Memorandum

U.S. Department
of Transportation

**Federal Highway
Administration**

6300 Georgetown Pike
McLean, Virginia 22101

Subject: **ACTION**: LTPP Directive I-85
Guidelines for Manual Upgrades to LTPP IMS Quality Control Checks

Date: March 24, 2003

From: Eric Weaver
Long Term Pavement Performance Team

Reply to
Attn of: HRDI-13

To: Dr. Frank Meyer, PM - LTPP North Atlantic Regional Contract
Mr. Tom Wilson, PM - LTPP North Central Regional Contract
Mr. Mark Gardner, PM - LTPP Southern Regional Contract
Dr. Sirous Alavi, PM - LTPP Western Regional Contract

Attached is the Long Term Pavement Performance (LTPP) Program IMS Directive I-85: Guidelines for Manual Upgrades to LTPP IMS Quality Control Checks. The directive addresses treatment of the results of the automated QC checks on data after entry into the IMS. Please ensure that all personnel involved with the IMS are aware of this new directive.

Should you have any questions or would like to discuss this directive, please do not hesitate to contact me at 202-493-3153.

Attachment

LONG TERM PAVEMENT PERFORMANCE PROGRAM DIRECTIVE

For the Technical Direction of the LTPP Program

Program Area:	IMS	Directive Number:	I-85
Date:	March 24, 2003	Supersedes:	NA
Subject:	Guidelines for Manual Upgrades to LTPP IMS Quality Control Checks		

Background

The Quality Control (QC) checks in the LTPP Information Management System (IMS) were developed to address a variety of issues related to data quality and data storage within a distributed-relational database structure. Quality control checks of the data are conducted at every step in the data collection and processing process. This directive addresses treatment of the results of the automated QC checks on data after entry into the IMS.

The IMS QC checks are an automated series of data checks. Levels using a hierarchical alpha code starting with A and ending at E designate the checks. The level designations are described below.

Level	Description
A	Starting point. When records are first input into the IMS they are assigned a status of A. Records failing the level B or level C checks will have a status of A (or B, if a B level check exists). At present, data for SPS supplemental test sections, which by policy are not subjected to QC checks, are left at level A in most tables.
B	An old check that is being replaced in some modules. Originally, level B was a dependency check on the availability of certain critical data contained in other tables. In some modules, this check has been phased out and replaced with level E checks and changes to the structure of the EXPERIMENT_SECTION table. There are cases where records with RECORD_STATUS=B exist due to restrictions imposed by the software used to perform manual upgrades.
C	Availability of critical data fields in a record. These are checks to see if certain data fields have non-null values. As an example, test section coordinates are required for all entries in INV_ID and SPS_ID. Some of the level C checks are conditional checks on several fields. For example, in MON_DEFL_DROP_DATA, of the 7 to 9 possible deflection values, at least 5 must be non-null. These checks are not performed on key

Level	Description
	fields and fields defined as non-null, since these fields must be populated in order to create a record.
D	Range checks on the values contained in single fields. While these are called expanded range checks, they are refined range checks on the reasonableness of the magnitude of a number or code value. When data is entered, its range must match the field format logic. For example, a value of 999 can not be entered in a field defined as NUMBER(2,0). These checks are more stringent than logical range values, but in some instances are set to a rather large range of values to encompass typical conditions. For example, the range of air temperature must accommodate conditions spanning from Arizona to Alaska. In other instances, the range limits are based on traditional practice in order to flag outliers and suspect values. For example, the percent longitudinal reinforcement in PCC pavements is limited to an upper value of 1% since it is very rare that pavements are built with even this very high level of steel reinforcement.
E	Relational checks between data elements in the same record and data elements contained in other records. Although previously described as intra-modular checks, these checks have been expanded to include record level inter-field and inter-modular checks. Some of the types of level E checks include: • Logical relationships between related values. For example, a minimum value must be less than or equal to the average, which must be less than or equal to the maximum. • Parent-child integrity checks. For example, every record in MON_DEFL_LOC_INFO must have a matching record in MON_DEFL_MASTER. • Range checks between related values. For example, the difference between the daily maximum and minimum air temperature must be less than 50° C. • Referential cascading parent-child level E relationships. For example, for records in MON_T_PROF_MASTER to reach level E, all matching records in MON_T_PROF_PROFILE must be at level E. • Compliance with LTPP rules and test protocols. Many level E checks are based upon LTPP rules for pavement-structure-material layer types, sequence and LTPP test protocols. For example, the surface layer of a GPS-3 test section should consist of portland cement concrete. • Computed parameter referential level E checks on records in source tables. For example, for records that contain results of FWD backcalculation computations to reach level E, matching data from the FWD deflection tables must also be at E.

