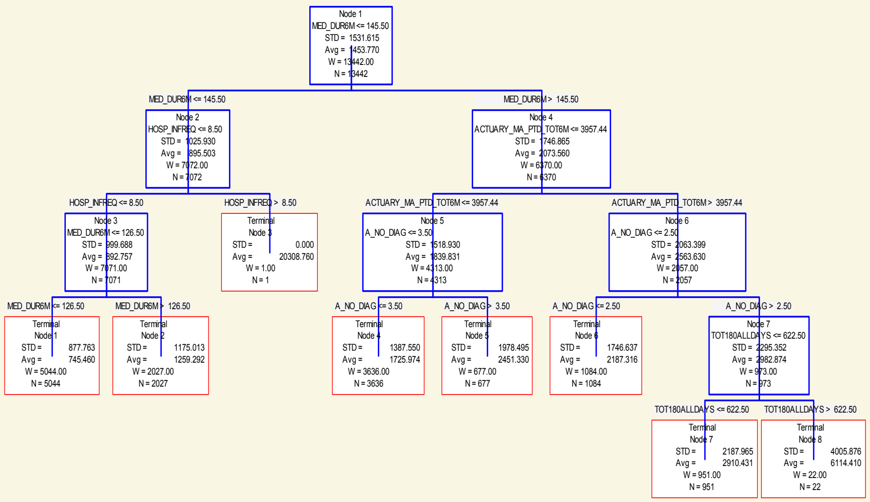
Washington State Department of Labor & Industries

Actuarial Services
Yuk W. Cheng, Joshua Ligosky and Bill Vasek



Actuarial Predictive Modeling

–

Early Case Reserve

22nd Jan., 2016

Outline

- What is predicted per claim?
- What type of data do we use?
- Important variables.
- How do we choose our models?
- What we cannot predict?
- How will we improve the prediction?
- What type of predictive analytic models do we use?
- Questions.

What is predicted per claim?

- Type of claim – e.g. PPD, Timeloss or Medical only
- Incurred cost of the claim at 21 months
- Eventual duration of time-loss and medical benefits.
- Note: Occurs each month (from 1 to 9 months)

What type of data do we use?

- Alpha-numeric information in our claims data systems.
  - Upfront injured worker information, e.g, Report Of Accident information, age, wage, medical diagnosis codes, etc.
  - Current status code of each open claim – [ppd, medical only, timeloss].
  - Current benefit amounts paid by type
  - Other available information as it becomes available, e.g., no. of medical visits, claim complexity, OIICS codes.

Important variables at different ages (in months)

var.imp.1m21m	var.imp.2m21m	var.imp.3m21m	var.imp.4m21m	var.imp.5m21m
clm_rptd_mult_job_flg (100%)	actuary_clm_stat_code2m (100%)	actuary_clm_stat_code3m (100%)	actuary_clm_stat_code4m (100%)	clm_rptd_mult_job_flg (100%)
actuary_clm_stat_code1m (73.49%)	actuary_af_ptd_tot2m (90.5%)	actuary_af_ptd_tot3m (90.3%)	actuary_tl_ptd_tot4m (89.84%)	clm_hrs_workd_per_day_qty (49.09%)
actuary_af_ptd_tot1m (61.73%)	actuary_tl_ptd_tot2m (90.42%)	actuary_tl_ptd_tot3m (90.11%)	actuary_af_ptd_tot4m (89.82%)	clm_day_per_wk_workd_qty (48.98%)
actuary_tl_ptd_tot1m (61.73%)	actuary_tl_day_pd_qty2m (86.25%)	actuary_tl_day_pd_qty3m (86.49%)	actuary_tl_day_pd_qty4m (86.93%)	actuary_clm_stat_code5m (47.7%)
clm_hrs_workd_per_day_qty (41.51%)	clm_rptd_mult_job_flg (51.6%)	clm_rptd_mult_job_flg (58.88%)	clm_rptd_mult_job_flg (67.9%)	clm_mo_wage_amt (47.11%)
clm_day_per_wk_workd_qty (40.62%)	clm_phycn_est_tl_day_qty (22.46%)	actuary_ma_ptd_tot3m (31.6%)	actuary_ma_ptd_tot4m (38.73%)	clm_mo_comp_rate_amt (39.81%)
actuary_tl_day_pd_qty1m (17.43%)	clm_cmplx_code (7.59%)	clm_hrs_workd_per_day_qty (10.31%)	clm_hrs_workd_per_day_qty (14.24%)	actuary_af_ptd_tot5m (37.1%)
clm_mo_comp_rate_amt (13.09%)	clm_day_per_wk_workd_qty (7.34%)	clm_mo_comp_rate_amt (9.39%)	clm_mo_comp_rate_amt (13.89%)	actuary_tl_ptd_tot5m (36.47%)
clm_mo_wage_amt (12.27%)	clm_mo_comp_rate_amt (7.25%)	clm_day_per_wk_workd_qty (8.54%)	clm_mo_wage_amt (13.86%)	actuary_tl_day_pd_qty5m (34.34%)
clm_cmplx_code (6.78%)	clm_mo_wage_amt (6.16%)	clm_mo_wage_amt (8.51%)	clm_day_per_wk_workd_qty (12.5%)	actuary_ma_ptd_tot5m (24.09%)
clm_phycn_est_tl_day_qty (6.46%)	clm_hrs_workd_per_day_qty (5.9%)	clm_cmplx_code (8.25%)	clm_cmplx_code (11%)	injry_body_part_code (11.99%)
clm_prity_flg (6.35%)	actuary_ma_ptd_tot2m (4.62%)	injry_body_part_code (6.2%)	injry_body_part_code (7.07%)	injry_oiics_body_part_code (10.98%)
med.dur1m (4%)	injry_body_part_code (4.18%)	injry_oiics_body_part_code (6.02%)	injry_oiics_body_part_code (6.93%)	hosp_amt (7.2%)
injry_body_part_code (3.88%)	injry_oiics_body_part_code (3.85%)	med.dur3m (4.29%)	med.dur4m (6.82%)	visits_within_5m (6.84%)
visits_within_1m (3.67%)	med.dur2m (2.24%)	injry_nat_code (2.4%)	visits_within_4m (4.29%)	clm_cmplx_code (5.39%)

