

Assessing Newborn Screening Demographic Data Quality Using Correlations with Analytes

Travis Henry, Ashley Comer, Alankar Kampoowale,
Stanton Berberich

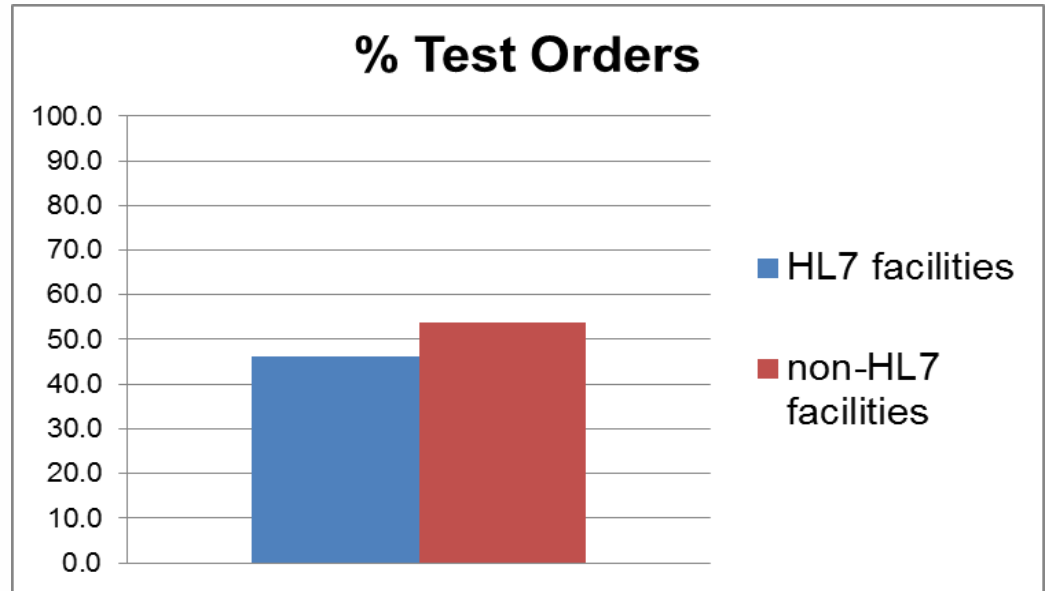
State Hygienic Laboratory at the University of Iowa

2019 APHL Newborn Screening and Genetic Testing Symposium

Chicago, IL
April 8, 2018

Background

- Approx. 45% of Iowa's test requests and associated demographic data are received using HL7 messaging



Background

- There are discrepancies between data provided electronically versus the paper collection form
- Data entry has adopted rules for conflicting data

The image shows a detailed paper form for blood specimen collection. It includes fields for patient information, specimen type, and collection site. A barcode is visible on the right side. At the bottom, there are two boxes labeled "PLACE THE BFL LABEL 'on Specimen Container'" and "FOR USE ONLY".

DO NOT REMOVE THIS COVER FLAP. IT IS FOR THE PROTECTION OF THE SPECIMEN AND THE SPECIMEN HANDLER.

PLEASE MAKE SURE THAT THE BLOOD SPECIMEN IS COMPLETELY FULL.

AND PROTECT BFL FLAP IS IN PLACE BEFORE SUBMITTING SPECIMEN.

Use and touch sample area only and use it straight.



VS



Background

- Is there a way to assess data accuracy received either electronically or by paper?

