CHAPTER IV

RECEIPT AND PROCESSING OF BASIC RECORDS

This chapter describes the procedures employed to review and examine the basic records for completeness and accuracy, and the methods used in converting the large mass of information for the 1,873,483 children in the study areas, including Canada and Finland, to punch cards in order to permit mechanical tabulation of the various characteristics of the study population.

An outline of the processing steps, arranged in chronological order, and a résumé of detailed instructions applicable to each of the operations are included in the Appendix. Chapter V describes the comparable processes employed in translating the extensive amount of data gathered on reported cases of poliomyelitis in the study population to punch cards for analysis and correlation with the basic study population data.

Although clerical help would not be needed until July when data would become available from the Field Trial areas, staff members of the Vaccine Evaluation Center began preparing for that phase of the study early in April. During a two-month period 236 clerical applicants were interviewed, screened carefully, and classified as to their abilities and qualifications. The majority of the applicants were seniors and graduate students at the University of Michigan, and represented many nationalities and professions. The group finally chosen for the work included students of law, medicine, and engineering as well as general academic students, teaching fellows, professors' wives, and primary and secondary teachers; many were working on postgraduate degrees in the social sciences and welcomed the opportunity to gain this practical experience in their field of interest; most of them displayed evidence of their pride in taking part in the project while earning modest funds at the same time.

A list was made of full-time and part-time workers from which people were called into service as they were needed to process data according to the volume received. At the time of peak clerical activity in July there were 118 persons employed as statistical clerks, supervisors, and typists.

During this same period of preparation, the office space assigned to VEC in the old maternity hospital building was being renovated and made ready for use. Coat racks and bulletin boards were installed and designation signs and room numbers were posted. Refrigeration was provided for storage of a sample of each lot of vaccine to be employed in the Field Trial, for use in any additional studies which might be indicated.

Office equipment - much of it needed on a short-term basis - was mainly borrowed or rented. Equipment which necessarily had to be purchased was in most instances obtained with a definite provisional resale understanding to ensure economy. Many needed items were obtained at reduced prices from the University of Michigan surplus stores and from Government Surplus. The U.S. Bureau of the Census supplied collapsible clerk-tables and portable bins on a loan basis. Used files, desks and chairs were procured, and typewriters, calculators, and adding machines were obtained.

Office supplies were purchased and the many Field Trial forms designed and printed. Receipt and control cards were made ready so that this unit of the procedure was set up well in advance of the receipt of documents from the

field; a small staff of supervisors and control clerks began work in June in order to be prepared to receive and begin the processing of materials as they arrived from the field.

PROCESSING AND TABULATION

Before the mass of information was converted to punch cards for mechanical tabulation, basic records were examined to verify accuracy and completeness of required data, and a control procedure and system of filing were devised.

RECEIPT, CHECK-IN AND CONTROL PROCEDURES

In advance of the actual receipt of clinic records (FT-3, Registration Schedule, and FT-4, Vaccination Record), the Vaccine Evaluation Center prepared a form (VEC-1, Check-In and Control Record) on which the basic documents were listed when received (see forms in Appendix).

After these appropriate entries were made for each of the 15,000 schools in the study, the records were filed and used to control the flow of documents through the various processing phases. Registration Schedules for each school, in numerical order, were placed in a letter-size folder labeled by state, county, city, town or township, and the school name and five-digit code.

The Vaccination Records were arranged in the order in which the children were listed on Registration sheets and placed in brown manila envelopes marked with school name and code to be filed with the folder ready for transmittal to the next step in the processing.

Movement of materials through the various stages of operation and the role of the control form (VEC-1) in maintaining orderly transfer of the school record folders are discussed in detail in Form VEC-20, included in the Appendix.

The files were set up and maintained in four major sections as follows:

1. Current or active files used by the Control Section of the staff for materials awaiting final disposition.
2. Permanent files for school records (Form FT-3).
3. Permanent files for Vaccination Records (Form FT-4).
4. Metal card files for the basic record punch cards prepared for each of the children in the study.

All transmitted work sheets and code sheets, listings and tabulations were filed in clothbound ring binders. A minimum supply of each of the forms and instructional material used in the processing of the records was kept, and basic school records were microfilmed, for ultimate permanent record filing.

