

Data Management for Toxicological Studies

by Ikuo Horii

Organized data management increases the reliability of statistical analysis. The basic purpose of data management is to assure the integrity and the quality of data. To assure data validity, establishing a checking system, such as data audit, would be desirable at the following points: protocol design, supervision of study schedule, definition of data, data collection, choice of tests and procedures, verification, data checking, data recording, data handling, data analysis, and data validation. To process an enormous amount of data on a multitude of items, use of a computerized system would be advantageous. The data processing system in toxicological studies should be based on a protocol-driven system, which gathers and records the data accurately. The main functions that are to be handled by computer are data collection, recording and retrieval via terminals, and statistical analysis of data and assembling of reports. One should be able to validate whether the computer system would perform its intended function accurately, reliably, and consistently. This paper discusses the basic considerations of data management and provides examples of the state of the computerized data management system and its validation.

Introduction

The fundamental significance of data management lies in assuring the integrity and the quality of data. If the integrity and quality of data are not assured, statistical analysis of the data will not be reliable, no matter what statistical procedure is used. On the other hand, it is important to comprehend the data characteristics and timing of data generation in selecting appropriate statistical procedures (1).

Toxicological studies for the assessment of drug safety require various kinds of tests and observations on a multitude of items in a large number of animals for long periods. Recently, to avoid deviation or error that may occur during data gathering and processing, computerized data processing systems have been developed for managing toxicological data. Good Laboratory Practice (GLP) regulations have also required that the non-clinical laboratory studies (toxicological studies) are accurately conducted, recorded, monitored, and reported in accordance with protocol and standard operating procedures (2). The present report discusses the basic considerations of data management, introduces a computerized system that incorporates GLP regulations (3,4), and reviews computer system validation (unpublished data, 5).

Basic Considerations of Data Management

Toxicological studies may use single doses, repeated doses, stop-start dosing, etc., for a variety of end points.

Table 1. Data volume in 13-week toxicity study in rats.

Data volume in one animal	
Body weight	30
Food consumption	60
Clinical sign	300
Blood chemistry	20
Hematology	20
Urinalysis	15
Functional test	20
Organ weight	25
Gross autopsy findings	10
Histopathology	50
Data volume in one experiment	
550 × 40 × 5 = 110,000	
(Data in one animal) (animal/dose) (dose)	

In each experiment, these studies have many test items to be observed, tested, or measured. Regarding the data volume of, for instance, a 13-week toxicity study in rats, the potential number of data evaluations is 550 per animal, and totally about 100,000 data points will be considered in one experiment (Table 1).

In practice, several types of experiments are performed in parallel in one laboratory. Under these conditions, many possible errors in data evaluation may arise. Complicated schedules followed in various studies, different tests, end points, large numbers of animals, and samples can conspire to increase errors. (Fig. 1).

Recently, to avoid errors during the data handling, computerized data processing systems have been developed to manage toxicological data.

GLP regulation has also required that nonclinical laboratory studies (toxicological studies) be accurately conducted, recorded, monitored, and reported according to protocol and standard operating procedures.

The following basic functions in various forms of toxicological

Department of Toxicology and Pathology, Nippon Roche Research Center, 200 Kajiwara, Kamakura, Kanagawa Prefecture, 247 Japan.

This paper was presented at the International Biostatistics Conference on the Study of Toxicology that was held May 13-25, 1991, in Tokyo, Japan.

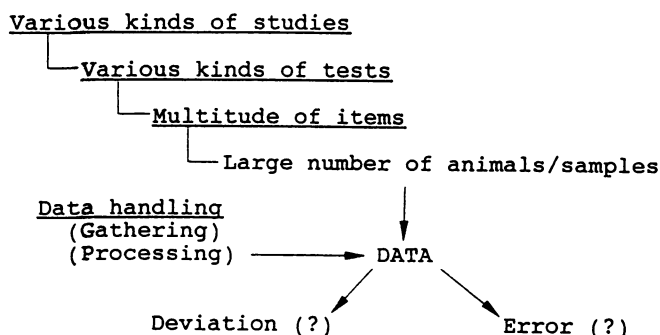


FIGURE 1. Data volume in 13-week toxicity study in rats.

studies (e.g. single and repeated administration toxicity, reproductive-toxicity, specific toxicity, carcinogenicity, etc.) should be considered for data management. A checking system for data audit should be established according to the following guidelines:

- Situation of protocol (how to refer to the protocol in practical settings)
- Supervision of study schedule (how to control the schedule)
- Definition of data (clarification of raw data)
- Procedure for data collection (how to collect data accurately)
- Guidance of test item and its procedure (to match standard operating procedures)
- Verification/check of data (who and how to check the data)
- Data recording (how to record data accurately)
- Data review (for easy retrieval)
- Monitoring the study (to establish the inquiry system)
- Qualification of data handling (who handles the data)
- Data analysis (to introduce relevant processing)
- Data validation (to assure the integrity of data)

An important point to consider is the reference to the protocol in any practical setting. The intention or purpose of the study, its schedule, and its contents should be made clear, and they should be clarified and carefully considered during data handling, gathering, or processing.

Supervision of study schedule refers to schedule control; namely, a practical schedule managed according to established protocols.

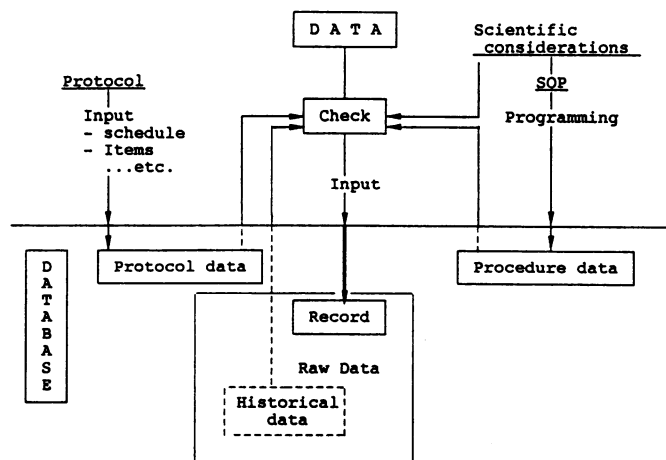


FIGURE 2. Data collection/recording in accordance with standard protocol.

Definition of data clarifies what the raw data are. For correct data processing and statistical analysis, we should deal with raw data directly, not secondary processed data. If we use the secondary processed data for further evaluation, verification of raw data and the secondary processed data should be performed.

Gathering data accurately is the basis of data handling. When validating data gathering by sensors such as analyzers or keyboard, attention should focus on avoiding errors or generating artificial changes in the data.

To develop a unified format to gather the data, it is useful to generate data of consistent quality. The standard operating procedure should be continually updated and improved as scientific and technological advances occur. Verification of data should be automatically systematized. At the time of data input, both scientific and computerized check systems should be employed. From the generation of data to the recording of data, the check system shown in Figure 2 may be employed, especially for computerized systems. No systems for data handling could easily or sufficiently manage the raw data check in real time without computerized support. For the data check before recording, previous protocol data and historical data should be referenced and a scientific check by the scientist should also be employed. All these check systems should be engaged by referring to protocol procedure data and historical data.

Data recording means simply recording data accurately. Final raw data are input in a uniform database that can be employed for further data processing.

The computerized retrieval system should be easily accessible from the unified database at any time.

Monitoring under the access and inquiry system is important to assure the integrity of the study. Furthermore, qualification of data handling is important to emphasize the responsibility for data handling and data security. Proper qualification can elevate the quality of the data alone.

Under the background protocol or process mentioned herein, the integrity and the quality of data would be fairly well assured. Using these data if one employs relevant statistical analyses, the assessment of health risks and other safety evaluations would be improved. From the perspective of toxicologists, three basic considerations about statistics in toxicology are important. First, statistical tests are performed under the premise that the samples are completely random in order to be free from biases. Second, accurate statistical tests should be done using the randomization tests such as the Pitman test so that the analyses are not performed by approximate methods based on the erroneous assumptions. Third, both the biological significance and the statistical significance should be considered before concluding that the toxicological effect is significant.

To properly conduct data management, the use of computerized systems would be profitable, and the approach should be applied with the functions mentioned previously.