The QC status of a record is stored in a field named RECORD_STATUS in most of the tables in the IMS. This field typically contains a letter A, B, C, D or E, indicating the QC level for that record. When a record is first created the RECORD_STATUS field is set to A. The QC checks are performed using computer programs that operate on groups of related tables. When a record with RECORD_STATUS=A passes all level C checks, the QC program will set RECORD_STATUS equal to C (or B, if B level checks exist). This process continues until the record passes all level E checks, then RECORD_STATUS is set

to E. Thus, a RECORD_STATUS = E implies that a record has passed all QC checks. This is often called "level E" data.

The exceptions to population of RECORD_STATUS with a letter are in the EXPERIMENT_SECTION and TRF tables. In EXPERIMENT_SECTION, a * indicates that the record is for a SPS supplemental test section, and a ! indicates that the test section has been released. In the TRF tables, ! is used to indicate records that have been superseded with a newer record with a higher modification number.

Because some QC checks were designed to flag potential data errors, failing an automated check does not mean that the data is necessarily bad or erroneous. Some checks are intended as alerts that data is unusual and should be inspected. In order to provide a mechanism to upgrade a record that does not contain errant or improper data, a manual upgrade procedure was created, which is the process of moving a record from a lower level QC to the next higher level. An external IMS program called BROWSER provides a semi-automated process for entering comments on why records were or were not upgraded and to optionally create SQL statements to perform the QC upgrades.

General Guidance on Manual Upgrades

Manual upgrades should be regarded as the last course of action for failing records. Records failing a QC check should always be investigated, and if possible, changes or corrections should be made to data entries to correct the error (if an error actually exists). There are situations when a record failing an automated QC check should not be manually upgraded; it should be left at a lower QC level. In some cases, records failing QC checks should be deleted from the IMS. In other cases, the QC checks need to be changed because of a programmatic policy change, programming blunders, errors or oversights.

The "IMS Quality Control Checks" document, commonly referred to as the "QC Manual" is the primary source of information on what QC checks are performed and when manual upgrades are not appropriate. The QC Manual defines all of the automated checks through which data are evaluated for existence, value, and relationships with other data elements. These checks serve as the first reference in the decision to accept the data since they indicate the normal expected outcome of the LTPP data collection process. The value checks represent a range of reasonable values, not a "mean and three standard deviations" type of outlier detection process. As an example, the area of fatigue cracking can be zero, but it can never be larger than the total area of the test section.

The relationship checks generally apply physical laws, mathematical properties or relational database design requirements. Examples of relationship checks include the following: asphalt resilient modulus cannot increase with increasing temperature; an average must be between the minimum and the maximum inclusive; and for a valid backcalculation result to exist, the layer structure used for the calculation must match the layering from materials testing.

If a check on a data item is not shown in the QC Manual, then there may be no check on that field. As the QC Manual is revised, more explicit statements will be added on which tables, rows, and columns can or cannot be considered for manual upgrade and why.

QC Error Resolution

When a record does not pass a QC check, the data in the record and related records should be examined. Some types of possible errors that can be corrected include:

- Transcription errors.
  Transcription errors are an inherent problem with any manual data entry system. When a record fails a QC check, this should be one of the first errors investigated. While it may not be possible to double-check all data entered, it is recommended that approximately 10% be double-checked to expose possible patterns in error occurrences. If a pattern is recognized, then this may aid in finding the source of the errors, which may involve a 100% review.

