

Hawaii Health Information Corporation

---

Enhancing Hawaii Hospital Information Content (eHHIC)

Deliverable 4

# Linking Lab Data to Inpatient Discharge Data

---

## TABLE OF CONTENTS

I.	OBJECTIVE.....	3
II.	METHOD.....	4
	a. Validation.....	4
	i. Edits.....	4
	b. Linking Laboratory Data.....	4
	i. Preliminary Linking.....	4
	ii. Deterministic/ Probabilistic Linking.....	4
	a. Linking Iterations.....	5
	iii. Linking with Limited Identifiers.....	7
	iv. Exceptions.....	8
	a. Exclusions.....	8
	b. Non-Linked Identifiers.....	9
	c. Emergency Room Discharges.....	9
	c. Overall Occurrence.....	9
	i. Inpatient Discharges Linked to $\geq 1$ Lab Record.....	9
	ii. Inpatient Discharges Without Labs.....	10
	iii. Laboratory Records Linked to Inpatient Discharges.....	11
III.	CONCLUSION.....	11
	APPENDIX A: Field/Edit Specifications.....	12
	APPENDIX B: Data Elements used for Deterministic/Probabilistic Linking.....	13
	APPENDIX C: Iterations for Deterministic/Probabilistic Linking.....	14
	APPENDIX D: Excluded Records, by Reason of Exclusion.....	15
	APPENDIX E: Inpatient Discharges with $\geq 1$ Lab Record, By Hospital and Calendar Year.....	16
	APPENDIX F: Discharge Records with No Lab Record, By Hospital and Calendar Year.....	17
IV.	SIGNATURES.....	18

---

## I. **Objective**

To link thirty-two specific laboratory test results from CY 2008-2011 with Hawaii Health Information Corporation's (HHIC) hospital discharge data. This was achieved by:

- A. Establishing specific edits on key matching variables
- B. Performing data linkage iterations utilizing deterministic/probabilistic methods
- C. Processing of Exception Records such as excluded and non-linked records

## II. Method

### A. Validation

#### 1. Edits

After a comprehensive quality review of the data, 30,668,969 submitted lab records for CY 2008-2011 were loaded into a staging environment for further validation and ultimately for linking to a hospitalization-related discharge record.

Validation included performing edits on key variables (such as Admission Date, Account Number and Sending Facility) within the lab results. These field edits ensured the integrity of the dataset to be used in the linking process. A complete list of key variables and the edit specifications that were applied are shown in Appendix A.

Laboratory data for one facility, Molokai General Hospital, could not be processed for linking purposes as key variables used in the linking process were not supplied in the data extract. Key linking variables included account number, admission date and medical record number. Of the 30,668,969 submitted lab records, 1,212,691 (4%) laboratory records were excluded due to missing key variables. A total record count of 29,456,278 lab records were used in the preliminary linking process.

### B. Linking Lab Data to HHIC Hospitalization Data

#### 1. Preliminary Linking

A preliminary linking process was performed on 29,456,278 laboratory records based on the account number and facility code provided in the lab record against the inpatient discharge account number and facility code. This method however did not result in a sufficiently high match percentage. Match percentages varied 50% to 90% across all facilities. As a result, a new approved method was devised to perform a more complete linking process between the lab result and HHIC's Inpatient Discharge Data.

#### 2. Deterministic/Probabilistic Linking

Additional linking methods were established to perform a more detailed deterministic/probabilistic approach based on nine variables that were identified within the submitted lab data that were in common with the

corresponding Inpatient discharge record. Appendix B lists the variables used for this approach.

As shown in Appendix C, multiple linking iterations were performed to link the laboratory data to the inpatient discharge data. The first seven iterations complemented each other and addressed different data issues such as variances noted in the account number between the discharge record and the laboratory record. The eighth and ninth iteration were used to address possible matches not linked in iterations 1 through 7. A brief description of the iterations performed follows.

*a. Linking Iterations (Refer to Appendix C)*

Iteration 1

The first iteration linked Lab Data and HHIC Inpatient discharge records based on the following fields being exact matches: Account Number, Medical Record Number, Date of Birth, Gender, HHIC Admission Date to Lab Observation Date, First Name (up to first space if multiple names are present), Last Name, and Hospital ID/Sending Facility. This iteration yielded the greatest volume of linked laboratory records with a total of 17,627,513 records linked to a hospitalization discharge record.

