Literate Programming Tools for Continuous Quality Improvement

William H Beasley, David E Bard, & Thomas Wilson
OUHSC, Developmental and Behavioral Pediatrics

Background

Continuous Quality Improvement (CQI)

- Our project serves as the external evaluators for the Maternal, Infant, and Early Childhood Home Visiting (MIECHV) grants, administered by the OK State Dept of Health (OSDH).
- HRSA defines CQI as "a continuous process that employs rapid cycles of improvement." Accordingly, our data system is designed to provide fresh and frequent feedback to the OSDH administration, and its contracted agencies.
- This poster describes the tools used in our evaluation; all are freely available to both OUHSC researchers and OUHSC clinicians.

Literate Programming

– Combines text, graphs, and statistics in a coherent document that is more easily understood by human readers.
– Uses automation to eliminate the need to repeatedly copy & paste analytic results after underlying data sources are updated.

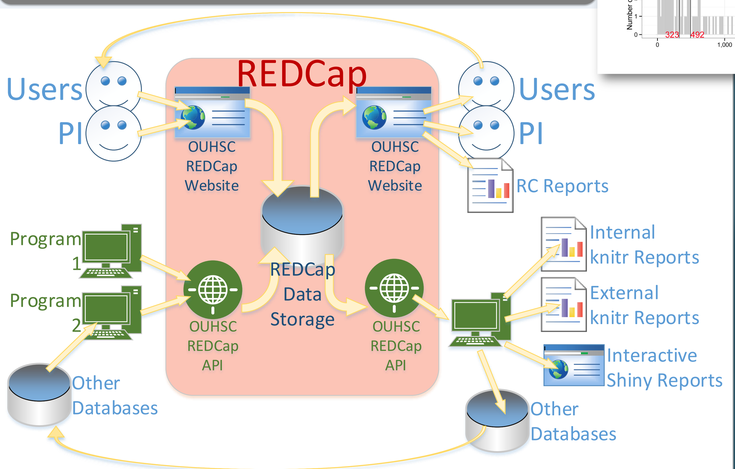
Reproducible Research

– Facilitates scientific replication.
– Disseminates techniques to other subfields.
– Promotes cumulative research.

Tools

Statistical Analysis & Data Manipulation: R & RStudio
Databases: REDCap & SQL Server
Report Creation & Delivery: knitr & Shiny (R packages)
Parallel Collaboration: GitHub and SharePoint

Workflows with REDCap and Other Databases



Collaboration and Communication
within these groups

The 4 statisticians on the project. Version control software (GitHub) allows parallel development of the analytical code. A multi-tiered data access layer allows substantial code reuse, which makes the software development more agile and efficient.

The 20 members of the OUHSC team. Centralized databases (REDCap and SQL Server) allow participant data to be collected and processed securely. Scheduling features within REDCap allow data collectors to manage appointments, and reschedule missed visits. Data collectors carry mobile 3G/4G hotspots to upload and download information during interviews.

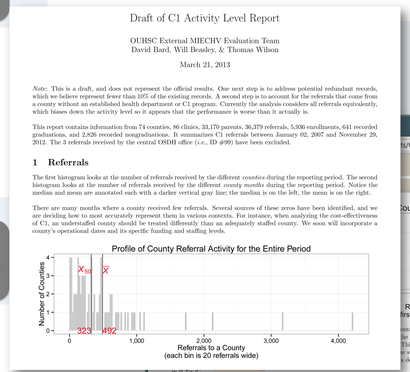
The 3 partnering organizations (OSDH, WIC, OHCA). Secure data exchanges allow us to receive their subject-level and agency-level data. Automation and literate programming allow us to distribute summary reports frequently.

Academics in different fields (particularly at OUHSC). Open-source and popular software tools allow us to more easily generalize the workflow to fields that are outside of developmental & behavioral pediatrics. Specifically, we adopt aspects of previous successful clinical research and CQI projects, as well as share the features of our study that might be well-suited for later projects.

Researchers in other states pursuing similar goals. Literate programming allows us to more easily create manuscripts and supplement publications with sanitized datasets that facilitate reproducible research.

Static Reports

When a user clicks the 'render' button, the report engine calls an underlying R script that (a) queries the database, (b) manipulates the data, and (c) runs the analyses. These analyses are returned to the reporting engine, which formats the results and creates a pdf for human consumption.