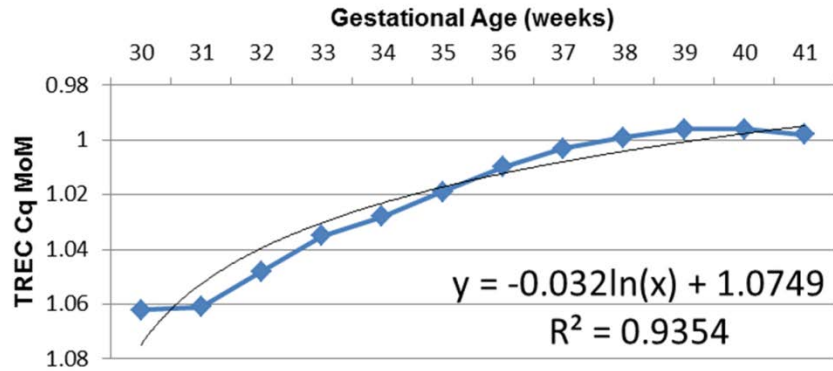


Analyte Correlations

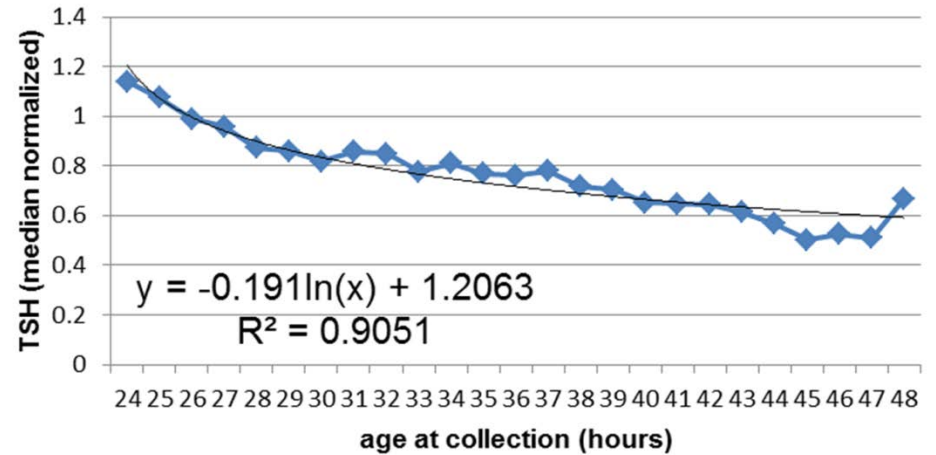
- Known associations between demographic data and analytes
- Built in “standard curve” based on population level data to assess possible outliers
- Gestational age and TREC
- Age at time of collection and TSH

Analyte Correlations

Gestational Age and TREC Cq MoM



Age at collection and TSH



Purpose

- Can we use known analyte and demographic data associations to identify outliers in demographic data?
- After we identify outliers, are there similarities in the outliers that can help us understand possible errors in demographic data?
- For example, Iowa receives 45% of data from facilities using HL7 messaging; could that be a source of error?