Iteration 2

The second iteration removed the patient's first name as a matching variable. This iteration addressed the issue of nicknames and middle names used as first names and resulted in 188,687 laboratory records linked to an inpatient discharge record.

Iteration 3

The third iteration removed the patient's last name as a matching variable in addition to the patient's first name. This iteration addressed the issue where patients changed their last names. In combination with iteration two, this iteration addressed transposed first and last name issues. This resulted in 178,426 laboratory records linked to an inpatient discharge record.

Iteration 4

The fourth iteration resulted in a total of 379,316 lab records linked to an inpatient discharge record. This iteration removed Account Number but reinstated First Name and Last Name as matching variables. This iteration addressed an issue with the Account Numbers submitted by The Queen's

Medical Center (QMC): The account number submitted by QMC was not the same account number as noted in HHIC's Inpatient Discharge dataset. QMC transmits a Contact Serial Number (CSN) to the laboratories that is specific to the visit. This identifier is then referenced as the account number between the hospital and the laboratory. The identifier transmitted to HHIC in the hospital discharge record is the hospital account record (HAR) or billing number and is referenced as the account number between HHIC and the hospital.

#### Iteration 5

The fifth iteration removed Medical Record Number and reinstated Account Number as a matching variable. This iteration addressed Medical Record inconsistencies due to formatting variances such as medical record numbers that contained leading 0's or '-', such as in the case of data submitted by Castle Medical Center.

This iteration also addressed an issue with incorrect Medical Record Numbers (MRN) submitted in Hawaii Pacific Health (HPH) Blood Gas Lab results. Hawaii Pacific Health is a hospital system that consists of four facilities. HPH maintains a system-wide (enterprise) medical record number in addition to a facility specific MRN. The Medical Record Number submitted by HPH (enterprise MRN) was not the same MRN in HHIC's Inpatient Discharge dataset (facility specific MRN). As a result, lab results did not match based on this variable. A large number of HPH Blood Gas Lab results were therefore matched under this iteration.

Approximately 173,495 lab records were linked during this iteration.

#### Iteration 6

The sixth iteration removed Admit Date but reinstated Medical Record Number a matching variable. The Medical Record Number and Account Number fields alleviated the uncertainty left behind from the removal of the Admit Date as a linking variable. The Medical Record Number and Account Number combination were a distinct combination that uniquely identified a particular inpatient discharge. This iteration accounted for the linking of 1,316,777 lab records to an inpatient discharge record.

### Iteration 7

The seventh iteration removed Birth Date and reinstated Admit Date as a matching variable. This iteration addressed default birth dates when one wasn't provided. A total of 123,751 lab records were linked during this iteration.

### Iteration 8

The eighth iteration utilized LinkageWiz<sup>1</sup> to perform probabilistic linking on all variables as noted in Appendix B. Based on these variables, a combined weight score of 40 or above was considered to be a match. All variables were evaluated to link any remaining lab results not successfully matched in the first seven iterations. Many records contained errors on two of the fields, resulting in the records not being linked in the first seven iterations.

For example, it was discovered that the gender variable was a main contributor to mismatches as several records contained an 'unknown' gender. The date of birth listed for one facility (The Queen's Medical Center) also contained a large number of '01/01/1901'; a default value and could not be used for linking purposes.

A total of 103,877 additional records were linked probabilistically for this iteration.

### 3. Linking with Limited Number of Identifiers (Iteration 9)

Blood gas results from one hospital system, Hawaii Medical Center<sup>2</sup>, needed to be handled differently. Shortly after receiving the data, the hospital system filed bankruptcy and closed. This presented a challenge in that a resource was unavailable to provide answers to questions present within the dataset.

The dataset, consisting of a total of 221,434 lab records from CY2008 through December 04, 2011, did not have an adequate number of identifying data variables. Therefore, a business rule was established that five fields (Sending Facility Code, Medical Record Number, Date of Birth, Gender and Patient Last Name) needed to match exactly in order to link the an HHIC Inpatient Discharge record to a lab record. Of the 221,434 lab records, this resulted in the linking of 131,094 lab records to a hospitalization discharge record and is described in Appendix C as Iteration 9.