Introduction of Computerized System

In the course of conducting toxicological studies, proper guidance to prompt investigators for the correct sequence of testing steps enforces the accurate conduct of the studies according to the standard protocol. When the experimental results are

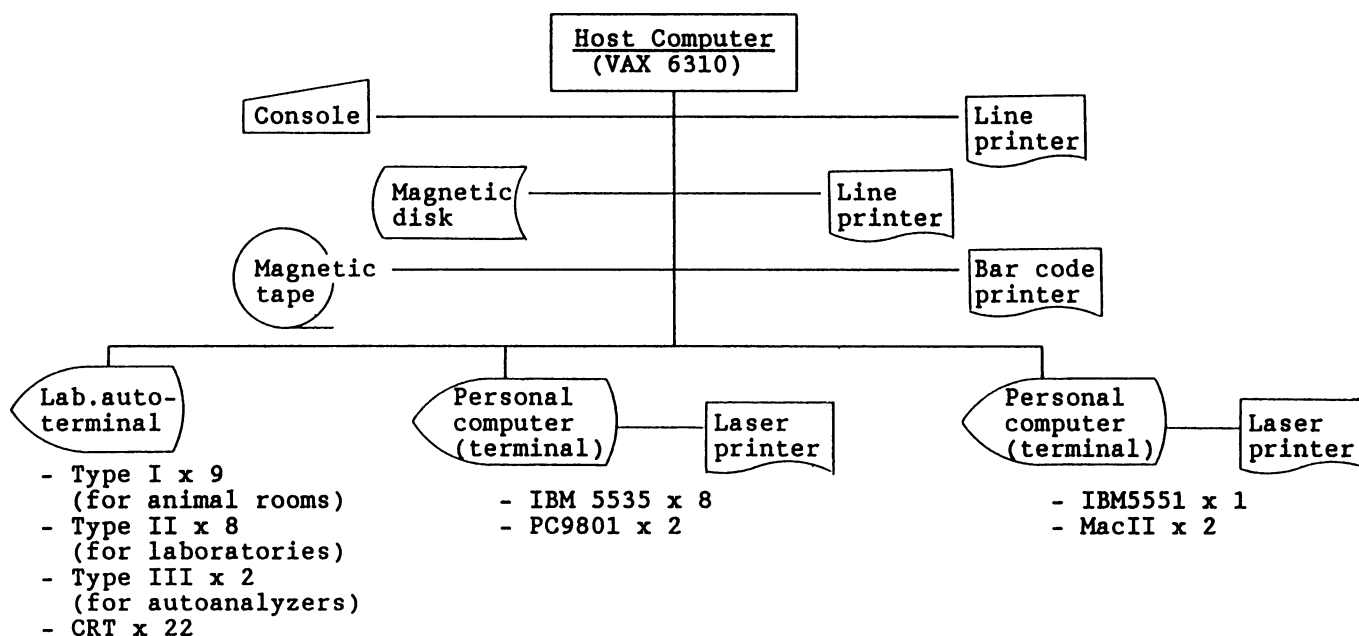


FIGURE 3. System configuration in the Nippon Roche Research Center/Toxicology and Pathology (NRRC/TP) program (hardware).

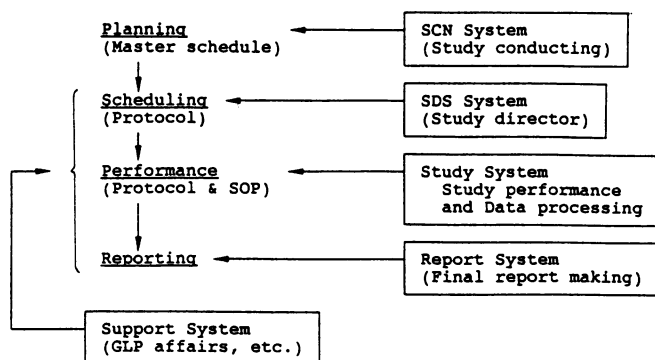


FIGURE 4. Total computer system in NRRC/TP (software).

received through computer terminals, the computer system should check the data against the standard protocol and against the history of previous results. The investigator can immediately recognize any errors and can correct them before the information enters the experimental database.

The computer system, which promptly processes the generated data by a combination of time-sharing database processing and a real-time multiprogramming system, was introduced as the total computer system in Nippon Roche Research Center (NRRC) and has been functioning successfully for more than 10 years. The aim of this system is to conduct the study accurately, to record the data, and to report results based on protocol and standard operating procedures under the GLP regulation.