- Improper referential data entry in another record.
  Because LTPP data are obtained from multiple sources, it is possible that a field used for referential links between tables will not have been properly recorded. LAYER_NO is a prime example of this type of correctable problem. There are times when a material testing laboratory may assign a LAYER_NO that is later changed in the IMS due to factors unknown to the laboratory contractor. This can cause a mismatch of material types in the layer tables. This type of error can be easily corrected by assigning the correct LAYER_NO in the mismatched record. Other types of examples include improper entries in CN_ASSIGN_DATE in EXPERIMENT_SECTION, missing or improper entries in *_LINK tables, and missing records in TST_HOLE_LOG and TST_SAMPLE_LOG. Notes should be made on the original data sheets to reflect the same information as record in the database in cases where layer designations are changed over time.

- Improper data acquisition or interpretation.
  In some cases, the supplier of the data may not have understood the intent or basis for the needed data element. These types of errors are usually associated with level-D range checks. In these cases, the only recourse is to contact the data supplier and search for the correct value. For example, the percentage of longitudinal reinforcement steel in PCC pavements should never exceed 1%. When an agency has reported numbers in excess of this value, Regional Support Contractor staff should discuss the issue with agency contacts to decide if the correct value can be determined from the available records. In some cases, it may also be possible to resolve issues using photographs or direct field measurements.

- Errors, oversights, and blunders with interpreted data.
  There are instances where it is possible to reinterpret data from the raw measurements. An error in distress data acquired via photographic-based measurements is an example of a potentially correctable error since the photographs

can be re-interpreted. When errors or problems are discovered in transverse profile or distress measurements, the apparent errors should be referred to the data collector for possible correction.

- Potentially rectifiable data.
  Longitudinal and transverse profile data provide opportunities where erroneous data in the IMS might be rectified. For longitudinal profile data, other measurement runs on a section may be available to replace runs containing spikes or other apparent equipment errors with other runs performed on the same day that do not contain such errors. Alternatively, on SPS projects, sub-sectioning of the raw data files can be corrected for apparent DMI drift. Manually collected transverse profile data in which the measurement width varied along a section may be able to be salvaged with reinterpretation of the raw data.

- Conflicts between two or more linked fields in a record.
  For example, if an admixture amount is provided, then the corresponding admixture type code must not be null or "no admixture."

During the QC error resolution process, it is also important to identify errors that are not possible to rectify. Some examples include:

- Equipment measurement errors.
  A record failing a QC check due to an identifiable equipment measurement error should not be manually upgraded. Instead, the errant data element(s) should be deleted from the IMS. In records with multiple measurement fields, the "bad" data element should be set to null. In cases where *all* of the measurement data elements in a record are affected by an equipment malfunction, the record should be deleted. Where a record contains multiple measurements from different sensors and the null (error) data columns cause it to fail a QC check, manual upgrade may be appropriate.

- Required data not available.
  There are instances when a required data element was not collected or was collected improperly and it is no longer possible to obtain or measure the data. This can lead to a test section being removed from the LTPP study, taken out of study, or recognition that it is not possible to obtain the required data element. If it is recognized that the required data element unattainable then corrective action could include comment without upgrade, or comment and upgrade, depending on significance of the related data element.

- Indeterminable problem that requires investigation.
  When new tables are added to the IMS or new QC programs are issued, some records may fail a QC check for no discernable reason (e.g. equipment failure). In these cases manual upgrades should not be performed until the exact cause for the problem can be determined. If the problem is related to a programming or table error, a Software Performance Report (SPR) should be prepared, which will document the problem and steps can then be taken toward resolution.

The following are examples of errors identified by the QC programs that should not be manually upgraded:

- Any error in the EXPERIMENT_SECTION table.

Manual upgrades to RECORD_STATUS in the EXPERIMENT_SECTION table are prohibited. Due to the importance of this table in the structure of the IMS, EXPERIMENT_SECTION records failing the automated QC checks can only be corrected by a change in a data element within the record or in a related record in another table. In some cases, a change to an error discovered in the QC program is needed. In any case, manual upgrades to RECORD_STATUS in this table are futile; the QC program always resets RECORD_STATUS to "A" every time it is run.

- Referential level E checks between computed parameters and their source data tables.