---

<sup>1</sup> Linkagewiz is a data matching, de-duplication and data cleansing tool used to link records across multiple databases and to identify duplicate records.

<sup>2</sup> Hawaii Medical Center is a hospital system consisting of two facilities – Hawaii Medical Center East and Hawaii Medical Center West

#### 4. Exceptions

To accommodate the providers and to stay within the timeframe of the project, it was important that HHIC remain flexible. Some facilities did not have the resources to modify their programs to transmit only the requested 32 specific lab tests. In order to maintain participation in the study, HHIC allowed those hospitals to send all labs for all patient classes. A total of 9,233,342 laboratory records were not linked to an inpatient discharge record as these records did not meet the data quality criteria and were rejected.

##### a. Exclusions

Overall 5,303,874 records were excluded for one or more of the following reasons:

- Lab Test is not one of the 32 requested labs
- Observation Value is Missing
- Observation Date is Missing
- Observation Result Status is not Final ('F') or Corrected ('C')
- Microbiology Records

##### 1. Filtering transmitted data to the requested 32 lab tests

Two facilities did not have the resources to modify their programs to restrict transmission to the requested 32 lab tests. Records that contained tests not among the 32 requested labs of interest were excluded. As shown in Appendix D, this particular filtering process comprised the majority of the exclusions.

##### 2. Missing Values

Records were excluded if missing values did not exist for key variables in the lab test.

- Laboratory test (observation value): The lab value is required to assign the adjusted risk of mortality (ROM) scores. If the observation value was missing, the lab test was excluded.
- Admission date: Accurate linking of the lab record to the hospitalization-related discharge record is dependent on linking the lab observation date to the discharge record's admission date. Records missing this key linking variable would therefore prohibit the linking of lab data and assignment of the adjusted risk of mortality and were thus excluded.



### 3. Observation Result Status

The lab result status of 'F' (final) or 'C' (corrected) guaranteed that the completed test result was received. Other result statuses ('P' – pending and 'I' – incomplete) were therefore excluded, as only the final or corrected test result was relevant to the study.

### 4. Microbiology Records

After reviewing the test results and values, we excluded the Microbiology tests (blood culture, LOINC 600-7; urine culture, LOINC 630-4; and sputum culture, LOINC 6460-0) from the study as their non-numeric observation values rendered the test results unquantifiable.

An overall summary of the excluded records is found in Appendix D.

#### *b. Non-Linked Records*

An additional 2,467,067 laboratory records were not linked as the observation date on the laboratory record was not within the dates associated to the inpatient discharge record.

#### *c. Emergency Room Discharges*

All hospitalization-related laboratory results within the time frame of the study were requested. Two facilities did not have the resources to modify their programs to restrict transmission to the requested 32 lab tests for specific patient classes and submitted lab data for all patient classes. This resulted in a total of 1,462,401 laboratory records that linked to an Emergency Room discharge record and no corresponding hospitalization discharge record. These records were excluded from the matching process.

### *C. Overall Occurrence*

A total of 389,348 out of 450,756 hospital discharge records from CY 2008-2011 were linked to one or more lab records (86.4%); resulting in 61,408 hospital discharge records with no lab record. A total of 20,222,936 laboratory records out of 30,668,969 submitted lab records were linked to 389,348 inpatient discharge records (65.9%).

#### *1. Inpatient Discharges Linked to $\geq 1$ Lab Records*

Table C.1 describes the Inpatient discharges linked to one or more lab records per year.

**Table C.1 Inpatient Discharges Linked to  $\geq 1$  Lab Record by Year**

Year of Discharge	Discharges linked to $\geq 1$ lab record	Total Discharges <sup>3</sup>	Pct
2008	95,796	112,542	85.1%
2009	98,713	113,160	87.2%
2010	97,330	111,616	87.2%
2011	97,509	113,438	85.9%
TOTAL	389,348	450,756	86.4%

A detail listing of Inpatient Discharge records linked to  $\geq 1$  lab records, by hospital and by calendar year(CY2008-2011) can be found in Appendix E.