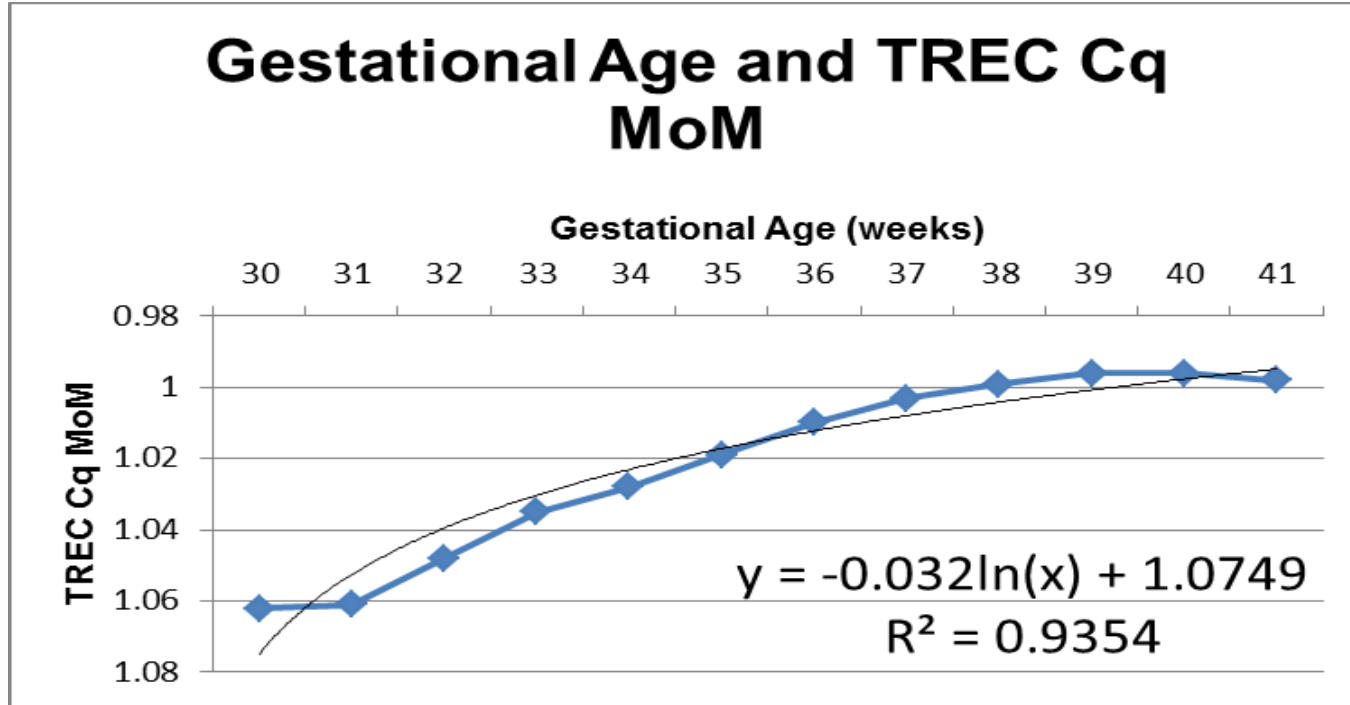
Methods

- Prior to statistical analysis, normalization and/or scaling is needed to account for batch effects and test dynamic range
- Iowa currently median normalizes and reports TREC Cq
- TSH data was ordered chronologically and analyte values were median normalized using a 1000 specimen window

Methods

- Assessment of outliers was completed using the median and interquartile range (IQR)
- Outliers were identified as anything less than or greater than 1.5 times the IQR when compared to the 25th and 75th percentiles
- Robust for identification of outliers

GA and TREC



GA and TREC

Gestational Age (weeks)	30	31	32	33	34	35	36	37	38	39	40	41
Number of values	67	96	137	230	427	633	1402	3231	5920	14952	8027	2253
Minimum	0.97	0.931	0.803	0.887	0.843	0.794	0.834	0.71	0.734	0.663	0.768	0.693
25% Percentile	1.017	1.04	1.024	1.012	1.007	0.9965	0.9908	0.984	0.981	0.978	0.979	0.98
Median	1.062	1.061	1.048	1.035	1.028	1.019	1.01	1.003	0.999	0.996	0.996	0.998
75% Percentile	1.096	1.085	1.072	1.059	1.052	1.041	1.03	1.022	1.018	1.015	1.016	1.016
Maximum	1.215	1.156	1.212	1.218	1.183	1.161	1.181	1.181	1.186	1.383	1.163	1.14
Interquartile range (IQR)	0.079	0.045	0.048	0.047	0.045	0.0445	0.0392	0.038	0.037	0.037	0.037	0.036
25% Percentile - 1.5 IQR	0.8985	0.9725	0.952	0.9415	0.9395	0.92975	0.932	0.927	0.9255	0.9225	0.9235	0.926
75% Percentile + 1.5 IQR	1.2145	1.1525	1.144	1.1295	1.1195	1.10775	1.0888	1.079	1.0735	1.0705	1.0715	1.07

GA and TREC

Gestational Age (weeks)	30	31	32	33	34	35	36	37	38	39	40	41
Number of values	67	96	137	230	427	633	1402	3231	5920	14952	8027	2253
Minimum	0.97	0.931	0.803	0.887	0.843	0.794	0.834	0.71	0.734	0.663	0.768	0.693
25% Percentile	1.017	1.04	1.024	1.012	1.007	0.9965	0.9908	0.984	0.981	0.978	0.979	0.98
Median	1.062	1.061	1.048	1.035	1.028	1.019	1.01	1.003	0.999	0.996	0.996	0.998
75% Percentile	1.096	1.085	1.072	1.059	1.052	1.041	1.03	1.022	1.018	1.015	1.016	1.016
Maximum	1.215	1.156	1.212	1.218	1.183	1.161	1.181	1.181	1.186	1.383	1.163	1.14
Interquartile range (IQR)	0.079	0.045	0.048	0.047	0.045	0.0445	0.0392	0.038	0.037	0.037	0.037	0.036
25% Percentile - 1.5 IQR	0.8985	0.9725	0.952	0.9415	0.9395	0.92975	0.932	0.927	0.9255	0.9225	0.9235	0.926
75% Percentile + 1.5 IQR	1.2145	1.1525	1.144	1.1295	1.1195	1.10775	1.0888	1.079	1.0735	1.0705	1.0715	1.07