EDITING AND CODING

Following the check for completeness of registration, the basic records for each school were edited in order to ensure accuracy of information on each child in the study and for consistency between schools and trial areas. The participation status and the lot numbers of the material with which the child was inoculated at each clinic were coded on a concealed basis, together with the reason, if known, for failure to complete the series of three inoculations and any reactions to the inoculum noted on the records.

Specific instructions (Form VEC-23) covering this operation are contained in the Appendix. All editors and coders were given special training for the operation, and the operating units were set up in groups of twenty people under an experienced supervisor. The records were edited and coded as quickly as possible in order that errors and inconsistencies could be resolved through correspondence with the field before too great a movement in the study population could occur.

The general progress in obtaining and processing the basic records for the entire

Table 26

PROCESSING STATUS OF BASIC RECORDS THROUGH ALL STAGES OF OPERATION
(JULY 3, 1954 - FEBRUARY 20, 1955)

Operation	July 3		July 31		August 31		September 30		October 31		November 30		December 31		January 31		February 20	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Received at VEC	24,382	1.3	711,037	37.6	1,478,669	78.2	1,820,696	96.3	1,859,829	98.4	1,882,388	99.6	1,889,815	100.0	1,889,815	100.0	1,889,815	100.0
Held for Receipt Problems	-	-	26,512	1.4	113,039	6.0	12,111	0.6	7,082	0.4	1,126	0.1	-	-	-	-	-	-
Edited and Coded	-	-	491,605	26.0	1,273,721	67.4	1,579,394	83.6	1,851,946	98.0	1,881,262	99.5	1,889,815	100.0	1,889,815	100.0	1,889,815	100.0
Held for Editing Problems	-	-	86,864	4.6	252,058	13.3	305,385	16.2	-	-	-	-	-	-	-	-	-	-
Final Review Completed	-	-	201,505	10.7	1,017,292	53.8	1,186,939	62.8	1,833,121	97.0	1,881,262	99.5	1,889,815	100.0	1,889,815	100.0	1,889,815	100.0
Sent to Punch	-	-	201,505	10.7	813,770	43.1	1,186,939	62.8	1,829,309	96.8	1,881,262	99.5	1,889,815	100.0	1,889,815	100.0	1,889,815	100.0
Punching Completed	-	-	-	-	247,041	13.1	856,460	34.7	1,621,500	85.8	1,861,016	98.5	1,889,815	100.0	1,889,815	100.0	1,889,815	100.0
Punching Verified	-	-	-	-	24,351	1.3	414,344	21.9	1,242,374	65.7	1,861,016	98.5	1,889,815	100.0	1,889,815	100.0	1,889,815	100.0
Returned from Punch	-	-	-	-	24,351	1.3	122,514	6.5	907,213	48.0	1,861,016	98.5	1,889,815	100.0	1,889,815	100.0	1,889,815	100.0
Sent to Mechanical Edit	-	-	-	-	-	-	-	-	-	-	1,849,688	97.9	1,889,815	100.0	1,889,815	100.0	1,889,815	100.0
Returned from Mechanical Edit	-	-	-	-	-	-	-	-	-	-	1,612,009	85.3	1,734,513	91.8	1,829,916	98.8	1,889,815	100.0
Number of Areas Corrected	-	-	-	-	-	-	-	-	-	-	128	59.5	200	93.0	211	98.1	215	100.0
Number of Areas Tabulated	-	-	-	-	-	-	-	-	-	-	-	-	58	27.0	211	98.1	215	100.0

Percents based on a study population of 1,889,815 and 215 areas (which includes 211 U. S. Areas, 3 areas in Canada and 1 in Finland).

study population through the various operations is shown in Table 26.

The last of the basic records received from trial areas in the United States were coded and transmitted to punch late in October, 1954. Corrections and additional information were still needed from 49 areas to complete the basic records. The cards for these children were coded to reject in an edit tabulation and were corrected when the information was received from the field.

Inconsistencies found during the editing and coding operations required correspondence with 192 of the 211 Field Trial areas in the United States. The most frequent errors together with the percent of the total study population they affected were as follows:

1. Failure to submit FT-4 Vaccination Record when dates of vaccination were entered on FT-3 register 3,642-0.20%
2. Birth date missing 737-0.04%
3. Parent request "no" but record indicated child was vaccinated or gave blood 2,006-0.11%
4. Doubtful as to "yes" or "no" as recorded for parental request 4,216-0.23%
5. Discrepancy in vaccination dates between FT-3 and FT-4 forms . . 4,627-0.25%
6. More than one grade recorded on register and not separately identified 1,592-0.09%
7. Questionable as to whether child should be included in study population (name duplications, child moved away prior to vaccination, etc.) 1,055-0.06%

8. Failure to indicate vaccine lot number of FT-4 form 1,421-0.07%

As indicated, these inconsistencies represented an insignificant proportion of the total registration. On the whole, the Registration and Vaccination Records were very well prepared.