## 2. Inpatient Discharges Without Labs

Table C.2 describes the top five reasons for hospitalization where the discharge record could not be linked to a lab record for CY 2008-2011. Of the top five hospitalizations, almost 1 in 2 lab records were not received for Newborns and Mental Health Diseases and Disorders. Inquiries to the labs and hospitals are currently being performed as to the explanation of missing lab records for the hospitalizations listed below.

A detail listing of Inpatient Discharge records without labs by hospital and by calendar year (CY2008-2011) can be found in Appendix F.

**Table C.2 – Discharges without Lab Records: Top Five Reasons for Hospitalization**

Reason for Hospitalization	Discharges without Lab Records	Total Discharges	Overall Pct	Overall Pct by Hospitalization
Newborns	31,656	66,051	7.0%	47.9%
Mental Diseases and Disorders	6,808	15,317	1.5%	44.4%
Diseases and Disorders of the Musculoskeletal System and Connective Tissue	3,982	32,224	0.9%	12.4%
Pregnancy, Childbirth and the Puerperium	2,303	69,764	0.5%	3.3%
Diseases and Disorders of the Respiratory System	2,159	37,700	0.4%	5.7%
Miscellaneous Reasons	14,500	229,700	3.2%	6.3%
			<b>13.6%</b>	

<sup>3</sup> Total Discharges for participating hospitals (Refer to Deliverable 1: Appendix D)

### 3. Laboratory Records Linked to Inpatient Discharges

Table C.3 describes the transmitted laboratory records that were excluded from the study and the lab records that were linked to an inpatient discharge record:

**Table C.3 Transmitted Lab Records Excluded from Study**

<b>Labs</b>	<b>Lab Results</b>
Initially Transmitted (CY2008-2011)	30,668,969
Exclusions: Missing Identifiers (Molokai)	1,212,691
Exclusions: Records linked to discharge record outside the study period	2,467,067
Exclusions: Records Linked to ER Discharge	1,462,401
Exclusions: Not one of 32 labs Missing values for key lab variables Microbiology	5,303,874
Final Lab Record Count Linked to Inpatient Discharge Record	20,222,936

## II. Conclusion

Conformity was absent throughout the datasets from the participating facilities. Different datasets required different algorithms to link the laboratory data to a HHIC hospitalization discharge record. As a result, it required additional time and effort to customize the linking process for the facilities. Overall, key numeric clinical lab data were successfully linked to hospital administrative data.

### APPENDIX A – Field/Edit Specifications

Field	Edit Specification
Send Facility	Sending Facility is Required
Account Number	Account Number is Required
Medicare Record Number	MRN is Required
Birth Date	Birth Date is Required Birth Date cannot be greater than observation date Birth Date cannot be 120 years prior to observation date
Gender	Gender must be one of the following values (F, M, O or U)
Patient First Name	Patient First Name is Required
Patient Last Name	Patient Last Name is Required
Admission Date	Admission Date is Required If linked, admission date in file must be within 2 days of the admission date in discharge record
Discharge Date	If provided, must be greater than admission date
Patient Class	Patient Class must be one of the following values: I, E, O
Test Result	Test Result is Required
Observation Value	Observation values must be numeric
Observation Date	If linked, Observation Date must not be more than 3 days prior to admission date and must not be greater than discharge date If not linked,, Observation Date must not be more than 3 days prior to admission date.
Unit	Unit is required Must be acceptable unit for lab test/hospital
Reference Range	Reference Range is Required

**APPENDIX B: DATA ELEMENTS USED FOR DETERMINISTIC/PROBABILISTIC LINKING**

Data Variable
Sending Facility
Account Number
Medical Record Number
SSN
Date of Birth
Date of Admission/Lab Observation Date
Gender
Patient First Name (up to first space if multiple names were present)
Patient Last Name