Draft of C1 Activity Level Report

OUHSC External MIECHV Evaluation Team
David Bard, Will Beasley, & Thomas Wilson

March 21, 2013

Note: This is a draft, and does not represent the official results. One next step is to address potential redundant records, which we believe represent fewer than 10% of the existing records. A second step is to account for the referrals that come from a county without an established health department or C1 program. Currently the analysis considers all referrals equivalently, which biases down the activity level so it appears that the performance is worse than it actually is.

This report contains information from 74 counties, 86 clinics, 33,170 parents, 36,379 referrals, 5,936 enrollments, 641 recorded graduations, and 2,826 recorded nongraduations. It summarizes C1 referrals between January 02, 2007 and November 29, 2012. The 3 referrals received by the central OSDH office (i.e., ID #99) have been excluded.

1 Referrals

The first histogram looks at the number of referrals received by the different *counties* during the reporting period. The second histogram looks at the number of referrals received by the different *county months* during the reporting period. Notice the median and mean are annotated each with a darker vertical gray line; the median is on the left, the mean is on the right.

There are many months where a county received few referrals. Several sources of these zeros have been identified, and we are deciding how to most accurately represent them in various contexts. For instance, when analyzing the cost-effectiveness of C1, an understaffed county should be treated differently than an adequately staffed county. We soon will incorporate a county's operational dates and its specific funding and staffing levels.

Profile of County Referral Activity for the Entire Period

Number of Counties

Referrals to a County (each bin is 20 referrals wide)

Number of Monthly Referrals for each County

Number of Referrals per Month

Date (binned by month)

Number of Monthly Referrals for each County (zoomed)

Number of Referrals per Month

Date

Referrals to a County in a month (first bin is '0'; each bin is 1 referral wide)

Identifying Referral Count for each CountyID

Number of Referrals during Reporting Period (logged)

CountyID

The next two graphs show the number of referrals received over time, by county. Each county has a unique line. Each boxplot denotes a month's 25%, 50% and 75% counts of referrals to a county. For instance, when the middle of a box plot is at $y = 7$, then 50% of the counties received 7 referrals or fewer.

The gray lines appear again as the median (on bottom) and mean (on top). A curvy longitudinal line summarizes the state's trend during the reporting period. The two graphs are identical, except the second one zooms in on the vertical.

Section B: Pregnancy

Current and Previous Pregnancies

Are you currently pregnant?

Yes 10%

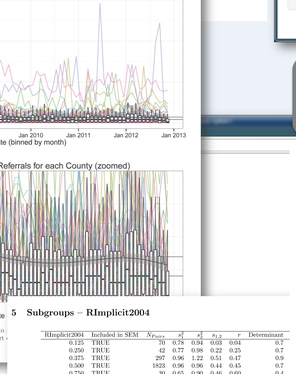
No 90%

Number of Responding Participants (n=73)

Weeks until due date

Number of Responding Participants (n=7)

Weeks


5 Subgroups – RImplicit2004

RImplicit2004	Included in SEM	N_{Pairs}	π_1^2	π_2^2	$\pi_{1,2}$	r	Determinant
0.125	TRUE	70	0.78	0.94	0.03	0.04	0.7
0.250	TRUE	42	0.77	0.96	0.22	0.25	0.7
0.375	TRUE	297	0.96	1.22	0.51	0.47	0.9
0.500	TRUE	1823	0.96	0.96	0.44	0.45	0.7
0.750	TRUE	30	0.65	0.90	0.46	0.60	0.4

Table 8: RImplicit2004 – Newer Version of Links

RImplicit2004	Included in SEM	N_{Pairs}	π_1^2	π_2^2	$\pi_{1,2}$	r	Determinant
0.125	TRUE	70	0.78	0.94	0.03	0.04	0.7
0.250	TRUE	42	0.77	0.96	0.22	0.25	0.7
0.375	TRUE	297	0.96	1.22	0.51	0.47	0.9
0.500	TRUE	1823	0.96	0.96	0.44	0.45	0.7
0.750	TRUE	30	0.65	0.90	0.46	0.60	0.4

Table 9: RImplicit2004 – Older Version of Links

6 Explanation of R Variants

Variant	Explanation
R	We recommend researchers typical use this version.
R_{Full}	The most complete version we have; doesn't exclude groups like R=0.
R_{Pass1}	Supposed to be fooled only by errors in the subject's/mother's knowledge
$RImplicit$	Uses only implicit items
$RImplicit_{Pass1}$	Uses only implicit items & supposed to be fooled only by knowledge errors
$RImplicit_{Mother}$	Uses only mother's implicit items (exists only for Gen2)
$RImplicit_{Subject}$	Uses only subject's implicit items
$RImplicit_{2004}$	The state of the links in 2004. Rodgers & Rowe for Gen1; Rodgers, Johnson &
$REExplicit$	Uses only explicit items
$REExplicit_{Pass1}$	Uses only explicit items & supposed to be fooled only by knowledge errors