The verification of the editing and coding was done by clerks who had demonstrated a high degree of accuracy in these processes and was conducted on a sample basis as rapidly as the coders were qualified. The procedures employed for sample verification are detailed in a later section of this chapter which outlines the quality control measures applied to the statistical operations.

Detection of Recording Errors Made in Field Clinics

The combination of the first two digits of the code with definite two-letter suffix patterns to identify vaccine or placebo was purposely introduced into the code scheme to provide an error detection device to bring to light erroneous entries made on the Vaccination Records. A total of 67,600 code combinations can be made by combining two digits with two letters (10 x 10 x 26 x 26 = 67,600). Of this total, the code scheme devised to identify the inoculum comprised only 200 specific code combinations. Therefore, the probability that a code recorded in error would coincide with one of the 200 legitimate code combinations was less then 3 in 1,000 and the probability was less than 2 in 1,000 that the error would give rise to a reversed status (vaccine for placebo or vice versa).

The clerks in the editing operation were provided with a list of the 200 proper code combinations and each Vaccination Record was compared with this list to make certain that the recorded code was correct. When the editing clerk discovered a recording error, it was usually a relatively simple matter to determine the correct code. In making this determination the clerk would use the knowledge that each child was to receive three injections of the same substance and that usually ten children successively listed on the Registration Schedule were inoculated from the same vial. If necessary the clerk could refer to file copies of the shipping orders supplied to the Evaluation Center by the manufacturer to determine what lots of vaccine and placebo were sent to a particular area. If the correct code could not be readily determined from the records available, correspondence was initiated with the local health officials of the areas concerned to enlist their aid in solving the problem.

Vaccine and Placebo Recodes

To reduce the number of digits required in the punch cards and to eliminate the complexity of alphabetical punching, the vaccine and placebo codes were recoded as the information from the Vaccination Records was transcribed to the Registration Schedules in the editing and coding process. The editors at VEC were provided with a listing of two-digit recodes for all of the possible five-digit and two-letter suffix codes used in the field. The first digit denoted the substance and the manufacturer, and the second digit the particular lot of material injected. The lists were so constructed and the particular digits of the recode so defined that the nature of the inoculum was still concealed in this phase of the processing.

The codes for the nature of the inoculum received were not introduced into the punched cards representing reported cases of poliomyelitis in the study population nor were tabulations or distributions by vaccination status completed until the final VEC diagnosis had been established for each case. In this manner, knowledge of the vaccination status of cases was kept concealed.

Methods Used to Correct Errors and Omissions

In the editing and coding operation, incomplete and inconsistent records which could not be immediately corrected were set aside for more detailed investigation. A number of clerks who had attained best rates

of production and lowest rates of error were selected to handle these problems; it was believed that by limiting this responsibility to a few people, competency would be more rapidly developed and the range of variation in making the required decisions would be narrowed. Specific instructions (Form VEC-23a) covering this operation which was called "Technical Review" are contained in the Appendix.

Staff members who had visited the trial areas throughout the nation observing clinics in action briefed the technical review clerks regarding the organization of the clinics and the various ways in which the forms may have been filled out. Particular attention was directed to those situations and areas where known deviations from the prescribed procedure had taken place, so that the correct interpretation could be readily made. The knowledge that in almost all instances the Registration Schedules were prepared from school records whereas Vaccination Records were prepared from Parental Request Forms was important in reconciling inconsistencies between the two records. Understanding of the various patterns of entries employed by the numerous persons who prepared the records in the field was invaluable in determining the significance of irregular entries, such as check marks meaning "yes" and "X's" or blanks meaning "no."

Technical review clerks made use of a number of sources of information in correcting, verifying, and completing the records. Reports from the Field Trial areas indicating dates of clinics and number of children inoculated, the manufacturers' records of shipment of lots of vaccine and placebo to the areas, and records indicating the code numbers of the particular material were used by the review clerks in analyzing problems which were referred to them. Many solutions were obtained by reference to the records of other schools from the same area. Correspondence received by the Evaluation Center showing that children may have moved in or out of an area during the course of vaccination, and letters concerning required substitutions of vials, served to clarify certain perplexing questions where unexpected lots were recorded.