GA and TREC

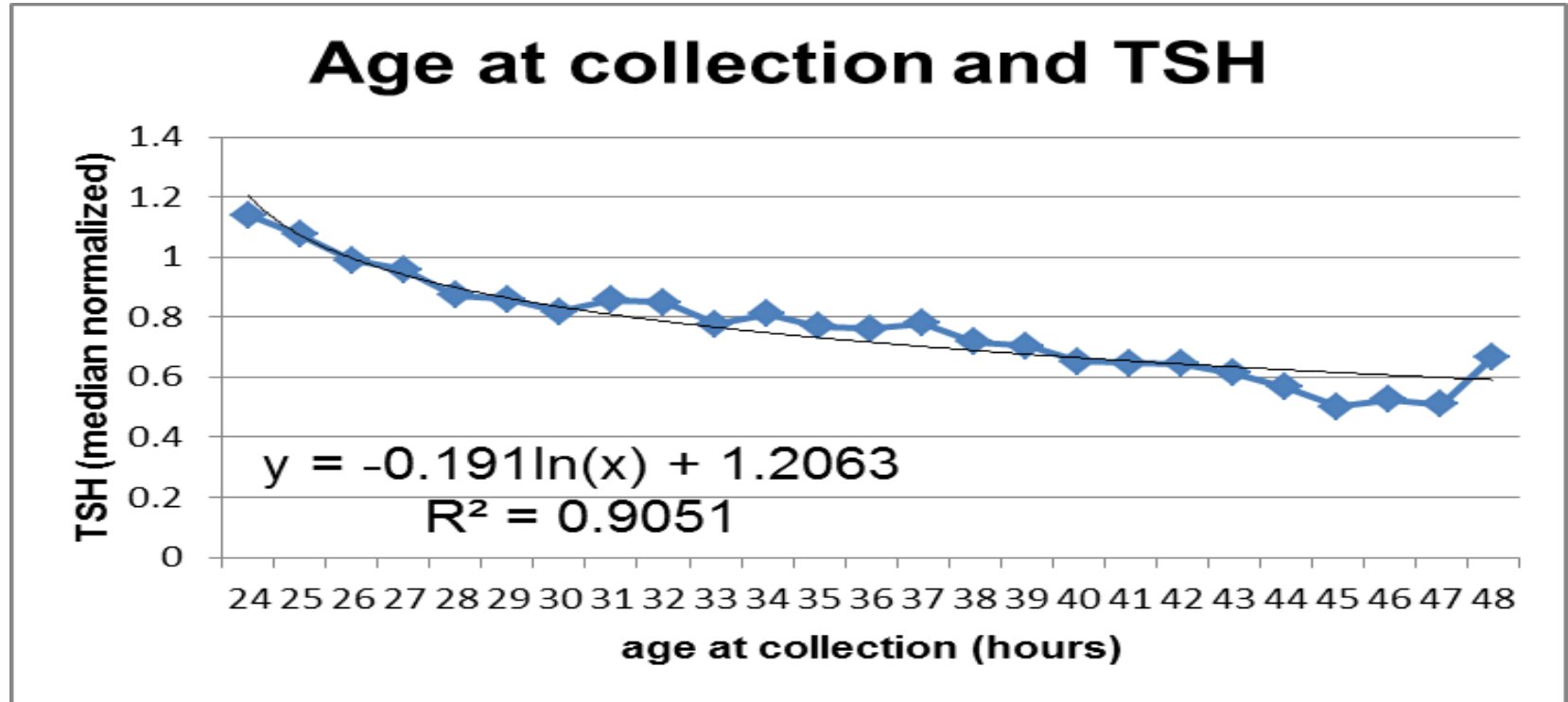
Gestational Age (weeks)	37	38	39
Number of values	3231	5920	14952
Minimum	0.71	0.734	0.663
25% Percentile	0.984	0.981	0.978
Median	1.003	0.999	0.996
75% Percentile	1.022	1.018	1.015
Maximum	1.181	1.186	1.383
Interquartile range (IQR)	0.038	0.037	0.037
25% Percentile - 1.5 IQR	0.927	0.9255	0.9225
75% Percentile + 1.5 IQR	1.079	1.0735	1.0705

GA and TREC

Gestational Age (weeks)	38	population
Number of values	5920	
Minimum	0.734	
25% Percentile	0.981	
Median	0.999	
75% Percentile	1.018	
Maximum	1.186	
Interquartile range (IQR)	0.037	% values </>
25% Percentile - 1.5 IQR	0.9255	1.27
75% Percentile + 1.5 IQR	1.0735	1.28

