DATA MANAGEMENT MANUAL

Health Professional License and Supplemental Survey Data

SCHOOL OF MEDICINE

BOWEN CENTER FOR HEALTH
WORKFORCE RESEARCH & POLICY

Version Date: February 9, 2021

Table of Contents

Introduction

Since 2015, the Bowen Center for Health Workforce Research and Policy (Bowen Center) has been commissioned by the State of Indiana to work in collaboration with the Indiana Professional Licensing Agency (IPLA) to draft and administer surveys which are administered electronically to licensed health professional during their license renewal period. These surveys collect seminal information regarding demographics, education & training, services provided, populations served, practice location and practice characteristics.

After administration of these supplemental surveys, data collected after the license renewal period is to be processed and stored in a longitudinal SQL database. The data are then used for reporting, data visualization and informing health policy and research. The coordination of this data management process has involved collaborative efforts from the Bowen Center and the Department of Biostatistics at Indiana University.

This manual outlines the procedures for storing, coding and formatting the license and supplemental survey data for the health professions data. The status of all health professions data should be tracked in the Health Professions Inventory Logbook on the J Drive. Changes to these procedures are reviewed and approved by the Bowen Center Research Analyst.

I. Storage of licensure and survey data files from IPLA

    a. License and supplemental survey data files are periodically transferred from IPLA to the Bowen Center using WINSCP. The schedule for these data transfers are predetermined by the IPLA. Current schedules for data extracts can be found in the following pathway: K:\FAMM\HealthWorkforceStudies\Evaluation (Current Projects)\AHEC - ISDH\Final Survey Instruments\Schedules

        i. On the scheduled extract data, an email from the IPLA will notify the Bowen Center when the most recent license and survey data files are ready for secure transfer.

        ii. IPLA will provide information on the licensed professionals included in the data files and instructions for accessing the data files on WINSCP.

    b. Original data files will be stored in the HWS Data folder under PLA Source Data, inside a new folder with the title "PLA Data Extract [Date of Data Extract]".

        i. Only designated Bowen Center team members will have access to the HWS Data fold and be able to transfer the licensure and survey data files.

II. Formatting Survey Data

    a. The raw survey data file contains survey data for all health professionals who renewed their license online. Each health professional will have one observation per question asked in the online survey.

    b. Import data into SAS and create a raw survey data set for a specific health professional by restricting by the license prefix of the specific health professional. See Appendix A for list of license prefixes for licensed Indiana health professionals.

    c. Transpose the survey question answers to create a data set in which each health professional had one observation. The transposed data set will contain the following variables:

      i.   Unique license number
      ii.   Profession ID
      iii. Profession Name
      iv. License Type ID
      v.   License Type Name
      vi. Health Professional's Full Name
      vii. Health Professional's Sort Name (Last Name, First Name)
      viii. Create Date
      ix. Update Date
      x.   Survey Answers (number of questions vary by health profession)

* The survey data file contains two license status variables (License Status ID and License Status Name) which should not be used. These variables reflect the license status at the time the survey was created, not the status of the license after it has been renewed. Only the license status variable in the license file should be used for the final licensure survey data. *

III. Coding Survey Data

a. All fields in the original survey data file are formatted as text fields and will need to be recoded and formatted. Each survey questions will be coded according to the coding definitions outlined in the corresponding health professional data dictionary. Coding schemes of variables commonly used in licensure surveys are available in Appendix B.

b. Reported practice hours will be used to determine practice FTE for all practice locations. Below is a table which outlines this conversion:

Hours per Week in Patient Care	FTE Conversion
0 hours per week	0.0
1 – 4 hours per week	0.1
5 – 8 hours per week	0.2
9 – 12 hours per week	0.3
13 – 16 hours per week	0.4
17 – 20 hours per week	0.5
21 – 24 hours per week	0.6
25 – 28 hours per week	0.7
29 – 32 hours per week	0.8
33 – 36 hours per week	0.9
37 – 40 hours per week	1.0
41 or more hours per week	1.0

c. Create the variable Respondent to indicate whether the health professional responded to the online survey.
    i. Health professional answered at least one survey question: Respondent=1
    ii. Health professional did not answer any survey question: Respondent=0

d. Quality Assurance
    i. Before dropping the original text answers from the new survey dataset, it is recommended to run a quality check on coding schemes by utilizing the PROC FREQ procedures in SAS.
    ii. Running a frequency table to compare the distribution of answers between the original text variable and the new coded variable. If there are discrepancies, ensure that spelling and formatting are corrected where necessary.

IV. Importing License Data

    a. The license data file contains the unique license number, mailing address, birth date, license status and issues and expiration dates for all Indiana health professionals. All fields in the original license data file are formatted as text fields and some will need to be reformatted in order to be used in the licensure survey data set.

    b. Import the license data into SAS and create a raw license data set for a specific health professional by restricting by the license prefix of the specific health professional. See Appendix A for list of license prefixes for licensed Indiana health professionals.
        i. Data will also be restricted by the license suffix. Only licenses ending with "A" will be included in the license data set.

