road Safety Session of the 2010 Annual Conference of the Transportation Association of Canada, Halifax, Nova Scotia.
Tarko, A.P., Li, M., Romero, M., J. Thomaz, J. (2014). A systematic approach to identifying traffic safety needs and intervention programs for Indiana: Volume I – Research report, in Joint Transportation Res3earch Program Publication, no. FHWA/IN/JTRP-2014/03.
Tjandra, S.A., 2014. Business Intelligence system for traffic data integration: linking roadway, collision and traffic flow data to improve traffic safety. NATMEC 2014 Improving Traffic Data Collection, Analysis, and Use, Chicago, Illinois.
Wellner, A., Qin, X., 2011. GIS-based highway safety metrics implementation and evaluation. Transport. Res. Record: J. Transport. Res. Board 2241, 1–9.
Xie, G., Hoeft, B., 2014. Freeway and arterial system of transportation dashboard: Web-based freeway and arterial performance measurement system. Transport. Res. Record: J. Transport. Res. Board 2271, 23–34.
Zhao, Y., Tian, Z.Z., Reider, C., 2017. Screening urban Road Networks for corridors with Promise. 94th Transportation Research Board Annual Meeting, Washington D.C.