When reference to all information available at VEC failed to provide the proper solution to a particular problem, the health officer was asked to examine all local records or to obtain information from the child or parents directly.

Despite all efforts made, the records for a small group of children could not be corrected, and it was decided to exclude them from the study population. None of these children was ever reported as a case of poliomyelitis, and their exclusion could have no significance in the reported results. The number of children involved and the reason for their exclusion are indicated below:

In observed areas, 1st or 3rd grade: dates of vaccination indicated on Registration Schedule - no Vaccination Record	36
Parental request "no": dates of vaccination indicated on Registration Schedule - no Vaccination Record	24
Vaccination Record indicated mixed injections of vaccine and placebo	748
Injections received but inoculum not identified	175
Total	983

Among 200,745 children in placebo areas who received three injections of vaccine, only 33 received material from more than one lot.

Quality Control and Error Detection

In the editing and coding operations, the Registration Schedules and Vaccination Records were thoroughly reviewed for completeness of information and for consistency within each form and between the two forms. In addition, certain of the items, namely, parental request, vaccine lot numbers, reported reactions, and reasons for failure to com-

plete inoculations, were coded in preparation for punching of cards for the basic study population. The clerks assigned to these operations were instructed to correct those errors for which the solution was readily apparent (for example, discrepancy in the age reported) and to set aside the records of those schools for which more detailed investigation was required.

To ensure that errors in the field reports were detected and eliminated so as to maintain a high level of accuracy in the coding operation, all work processed by the editing and coding clerks was subjected to a quality control procedure designed to determine whether it was being performed within satisfactory rates of accuracy. Error rates were computed on the basis of error lines per 100 lines verified. An error line was defined as a line of information in the Registration Schedule on which the clerk had made one or more coding errors and/or failed to detect or properly to correct one or more instances of inconsistency or incompleteness in reported information. This quality control procedure required complete verification of the output of each editing and coding clerk during an initial period until his work indicated that he had reduced his error rate to, or below, the acceptable maximum. Thereafter, his work was verified on a sample basis. Detailed records were maintained on the number of lines processed by each clerk and the number of error lines detected. All errors in editing or coding detected by the verifier were corrected. The section supervisor reviewed with the clerk each unit of his work which had been verified, pointing out the errors which had been uncovered and instructing the clerk in the correct treatment of each improperly handled item so that the same error would not be repeated.

After the first three weeks of operation, all coders whose error rates were 1 percent or less, based on 3,000 or more verified lines, were qualified for sample verification. Those coders for whom less than 3,000 lines were verified or whose error rate was more than 1 percent continued to have their work verified on a 100 percent basis. All coders who failed to qualify for sample verification within three weeks after the initial qualifying period were removed from the operation.

Verification was made of approximately one-third of the lines edited and coded by those clerks who qualified for sample verification. The lines to be verified were assigned daily by the section supervisor who made random selections of the pattern of lines to be inspected by each verifier from a set of 28 cards prepared in accordance with the Line Number Scheme. On each of these

LINE NUMBER SCHEME FOR USE IN SAMPLE VERIFICATIONS OF CODING AND EDITING

Starting Lines	Lines To Be Verified
1 and 19	1- 5 and 19-23
2 and 20	2- 6 and 20-24
3 and 21	3- 7 and 21-25
4 and 22	4- 8 and 22-26
5 and 23	5- 9 and 23-27
6 and 24	6-10 and 24-28
7 and 25	7-11 and 25-29
8 and 26	8-12 and 26-30
9 and 27	9-13 and 27-31
10 and 28	10-14 and 28-32
11 and 29	11-15 and 29-33
12 and 30	12-16 and 30-34
13 and 31	13-17 and 31-35
14 and 32	14-18 and 32-36
1 and 32	1- 5 and 32-36
2 and 31	2- 6 and 31-35
3 and 30	3- 7 and 30-34
4 and 29	4- 8 and 29-33
5 and 28	5- 9 and 28-32
6 and 27	6-10 and 27-31
7 and 26	7-11 and 26-30
8 and 25	8-12 and 25-29
9 and 24	9-13 and 24-28
10 and 23	10-14 and 28-27
11 and 22	11-15 and 22-26
12 and 21	12-16 and 21-25
13 and 20	13-17 and 20-24
14 and 19	14-18 and 19-23

cards, two blocks of 5 lines each were designated and the entire set included all lines so that from the shuffled deck there was equal chance that each or any line would be drawn for verification. This pattern was so devised as to obtain representation in the verified sample of lines on the lower half of the FT-3 sheet as well as on the upper half.