V. Merging and Formatting License and Survey Data Sets

    a. Merge the license and survey data set by unique license number

    b. If Respondent status is blank for a health professional *and* their license status is not active (sec_lic_status in 1, 15, 50) then remove the health professional from the new data set. These observations have not (or in some cases could not have) been renewed.

    c. Survey respondents with inactive licenses will remain in the data set. They will be removed when the inclusion and exclusion criteria is applied to create a survey sample for the data report. Appendix C outlines the inclusion and exclusion criteria for creating the survey sample.

    d. Formatting Date Variables
        i. There are five date variables (Create_Date, Update_date, Date_of_Birth, Issue_Date and Expiration_Date) which are formatted as YYYYMMDD through read into SAS as character strings.
        ii. A second variable will need to be created through concatenation so that the date has a format of MM/DD/YYYY.
        iii. After creating these variables, use the input command to then create numeric variables from the second date variable and formatted as mmddyy10.
        iv. After creating the new date variables, drop the original date variables and secondary character variables from the dataset.

    e. Create a new variable called SurveyStatus to indicate the survey status.
        i. If Respondent is equal to 1, then SurveyStatus equals "Respondent".
        ii. If Respondent is equal to 0, then SurveyStatus equals "Non-Respondent".
        iii. If Respondent is blank then, SurveyStatus equals "Offline".
        iv. Once the survey status variable has been created, drop the Respondent variable from the dataset.

    f. Formatting and Finalizing Datasets
        i. Using PROC FORMAT create a format catalog for each coded variable. The catalog should be saved in the permanent library for the health profession (this is usually a SAS folder in the license renewal year folder for this dataset).
        ii. When creating the permanent SAS dataset, reorder the coded and formatted variables according to the associated data dictionary.
        iii. Use the formats catalog to apply the appropriate format to each coded variable.

VI. Storage in the Longitudinal Database

a. After finalizing and formatting the permanent SAS dataset export a copy to an Excel file in the J drive. Health profession license and supplemental survey data are saved in the following folder: \\10.234.189.104\Extranet$\Health Workforce\Projects\PLA

b. A copy of the finalized codebook should also be saved to the J drive. The database administrator will update codebook with database mapping after importing the data into the SQL database.

VII. Address Cleaning and Geocoding

a. After data are transferred into the import database, the database administrator will export the practice and license address data to an Excel file for address cleaning and/or geocoding.

b. When the data include address, city, state and ZIP code, this file will be transferred to Melissa Services, Inc. using the SFTP site. The vendor will notify the Bowen Center when the address data has been cleaned and geocoded.

c. When the practice data only include state, county and ZIP Code, then data will be cleaned and standardized by both a research assistant and coding used in the database. Cleaning will include correcting missing data where necessary (e.g. correcting the state field when ZIP code and/or county are provided, including county where state and ZIP code are available, etc.)

d. After address cleaning and geocoding is completed, the new geographical information (match level, census tract and county assignment) data will be imported into the relational database.

Appendix A: Health Professional License Prefix

Health Profession	License Prefix
Physician: Allopathic	01
Physician: Osteopathic	02
Physical Therapist (PT)	05
Podiatrist	07
Chiropractor	08
Physician Assistant (PA)	10
Dentist	12
Dental Hygienist	13
Hearing Aid Dealer	17
Optometrist	18
Psychologist	20
Speech Pathologist	22
Audiologist	23
Pharmacist	26
Licensed Practical Nurse (LPN)	27
Registered Nurse (RN)	28
APN Prescriptive Authority	71
Occupational Therapist	31
Social Worker	33
Clinical Social Worker	34
Marriage & Family Therapist	35
Mental Health Counselor	39
Marriage & Family Associate	85
Addiction Counselor Associate	869
Addiction Counselor	860
Clinical Addiction Counselor Associate	879
Clinical Addiction Counselor	870
Mental Health Associate	88

Appendix B: Coding Scheme for Commonly Used Variables

Variable(s)	Category	Coding
Sex	Male	1
	Female	2
Hispanic_Latino	Yes	1
	No	2
Race	White	1
	American Indian or Alaska Native	2
	Native Hawaiian or Other Pacific Islander	3
	Black or African American	4
	Asian	5
	Multiracial (more than one race selected)	6
QualifyingDegreeState, MedicalDegreeState, ResidencyState	Indiana	1
	Michigan	2
	Illinois	3
	Kentucky	4
	Ohio	5
	Another State (not listed)	6
	Another Country (not U.S.)	7
PrimaryPracticeHours, SecondPracticeHours, ThirdPracticeHours, PrimaryHoursWorked, SecondHoursWorked, PrimaryHoursTotal, SecondHoursTotal, PrimaryHoursPtCare SecondHoursPtCare	0 hours per week	1
	1 – 4 hours per week	2
	5 – 8 hours per week	3
	9 – 12 hours per week	4
	13 – 16 hours per week	5
	17 – 20 hours per week	6
	21 – 24 hours per week	7
	25 – 28 hours per week	8
	29 – 32 hours per week	9
	33 – 36 hours per week	10
	37 – 40 hours per week	11
	41 or more hours we week	12