A high-level mini-computer (VAX6310) was installed as the host computer, and microcomputer-based terminals for data gathering and retrieval were located in laboratories, animal rooms, dissection rooms, etc. (Fig. 3). Under the control/supervision of the host computer, all toxicological studies are performed in real time. Prompt responses (i.e., response time of

the computer) are assured with the specified software written by Massachusetts General Hospital Utility Multi-Programming System MUMPS. The following functions in the computerized system are covered completely: *a*) Study conduct with protocol-driven system (scheduling and guidance on video display terminal); *b*) Accurate record and strict correction of data (conversation system between investigators and computer, checking the system with standard programmed protocol, automatic input from sensors: autoanalyzer and balance); *c*) Review of data and monitoring of study (data retrieval with terminal or printer, automatic inquiry system for monitoring data integrity); *d*) Report (progress and final reports, statistical analysis); *e*) Backup system (security of data, archiving of data). Figure 4 shows the system configuration.

In the computerized system, toxicological studies are conducted on the basis of protocol data that have been programmed into the database and performed using software-based programmed standard operating procedures. The other systems are accessed or "called" as managing subsystems that support the main study systems to facilitate smooth study performance.

Data gathering and retrieval in this system are illustrated in Figure 5. Through the computer terminal interface, protocol and schedule are assigned by the study director, data are input at the laboratory or animal room by the examiner, and data on managing affairs are also input at the office by the responsible person concerned. All input data can be easily retrieved through the terminal. As a final report, tables and figures with the results of statistical analyses are printed after relevant data processing. As with the other function, sheets, labels, and written reports are also printed.

Figure 6 shows the flow-chart of relationship in each system for study conduct, data gathering, and data reporting. When the plan for a nonclinical study relating to drug safety is designed by the testing facility management, a protocol is prepared by a study

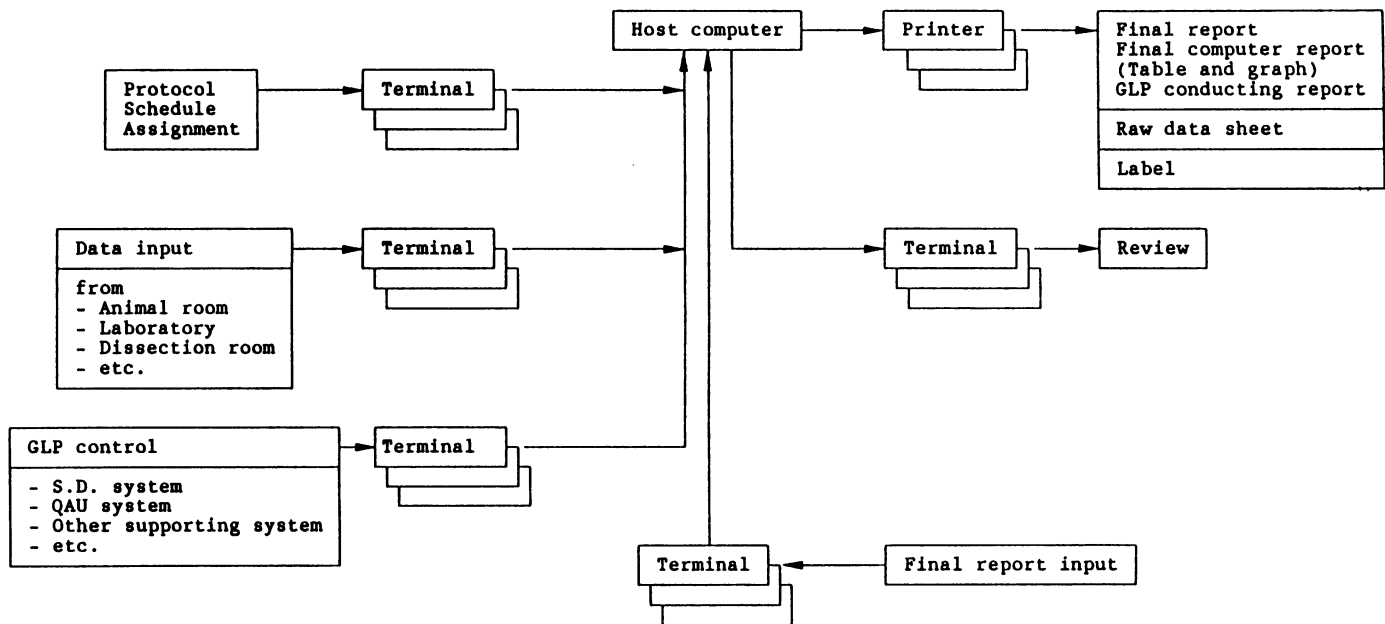


FIGURE 5. Outline of data gathering and retrieval (tox-DP system in NRRC).