**APPENDIX C: ITERATIONS FOR DETERMINISTIC/PROBABILISTIC LINKING OF LAB RECORDS TO INPATIENT DISCHARGE RECORD**

Iteration	Matched Records	% <sup>4</sup>	Account Number	Medical Record Number	Date of Birth	Gender	Admit Date	First Name	Last Name	Sending Facility
1	17,627,513	57.4	x	x	x	x	x	x	x	x
2	188,687	0.6	X	X	X	X	X		x	x
3	178,426	0.6	x	x	x	x	x			x
4	379,316	1.2		x	x	x	x	x	x	x
5	173,495	0.6	x		X	X	x	x	x	x
6	1,316,777	4.3	x	x	x	x		x	x	
7	123,751	0.4	x	x		x	x	x	X	x
8 <sup>5</sup>	103,877	0.3	x	x	x	x	x	x	x	x
9 <sup>6</sup>	131,094	0.4		x	x	x		x	x	x
TOTAL	20,222,936	65.8								

<sup>4</sup> % = Matches / "N" (N = 30,668,969; total number of submitted laboratory records)

<sup>5</sup> Probabilistic match on all 8 variables utilizing LinkageWiz

<sup>6</sup> Due to limited identifiers available for linking, deterministic linking was performed on five variables

**APPENDIX D: EXCLUDED RECORDS, BY REASON OF EXCLUSION**

<b>Exclusion Description</b>	<b>Count</b>
Test result is not one of the 32 labs of interest	4,023,862
Missing observation value	149,117
Missing observation date	36,293
Test result status is not "Final" or "Corrected"	11,333
Test Result is a Microbiology	1,083,269
<b>Total Exclusions, by Reason</b>	<b>5,303,874</b>

**APPENDIX E: DISCHARGE RECORDS WITH  $\geq 1$  LAB RECORD, BY HOSPITAL AND CALENDAR YEAR**

FACILITY ID	HOSPITAL NAME	2008			2009			2010			2011		
		Discharge with $\geq 1$ labs	TOTAL DISCHARGES	% D/C MATCHED TO LAB RECORD	Discharge with $\geq 1$ labs	TOTAL DISCHARGES	% D/C MATCHED TO LAB RECORD	Discharge with $\geq 1$ labs	TOTAL DISCHARGES	% D/C MATCHED TO LAB RECORD	Discharge with $\geq 1$ labs	TOTAL DISCHARGES	% D/C MATCHED TO LAB RECORD
120001	QUEEN'S	20,931	23,566	88.82	21,775	24,142	90.02	22,144	24,317	91.09	23,709	26,102	90.77
120002	MAUI	10,061	12,066	83.38	10,445	11,936	87.65	10,047	11,413	88.22	10,068	11,330	89.37
120005	HILO	7,578	8,683	87.27	7,524	8,401	89.66	6,946	8,002	86.89	6,702	8,322	86.64
120006	CASTLE	5,614	7,336	76.53	6,142	7,813	78.73	6,105	7,973	76.56	5,896	7,994	74.01
120010	HMC-EAST	3,242	3,407	95.16	2,613	2,711	96.39	2,407	2,486	96.82	1,895	2,005	80.51
120011	KAISER	10,381	11,634	89.23	11,295	12,575	89.86	10,708	12,318	86.96	10,975	12,577	87.30
120014	WILCOX	4,453	4,605	96.70	4,350	4,468	97.45	4,040	4,139	97.68	3,808	4,104	74.22
120019	KONA	3,043	3,750	81.15	2,717	3,313	82.16	2,684	3,217	83.62	2,824	3,269	86.51
120022	STRAUB	5,527	6,293	87.83	6,128	6,558	93.49	6,231	6,527	95.56	6,427	6,751	73.60
120026	PALI MOMI	5,220	5,608	93.08	5,845	6,041	96.76	6,165	6,342	97.19	6,151	6,346	73.86
120027	HMC-WEST	3,317	3,412	97.22	3,453	3,513	98.29	3,290	3,341	98.53	3,234	3,300	76.60
120028	NORTH HAWAII	2,008	2,872	69.92	1,697	2,388	71.19	1,866	2,454	76.20	2,047	2,640	77.57
121300	KAUAI VETS	927	1,262	73.45	920	1,210	76.20	936	1,146	81.85	918	1,143	80.58
121301	KAU	9	9	100.00	18	23	78.26	11	12	91.67	3	4	75.00
121302	KOHALA	8	9	88.89	3	4	75.00	0	0	0.00	0	1	0.00
121307	HALE HO'OLA HAMA	2	3	66.67	1	1	100.00	4	7	57.14	12	12	83.33
121308	KULA HOSPITAL	5	6	83.33	5	5	100.00	4	5	80.00	5	5	100.00
123300	KAPIOLANI	13,470	18,021	74.75	13,782	18,058	76.52	13,742	17,917	77.03	12,835	17,533	57.85
<b>TOTAL</b>		<b>95,796</b>	<b>112,542</b>	<b>85.1</b>	<b>98,713</b>	<b>113,160</b>	<b>87.2</b>	<b>97,330</b>	<b>111,616</b>	<b>87.2</b>	<b>97,509</b>	<b>113,438</b>	<b>85.96</b>