  Many of the present computed parameters are calculated externally to the IMS; the data are extracted, the computations performed, and the results of the computations are loaded into the IMS. Referential level E checks are used to ensure that the original source data still exist and are at level E. When a record in a computed parameter table fails a level E check because of this type of referential check, the record in the computed parameter table should not be manually upgraded. If the referential record status cannot reach level E, then the computations based on this data should also not be allowed to reach level E. Some of these type of computed parameter table relationships include:

  - MON_DEFL_FLX_BAKCAL_BASIN ⇒ MON_DEFL_DROP_DATA
  - MON_DEFL_RGD_BAKCAL_BASIN ⇒ MON_DEFL_DROP_DATA
  - MON_DEFL_*_BASIN ⇒ TST_LO5*
  - SMP_TDR_AUTO_DIELCTRIC ⇒ SMP_TDR_AUTO
  - SMP_TDR_MOISTURE_SUPPORT ⇒ TST_SS01_UG01_UG02
  - SMP_TDR_MOISTURE_SUPPORT ⇒ TST_UG04_SS03

- Parent-child table existence and level E referential checks.

  In groups of tables structured with a parent-to-child, or one-to-many record relationship, records in the parent or child table failing a check on the existence of a matching record in the related table, or check for a matching record at level E, should not be manually upgraded. For example, a record in MON_DEFL_TEMP_DEPTHS should not be manually upgraded to E if no matching record exists in MON_DEFL_MASTER. In this case, the record in MON_DEFL_TEMP_DEPTHS is considered an orphan record. Orphan records should be deleted. Likewise, records in MON_DEFL_MASTER should not be manually upgraded if they fail the test on the number of records in MON_DEFL_LOC_INFO with RECORD_STATUS=E. Some of the parent-child tables in the IMS include:

  - MON_PROFILE_MASTER – MON_PROFILE_DATA
  - MON_T_PROF_MASTER – MON_T_PROF_DEV_CONFIG

- MON_T_PROF_MASTER – MON_T_PROF_PROFILE
- AWS_PRECIPITATION_MONTH – AWS_DAILY_DATA
- MON_DEFL_MASTER – MON_DEFL_TEMP_DEPTHS
- MON_DEFL_TEMP_DEPTHS – MON_DEFL_TEMP_VALUES
- MON_DEFL_MASTER – MON_DEFL_LOC_INFO
- MON_DEFL_LOC_INFO – MON_DEFL_DEVICE_CONFIG
- MON_DEFL_LOC_INFO – MON_DEFL_DROP_DATA
- MON_DEFL_DEVICE_CONFIG – MON_DEFL_DEVICE_SENSORS
- MON_DEFL_FLX_BAKCAL_SECT – MON_DEFL_FLX_BAKCAL_POINT

- Internal computed parameters.
  New computed parameters procedures under development will perform computations and create records in computed parameter tables regardless of the RECORD_STATUS in the source data tables. In these instances, a level E check is always applied after the fact to ensure that computations performed using non-level E data do not make it to level E. This is an interim measure to provide non-level E computed parameters for supplemental test sections. Data from supplemental sections are released for many high profile national analysis projects. In the future, QC programs will be modified to check data for SPS supplemental test sections, however at present, these types of records should not be manually upgraded.

- Record in a table that contains only key fields and no "data" elements.
  For example, records in the estimated traffic tables that do not contain any traffic estimates are of no use in performance analysis and are candidates for deletion from the IMS. Many level C checks were designed to catch this circumstance.

The primary concern in the manual upgrade decision are those cases that fall into a "gray area" -- a problem record that is not correctable but might be useful to an analyst. The test that should be used to resolve these types of issues is:

Does the data in a record provide data points that can be used in mechanistically based, probabilistic, pavement performance analysis?

The following definitions can be used to aid in evaluating each problem record according to the intent of this test question:

- "Mechanistic-based" means that the data provides traffic-load, climate, structural, and material characterization information that can be used to model a pavement's response to load and climate effects.
- "Probabilistic" means that the data provides an estimate of the uncertainty or variability associated with variables used in pavement performance analysis.
- The objective of pavement performance analysis is to relate measured pavement condition and the response of dependent variables, to the factors that cause changes in pavement condition — the independent variables.