var.imp.6m21m	var.imp.7m21m	var.imp.8m21m	var.imp.9m21m
clm_rptd_mult_job_flg (100%)	clm_rptd_mult_job_flg (100%)	clm_rptd_mult_job_flg (100%)	clm_rptd_mult_job_flg (100%)
clm_day_per_wk_workd_qty (48.88%)	actuary_clm_stat_code7m (48.96%)	actuary_clm_stat_code8m (48.48%)	actuary_clm_stat_code9m (54.25%)
clm_hrs_workd_per_day_qty (48.87%)	clm_day_per_wk_workd_qty (48.6%)	clm_day_per_wk_workd_qty (46.61%)	clm_day_per_wk_workd_qty (47.32%)
actuary_clm_stat_code6m (48.41%)	clm_hrs_workd_per_day_qty (48.46%)	clm_hrs_workd_per_day_qty (46.51%)	clm_hrs_workd_per_day_qty (47.12%)
clm_mo_wage_amt (46.97%)	clm_mo_wage_amt (46.21%)	clm_mo_wage_amt (44.96%)	actuary_af_ptd_tot9m (43.52%)
clm_mo_comp_rate_amt (40.98%)	actuary_af_ptd_tot7m (39.45%)	actuary_af_ptd_tot8m (43.49%)	clm_mo_wage_amt (42.74%)
actuary_tl_ptd_tot6m (33.95%)	clm_mo_comp_rate_amt (39.13%)	actuary_tl_ptd_tot8m (40.66%)	actuary_tl_ptd_tot9m (39.73%)
actuary_af_ptd_tot6m (33.11%)	actuary_tl_ptd_tot7m (37.48%)	actuary_tl_day_pd_qty8m (39.07%)	actuary_tl_day_pd_qty9m (38.47%)
actuary_tl_day_pd_qty6m (32.59%)	actuary_tl_day_pd_qty7m (36.15%)	clm_mo_comp_rate_amt (34.28%)	clm_mo_comp_rate_amt (36.7%)
actuary_ma_ptd_tot6m (24.2%)	actuary_ma_ptd_tot7m (28.49%)	actuary_ma_ptd_tot8m (31.54%)	actuary_ma_ptd_tot9m (31.84%)
injry_body_part_code (12.3%)	injry_body_part_code (12.58%)	injry_oiics_body_part_code (16.32%)	injry_oiics_body_part_code (16.02%)
injry_oiics_body_part_code (11.89%)	injry_oiics_body_part_code (11.72%)	injry_body_part_code (16.28%)	injry_body_part_code (15.91%)
med.dur6m (8.34%)	visits_within_7m (10.16%)	visits_within_8m (11.79%)	visits_within_9m (14.87%)
hosp_amt (7.18%)	med.dur7m (9.17%)	hosp_amt (10.3%)	hosp_amt (11.55%)
visits_within_6m (6.59%)	hosp_amt (7.43%)	med.dur8m (5.26%)	med.dur9m (6.55%)

How do we choose our models?