A coder who was once qualified for sample verification remained qualified so long as his error rate was 1.5 percent or less, based on the lines verified and errors found for one full week's work. Coders who, once qualified, failed to maintain an error rate of 1.5 percent or less were returned to 100 percent verification for a minimum of one week. During the course of the operation no coder returned to 100 percent verification failed to requalify for sample verification after one week.

Verifiers assigned to sample verification were cautioned to concern themselves only with the lines specified for each sample; therefore, errors which occurred in lines outside the sample were not detected; errors discovered on sample lines were corrected.

Analysis of the production and error records maintained throughout the editing and coding operations indicate that the actual level of coding errors throughout these operations was less than 1 percent.

PUNCHING OF BASIC DATA CARDS

The punching of approximately two million basic study population cards was done on contract with the IBM Service Bureau at Detroit, Michigan, who guaranteed that an error rate of 2 percent or less would be maintained in the punching. A quality control procedure similar to that used in the editing and coding operations was applied to the Service Bureau personnel. Although detailed production and error records for the punching operations were not submitted, staff members of the Evaluation Center inspected the operation with sufficient frequency and detail to determine that the desired quality level was maintained. Occasionally the card punch operators detected errors which had slipped through the editing operations, and such errors were corrected.

To further reduce the number of residual errors, the first tabulation of the cards included a mechanical edit whereby each card was inspected electronically, and those which contained erroneous punches or inconsistent data were ejected from the file. This mechanical edit is further discussed in the following section of this chapter.

The Evaluation Center had planned to study a representative sample of the punch cards after mechanical edit to determine the nature and extent of the residual error in the data. Because of the low level of error attained as a result of the control procedures established, this study was not conducted.

TABULATION OF BASIC POPULATION DATA

Objectives of the tabulating program developed for the basic registration cards were: (1) to obtain distribution of the characteristics of the total study population by areas, age, sex, vaccination status, and vaccine lot numbers, and type of reaction reported if any; (2) to prepare a detailed distribution of these characteristics and others to serve as denominators in analyzing reported cases of poliomyelitis in the study population; and (3) to prepare a list from the cards of study children in placebo areas for each state showing the registration numbers of those receiving vaccine and those receiving placebo. It had been agreed that after the Summary Report these lists would be furnished to state health officers so that they and the parents could know which material each child had received.

The required tabulations involved four runs of the 1,829,916 registration cards through a series of IBM tabulating machines. The first three runs were tabulated on an IBM electronic statistical machine (101), and the fourth was prepared on an IBM accounting tabulator, (407). The first run of the cards was a mechanical edit to detect those cards which contained erroneous or inconsistent data.

Two methods of detecting errors were employed: The first method consisted of electronically inspecting each card and selecting those cards which contained impossible or inconsistent punches; such cards were then checked through the Registration Schedules and Vaccination Records to determine the required corrections. The second method of detection was related to cards which were not necessarily erroneous or inconsistent in an individual instance but would be suspect if the occurrence was beyond anticipated proportions. The mechanical edit, therefore, included the preparation of certain frequency distributions and, when the frequency of occurrence was higher than anticipated, a detailed investigation of the basic records was undertaken to determine the reliability of the recorded data. The number of cards selected as having erroneous or inconsistent data was extremely small, amounting to only 0.7 percent of the total study population. The actual number of cards rejected and corrected was 5,661 in placebo areas and 7,164 in observed areas.

The second and third runs of the cards provided the detailed characteristics of the study population and the denominator data needed in analyzing reported cases of poliomyelitis in the study population. The second run tabulated the population for each area and included distributions of the study population by parental consent and type of inoculation. The third run was tabulated for each vaccination status group in each area. There were six such groups for placebo areas and nine for observed areas. Detailed distributions of the study population by sex and age, sex and color, and urban and rural distributions were obtained.

The fourth run, prepared on type 407 accounting machine, provided a distribution of the vaccination status of the study children for each school participating in the Field Trial and for each lot of vaccine used. In addition, this tabulation provided the detailed listing needed to furnish the states with the registration numbers of the children in placebo areas who received vaccine or placebo.