**APPENDIX F: DISCHARGE RECORDS WITH NO LAB RECORD, BY HOSPITAL AND CALENDAR YEAR**

FACILITY ID	HOSPITAL NAME	2008			2009			2010			2011		
		Discharges without labs	TOTAL DISCHARGES	% D/C with No LAB RECORD	Discharges without labs	TOTAL DISCHARGES	% D/C with No LAB RECORD	Discharges without labs	TOTAL DISCHARGES	% D/C with No LAB RECORD	Discharges without labs	TOTAL DISCHARGES	% D/C with No LAB RECORD
120001	QUEEN'S	2,635	23,566	11.18	2,367	24,142	9.80	2,173	24,317	8.94	2,393	26,102	9.17
120002	MAUI	2,005	12,066	16.62	1,491	11,936	12.49	1,367	11,413	11.98	1,262	11,330	11.14
120005	HILO	1,105	8,683	12.73	877	8,401	10.44	1,056	8,002	13.20	1,620	8,322	19.47
120006	CASTLE	1,722	7,336	23.47	1,671	7,813	21.39	1,868	7,973	23.43	2,098	7,994	26.24
120010	HMC-EAST	165	3,407	4.84	98	2,711	3.61	79	2,486	3.18	110	2,005	5.49
120011	KAISER	1,253	11,634	10.77	1,280	12,575	10.18	1,611	12,318	13.08	1,602	12,577	12.74
120014	WILCOX	152	4,605	3.30	118	4,468	2.64	99	4,139	2.39	296	4,104	7.21
120019	KONA	707	3,750	18.85	596	3,313	17.99	534	3,217	16.60	445	3,269	13.61
120022	STRAUB	766	6,293	12.17	430	6,558	6.56	296	6,527	4.54	324	6,751	4.80
120026	PALI MOMI	388	5,608	6.92	196	6,041	3.24	177	6,342	2.79	195	6,346	3.07
120027	HMC-WEST	95	3,412	2.78	60	3,513	1.71	51	3,341	1.53	66	3,300	2.00
120028	NORTH HAWAII	864	2,872	30.08	691	2,388	28.94	588	2,454	23.96	593	2,640	22.46
121300	KAUAI VETS	335	1,262	26.55	290	1,210	23.97	210	1,146	18.32	225	1,143	19.69
121301	KAU	0	9	0.00	5	23	21.74	1	12	8.33	1	4	25.00
121302	KOHALA	1	9	11.11	1	4	25.00	0	0	0.00	1	1	100.00
121307	HALE HO'OLA HAMA	1	3	33.33	0	1	0.00	3	7	42.86	0	12	0.00
121308	KULA HOSPITAL	1	6	16.67	0	5	0.00	1	5	20.00	0	5	0.00
123300	KAPIOLANI	4,551	18,021	25.25	4,276	18,058	23.68	4,172	17,917	23.29	4,698	17,533	26.80
<b>TOTAL</b>		<b>16,746</b>	<b>112,542</b>	<b>14.9</b>	<b>14,447</b>	<b>113,160</b>	<b>12.8</b>	<b>14,286</b>	<b>111,616</b>	<b>12.8</b>	<b>15,929</b>	<b>113,438</b>	<b>14.04</b>

### III. Signatures

**Prepared by:** \_\_\_\_\_

Position Title: \_\_\_\_\_

Date: \_\_\_\_\_

#### **Approvals:**

Project Manager: \_\_\_\_\_

Date: \_\_\_\_\_

Co-Principal Investigator: \_\_\_\_\_

Date: \_\_\_\_\_