- Divide historical data into three sets: 1) training 2) validation and 3) test sets.
  – 1) Build models using the training set
  – 2) Use the validation data to find the optimal model.
  – 3) See how well the model can predict the test set.
- The validation errors gives an unbiased estimate of the predictive power of a model.

Estimated Accuracies based on the validation step

Age of claim (in month)	No. of claim used	Claim status								
		Fatality	TPD	PPD	TL	UC	MO	SAW	KOS	Total
1	54853	0 (0%)	0 (0%)	3105 (58.7%)	8250 (83.3%)	0 (0%)	36263 (96.81%)	323 (47.78%)	658 (48.67%)	48599 (88.6%)
2	72603	0 (0%)	0 (0%)	7668 (61.17%)	13280 (76.69%)	0 (0%)	36962 (95.63%)	355 (34.63%)	1697 (67.02%)	59962 (82.59%)
3	119761	0 (0%)	0 (0%)	17377 (70.52%)	23622 (76.06%)	0 (0%)	53992 (94.84%)	577 (34.65%)	2806 (63.28%)	98374 (82.14%)
4	47036	0 (0%)	0 (0%)	8542 (70.79%)	10272 (74.83%)	0 (0%)	16877 (92.66%)	222 (34.63%)	1177 (62.01%)	37090 (78.85%)
5	36289	0 (0%)	0 (0%)	10139 (76.26%)	6957 (64.64%)	36 (7.86%)	9026 (89.96%)	67 (14.92%)	608 (52.92%)	26833 (73.94%)
6	65573	0 (0%)	0 (0%)	19596 (77.68%)	12297 (64.55%)	81 (9.03%)	15261 (87.98%)	137 (17.91%)	1044 (50.51%)	48416 (73.84%)
7	29278	0 (0%)	0 (0%)	9488 (79.49%)	5203 (62.78%)	45 (10.25%)	6423 (88%)	79 (25%)	480 (52.23%)	21718 (74.18%)
8	25037	0 (0%)	0 (0%)	10405 (86.33%)	3530 (52.62%)	49 (12.01%)	4209 (85.67%)	71 (31.7%)	309 (52.11%)	18573 (74.18%)
9	45572	0 (0%)	0 (0%)	18821 (83.84%)	6631 (55.26%)	154 (20.42%)	7157 (82.56%)	162 (42.52%)	493 (44.86%)	33418 (73.33%)