Examples of problem records include:

- A record with only an explanation of why no useful data points were collected.
  The test is on the usefulness of the data in pavement performance modeling, not on the data collection contractor, laboratory contractor, or highway agency performance. The goal of LTPP operations is to build a research-quality database that supports many important types of pavement performance analyses. The preferred practice is to not include records in the database that only provide a reason why particular data items are not available. In many instances of this type, manual upgrades are not preferred. This type of information is more appropriate for the Ancillary Information Management System (AIMS), not the production research database. There are some exceptions in computed parameter tables, where such types of record serve a useful purpose. For example an non-interpretable code exists in SMP_TDR_MANUAL_DIELECTRIC.

- The data violates a LTPP "rule" but represents ground truth.
  Many level D and E QC checks evaluate compliance with LTPP rules for acceptable pavement structures, data measurement protocol, material types, etc. Where the verified ground truth data violates LTPP rules, but truthfully represents actual conditions, manual upgrades are often justified.

- A required associated data element was not collected.
  Occasionally, one of the required data elements associated with primary data is not measured or recorded. For example, the FWD infrared pavement-surface-temperature sensor malfunctioned or the manual pavement layer temperature gradient data was lost. In these circumstances, the failing record should only be upgraded after it has been confirmed that the associated data is truly missing and cannot be obtained.

- Although LTPP protocols were not followed, the data may still be useful in analysis.
  This can be a very difficult judgment decision when applied to laboratory test data supplied by state or non-LTPP contractor laboratories. The primary concern in this situation is the potential for improper interpretation. It must be remembered that data users combine test results in the same field into a suite of data for analysis. If a result does not represent the same physical quantity as the other data points in the same field, then the data point should not be included in the database. These types of data are candidates for inclusion in the AIMS.

Manual Downgrades

The QC checks are not perfect. Many of the checks are rudimentary and are not capable of identifying all types of errant or "bad" data. In some instances, there is a need to add new QC checks to the system to catch errant data. There are instances in which it is impossible to create an automatic check that can reasonably catch some types of anomalies. For example, consistency checks between related time series data have not been implemented within the Level E QC system. Since the objective of LTPP operations is to provide research quality data to support pavement performance-related data analysis objectives, it is also a

responsibility of the RSCs (Regional Support Contractors) to verify the integrity of data that passes automated QC checks. In those instances when errors are discovered in level E data, RSCs have the responsibility to notify FHWA and perform manual downgrades as appropriate.

To perform manual downgrades, the RSCs shall develop SQL scripts to reset the RECORD_STATUS to an appropriate level. The scripts shall be submitted to the TSSC for approval. Prior to submission of data for NIMS upload, the downgrade scripts shall be run by the RSCs. The TSSC shall consolidate the download scripts from all regions for use in checking the consistency of their application after an upload.

The appropriate levels for manual downgrades are:

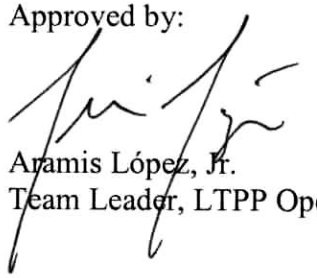
Level	Description
A	Entry(s) for important fields are missing, i.e. the record passes a level C check when important data fields not included in the level C check are missing or null.
C	The record passes a level D range check, but the magnitude of a value is suspect.
D	A record passes level E checks, but has relational inconsistencies with data in another table or field that could potentially be misinterpreted by data users.

Summary

The objective of the LTPP IMS database is to support engineering analysis of the performance of highways in North America. LTPP has created a system of checks in order to provide an indication of data quality to data users. There are situations when data failing automated QC checks should be upgraded and situations when data should not be upgraded. If there is doubt concerning an upgrade decision, RSC staff should always contact FHWA and TSSC to obtain advice and consent. Equally important to the manual upgrade decision is the manual downgrade responsibility when errant data at level E are identified.

Prepared by: TSSC

Approved by:



Aramis López, Jr.
Team Leader, LTPP Operations