GA	Facility ID	TREC Cq MoM
38	169	1.069
38	169	1.069
38	169	1.074
38	169	1.075
38	169	1.083
38	169	1.088
38	169	1.091
38	169	1.095
38	169	1.098
38	169	1.124
38	170	0.841
38	170	0.943
38	170	0.945
38	170	0.95
38	170	0.95
38	170	0.952
38	170	0.953
38	170	0.954

Age at time of collection and TSH



Age at time of collection and TSH

Age at time of collection (hours)	24	25	26	27	28	29	30
Number of values	15536	6396	3204	1891	1365	1052	854
Minimum	0.1964	0.1947	0.1982	0.1982	0.1982	0.2037	0.1947
25% Percentile	0.8319	0.7885	0.7284	0.6952	0.6316	0.6337	0.5885
Median	1.139	1.074	0.9903	0.9581	0.8739	0.8586	0.8189
75% Percentile	1.495	1.418	1.327	1.272	1.192	1.191	1.127
Maximum	51.55	27.56	44.11	46.3	4.212	4.755	5.968
Interquartile range (IQR)	0.6631	0.6295	0.5986	0.5768	0.5604	0.5573	0.5385
25% Percentile - 1.5 IQR	-0.16275	-0.15575	-0.1695	-0.17	-0.209	-0.20225	-0.21925
75% Percentile + 1.5 IQR	2.48965	2.36225	2.2249	2.1372	2.0326	2.02695	1.93475

Age at time of collection and TSH

Age at time of collection (hours)	24	25	26	27	28	29	30
Number of values	15536	6396	3204	1891	1365	1052	854
Minimum	0.1964	0.1947	0.1982	0.1982	0.1982	0.2037	0.1947
25% Percentile	0.8319	0.7885	0.7284	0.6952	0.6316	0.6337	0.5885
Median	1.139	1.074	0.9903	0.9581	0.8739	0.8586	0.8189
75% Percentile	1.495	1.418	1.327	1.272	1.192	1.191	1.127
Maximum	51.55	27.56	44.11	46.3	4.212	4.755	5.968
Interquartile range (IQR)	0.6631	0.6295	0.5986	0.5768	0.5604	0.5573	0.5385
25% Percentile - 1.5 IQR	-0.16275	-0.15575	-0.1695	-0.17	-0.209	-0.20225	-0.21925
75% Percentile + 1.5 IQR	2.48965	2.36225	2.2249	2.1372	2.0326	2.02695	1.93475

Age at time of collection and TSH

Age at time of collection (hours)	24	population
Number of values	15536	
Minimum	0.1964	
25% Percentile	0.8319	
Median	1.139	
75% Percentile	1.495	
Maximum	51.55	
Interquartile range (IQR)	0.6631	% values >/<
25% Percentile - 1.5 IQR	-0.16275	0
75% Percentile + 1.5 IQR	2.48965	1.90%

Age at time of collection and TSH

Age at time of collection (hours)	24	population
Number of values	15536	
Minimum	0.1964	
25% Percentile	0.8319	
Median	1.139	
75% Percentile	1.495	
Maximum	51.55	
Interquartile range (IQR)	0.6631	% values >/<
25% Percentile - 1.5 IQR	-0.16275	0
75% Percentile + 1.5 IQR	2.48965	1.90%

Age at time of collection and TSH

Age at time of collection (hours)	24	population
Number of values	15536	
Minimum	0.1964	
25% Percentile	0.8319	
Median	1.139	
75% Percentile	1.495	
Maximum	51.55	
Interquartile range (IQR)	0.6631	% values >/<
25% Percentile - 1.5 IQR	-0.16275	0
75% Percentile + 1.5 IQR	2.48965	1.90%

facility ID	n above IQR	total	%	HL7?
3	7	206	3.40	no
11	5	193	2.59	no
17	73	2509	2.91	yes

Summary

- Implementing HL7 messaging can introduce additional QA considerations
- Assessing outliers can provide a snapshot of data quality (rare disorders)
- A significant increase in outliers over expected frequency may indicate a issue
- If an issue is identified, more investigation is needed
- Data analytics requires data pre-processing
 - Normalization
 - Determination of possible bias(es)

Acknowledgements



Iowa Department of Public Health

Department of Pediatrics, University of stead
Family Children's Hospital

Mission: The State Hygienic Laboratory at the University of Iowa protects and improves quality of life by providing reliable environmental and public health information through the collective knowledge and capabilities of our organization.