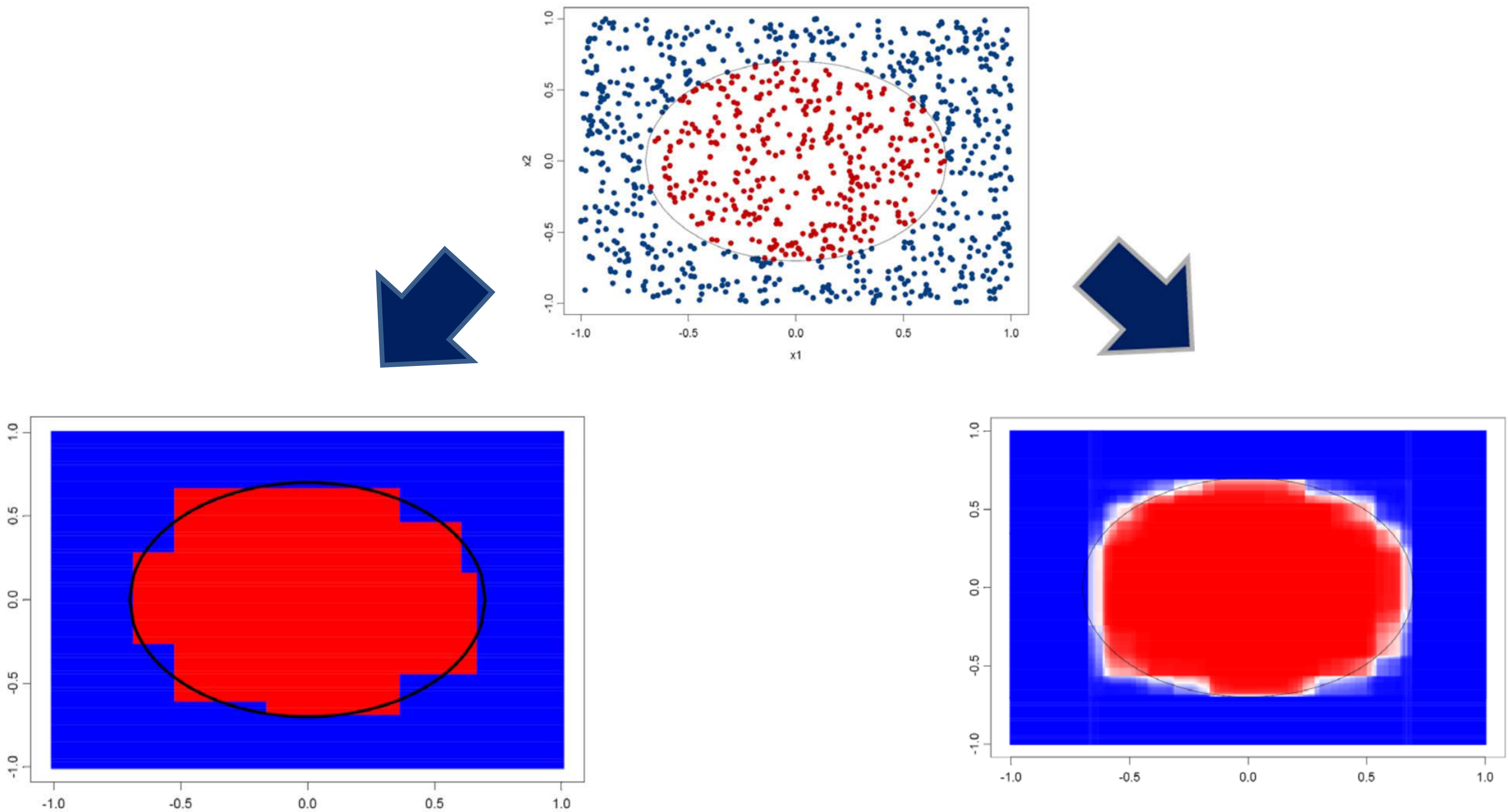
What we cannot predict?

- Accuracy of predicting fatality claims (0%).
- Accuracy of predicting TPD claims(0%).
- Accuracy of predicting PPD claims(59%-86%).
- Accuracy of predicting timeloss claims(55%-83%).
- Accuracy of predicting medical only claims(83%-97%).

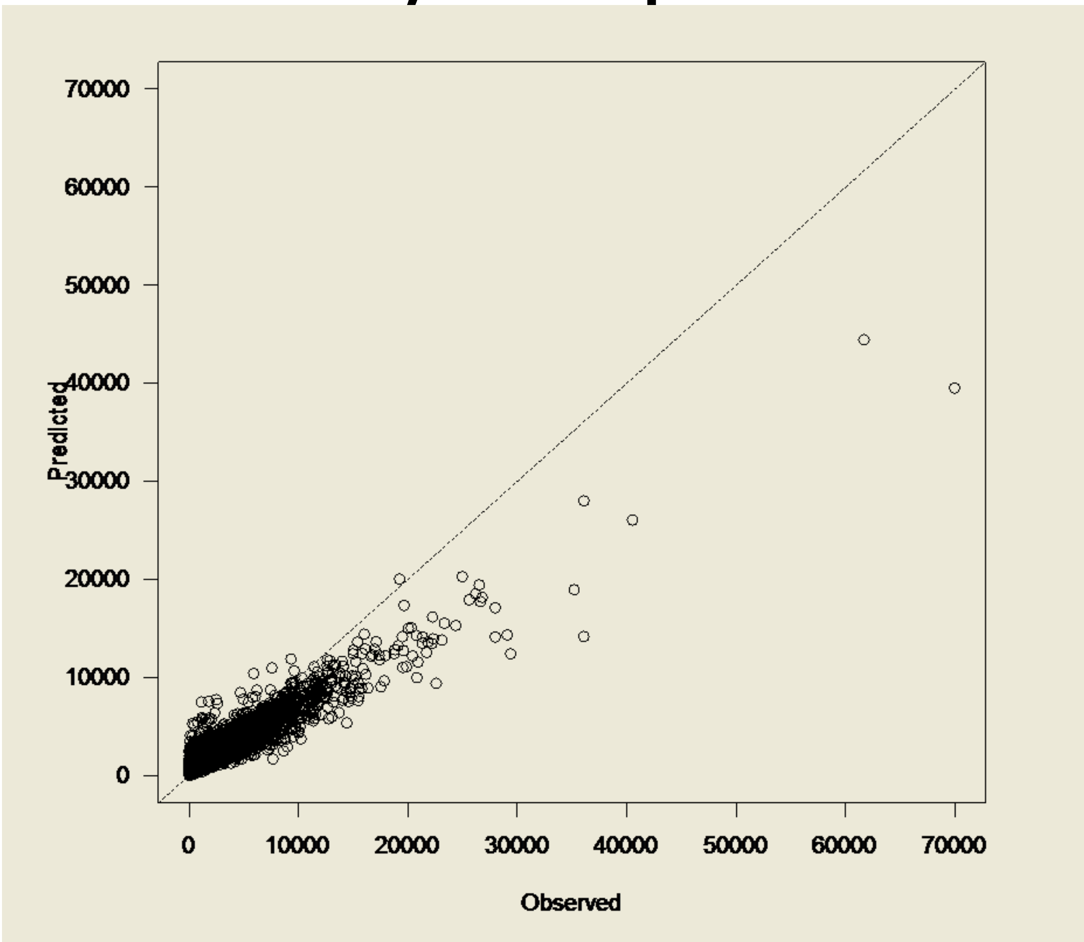
How will we improve the prediction?