director who is responsible for the overall study conduct. After the protocol is registered on the computer, a master schedule is made automatically in the computer system and shown on the video display to facilitate daily work guidance. In addition, all but information on assignments necessary for conducting the study not covered by the protocol is always entered by the study director during the experimental period.

Referring to the protocol assignment, time schedule, and standard operating procedures, the computer system provides daily guidance via video display as to what items shall be tested. This function effectively enforces protocol and standard operating

procedure adherence, and is strictly required by GLP regulation.

Regarding guidance of animal/sample number and any other detailed items for each data input, the investigator confirms animal and sample number and input items on the video display. If the investigator enters the wrong animal number or item, the video display does not show any response and data entry is refused. Input data undergo a range-check and data beyond the range are also highlighted or refused by the computer system.

For an identification system of animal and sample number, test items and data should be coupled with corresponding animal and sample ID, such as keyboard, magnetic card, bar code label, etc.

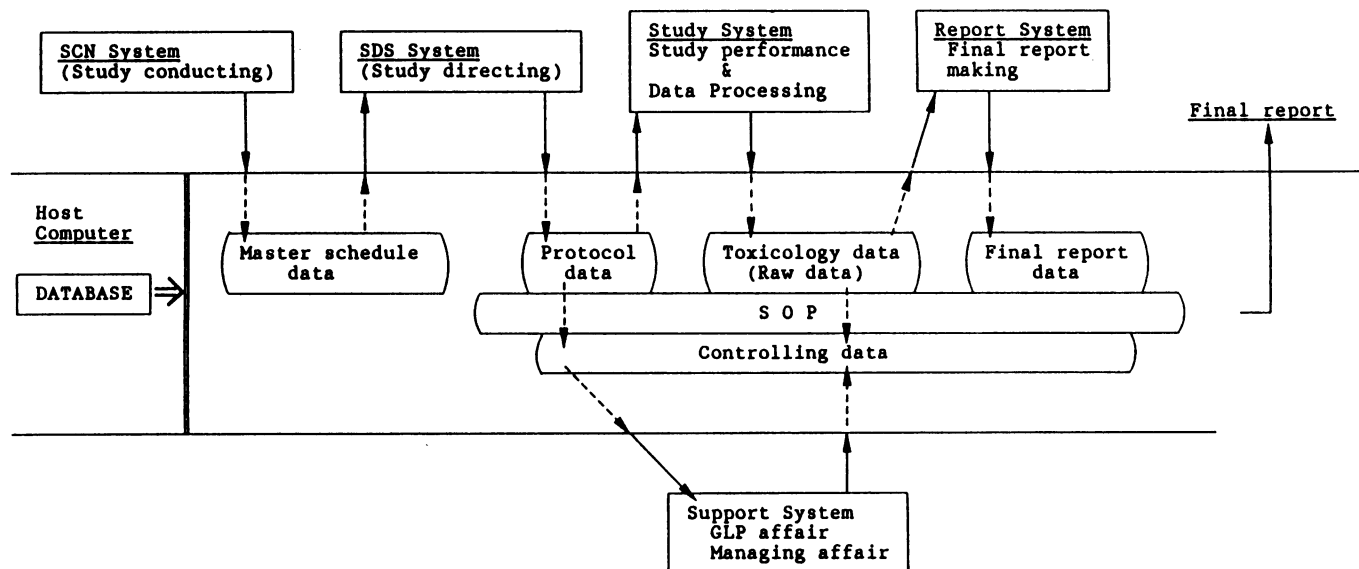


FIGURE 6. Flowchart of total computer system in NRRC/TP.

Of these, input by bar code reader is often employed along with video display of animal and sample ID because it enables handy, economical and reliable operations in this case.