More than 12 months 42%

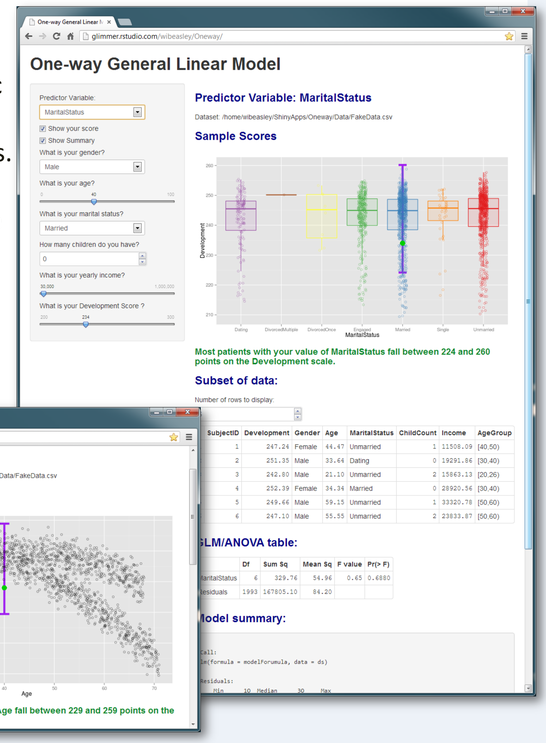
12 months or less

Number of Responding Participants (n=66)

Interactive Reports

Interactive web-based reports can be publically exposed if the content is exploratory in nature, and does not contain PHI. An interactive format allows readers to explore multidimensional spaces, more quickly than if all possibilities were contained in a large paper report.

Here is how a diagnostic prediction model could be presented to patients. Their manually-entered values are passed to the prediction engine, whose individualized results are presented with text, graphs, and tables.


One-way General Linear Model

Predictor Variable: MaritalStatus

Dataset: /home/wibeasley/ShinyApps/Oneway/Data/FakeData.csv

Sample Scores

Most patients with your value of MaritalStatus fall between 224 and 260 points on the Development scale.

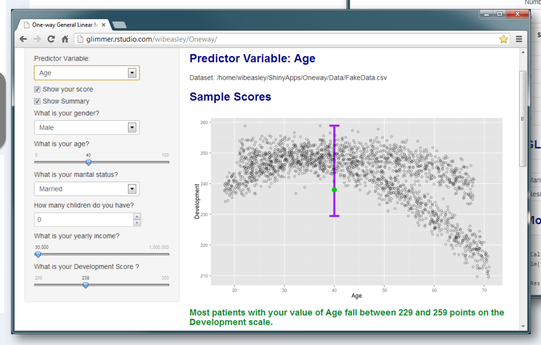
Subset of data:

SubjectID	Development	Gender	Age	MaritalStatus	ChildCount	Income	AgeGroup
1	247.24	Female	44.47	Unmarried	1	11508.09	[40,50)
2	251.35	Male	33.64	Dating	0	19291.86	[30,40)
3	242.80	Male	21.10	Unmarried	2	15663.13	[20,26)
4	252.39	Female	34.34	Married	0	28920.56	[30,40)
5	249.66	Male	59.15	Unmarried	1	33320.78	[50,60)
6	247.10	Male	55.55	Unmarried	2	29833.87	[50,60)

GLM/ANOVA table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
MaritalStatus	6	329.76	54.96	0.65	0.6880
Residuals	1993	167805.10	84.20		

Model summary:


One-way General Linear Model

Predictor Variable: Age

Dataset: /home/wibeasley/ShinyApps/Oneway/Data/FakeData.csv

Sample Scores

Most patients with your value of Age fall between 229 and 259 points on the Development scale.

Additional Software Details

Like most enterprise-level databases REDCap provides:

- User-specific authorization and HIPAA-compliant audit logs.
- A fault tolerant engine for improved data integrity.
- A programmatic interface that permits other software to input and extract data (once security credentials are authenticated).

However, REDCap is not a requirement for the CQI and reporting tools presented here. These tools are contained in R (an open source statistical software program), and can accommodate information from almost any database.

Further Information

If you have questions, or would like assistance applying this system (or a subset of its tools) to current or future projects, please contact: william-beasley@ouhsc.edu, thomas-wilson@ouhsc.edu, or david-bard@ouhsc.edu.