For data recording, real-time data acquisition from terminal or instruments is employed. Data on body weights, feed consumption, and organ weights are recorded directly in the computer from autobalances. Data on hematology and blood chemistry from hematological and biochemical autoanalyzers, clinical signs and urinalysis from note tablets, test animal and sample identification from bar code reader, and dates and times from the time clock on host computer are all entered automatically.

Prior to data archiving in a permanent computer file (or database), however, all the data undergo corroboration as to their correctness and integrity. Both the computer system and investigator participate. Thus, errors in key punching and transcription can be eliminated.

When data require correction or amendment, any data in the experimental database must not be changed by anyone other than the study director responsible for overall study conduct. The study director notes the corrected data, along with reasons for the amendments, original entries, dates, and signatures from the study notebook.

When reviewing the data, the laboratory management, study director, investigator, and QAU, can perform a review of the data of on-going experiments, via video display terminals with key word entries and dates and test items to ensure adherence to standard protocols and also to confirm the data.

For the study report, the computer system provides progress reports periodically during the experimental period for assessing the status of the experiment. After completion of the study, the computer system provides a final report with statistical results for data evaluation.

After completion of the study, generated data are saved on magnetic tapes that are stored in the laboratory archives. These data can also be stored and retrieved from computer disk files from the magnetic tapes when reinterpretation of the data is of interest.

As a data back-up, the contents of the disk files are copied onto the magnetic tape after completion of daily processing. The data obtained on each day are recorded both on the disk and on the magnetic tape. The work schedule for the following day is recorded on cassette tape at the terminals. If the host computer system is down, the experimenters will be able to practice the experimentation via schedules on the video display terminals from cassette tape.

Computer System Validation

Computer system validation to assure the justification of the toxicological data should cover all the stages of computer systematization from the developmental stage to practical use (Table 2). Prior to performing system validation, basic specifications of the system (structure of file/database, functions to be applied, and procedures for study conduct) should be checked in detail.

As an initial validation at the developing/installing stage, which includes retrospective validation, all the documents concerning system design, programming and installing procedure, records

Table 2. Computer validation.*

Developing stage (Initial validation)	Practical use
Design	Daily check and confirmation Validity of data processing
Programming	Security check Counter plan for system down
Installation	Management of changes in software and hardware
System test *Hardware *Software	Confirmation by responsible person Education and training of user
Validation record	

*Basic specification of system (structure, function, conducting).

of system testing (hardware and software), and other validation records should be compiled properly.

For practical use, system validation should meet a daily check/confirmation. The validity of the data processing, a security check, a counterplan for system down-time, management of changes in software/hardware, confirmation of maintenance by responsible person, and education and training of user are all necessary.

The checkpoints of computer validation are classified as follows:

- a) Hardware and computer room—location
 - access to host computer
 - magnetic media as archives
- b) Development of computer system—software (developed in-house, vendor-supplied, existing) system design specification
 - system configuration
 - programming
 - program testing
 - validation testing
 - documentation
- c) Operation and maintenance

Regarding the items noted above, appropriateness of location of host computer and terminals, adequacy of design and capacity to function, and procedures for operation and maintenance would be checked at the time of inspection for computer validation.

For a more detailed description regarding these various issues, see the check lists of "GLP inspection of computer system" that have been delivered by the Ministry of Health and Welfare in Japan (6).

In conclusion, it is important to emphasize that proper data management elevates the reliability of statistical analysis.

The author wishes to thank Messrs. H. Shiozaki and E. Uchida for developing the total computer system in NRRC/TP, and Dr. T. Takizawa for his advice on system managing.

REFERENCES

- Horii, I. Audit of experimental data. *J. Jap. Soc. Qual. Contr.* 17: 136-149 (1987).
- Udaka, K. GLP and its practice. *Mutag. Toxicol.* 4: 4-21 (1981).
- Udaka, K., and Horii, I. Total lab-automation system in toxicology. *Mutag. Toxicol.* 5: 549-568 (1982).
- Udaka, K., and Horii, I. Introduction of software in toxicological data processing. *Experiment. Med.* 3: 100-104 (1985).
- Clark, S. Computer systems validation: an investigator's view. *Pharmaceut. Technol.* 65: 62-66 (1988).
- Ministry of Health and Welfare (supervised). GLP inspection of computer system. In: *GLP Kaisetsu. Yakuji Nippoh-sha, Tokyo, 1989, pp. 157-160.*