- Update all the trained models in every quarter with new data.
- Incorporate additional predicted variables when the data is available, e.g., medical bill type, nursing consultant information.
- Update the test models more frequently when data is available.

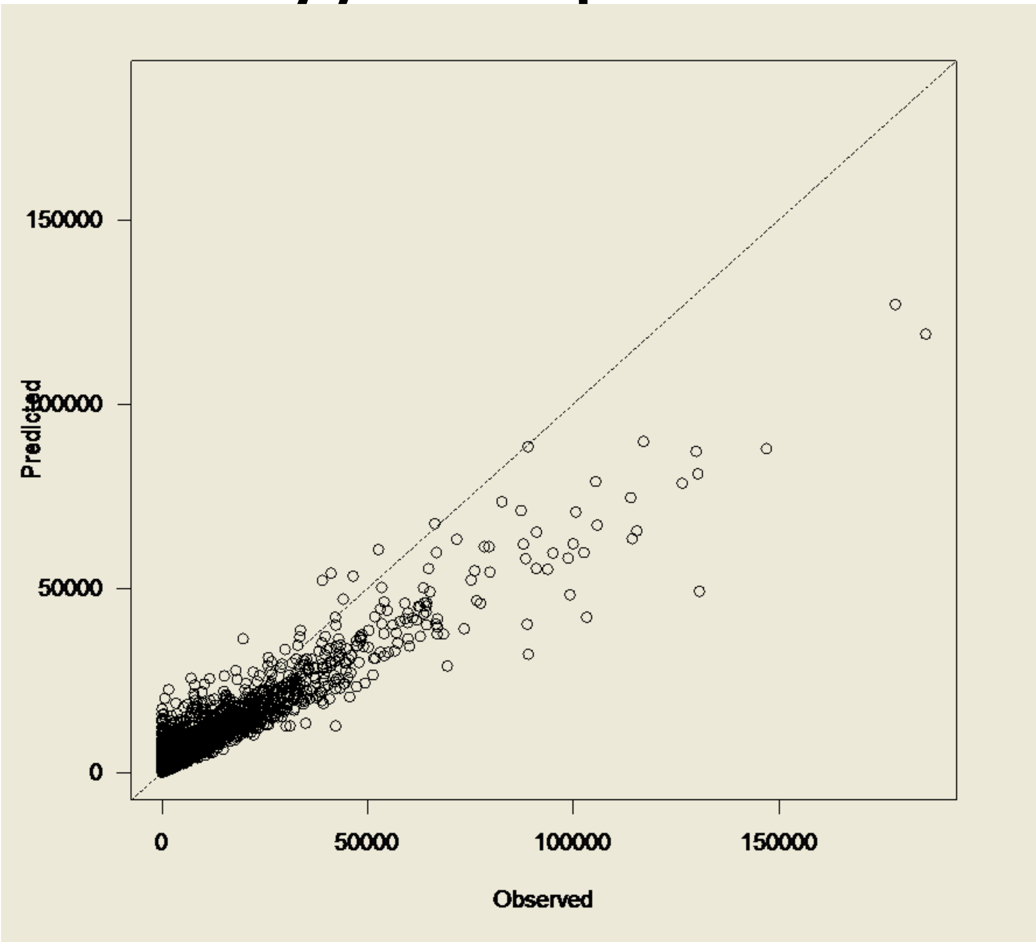
What type of Predictive Analytics models do we use? Example: CART and Random Forest



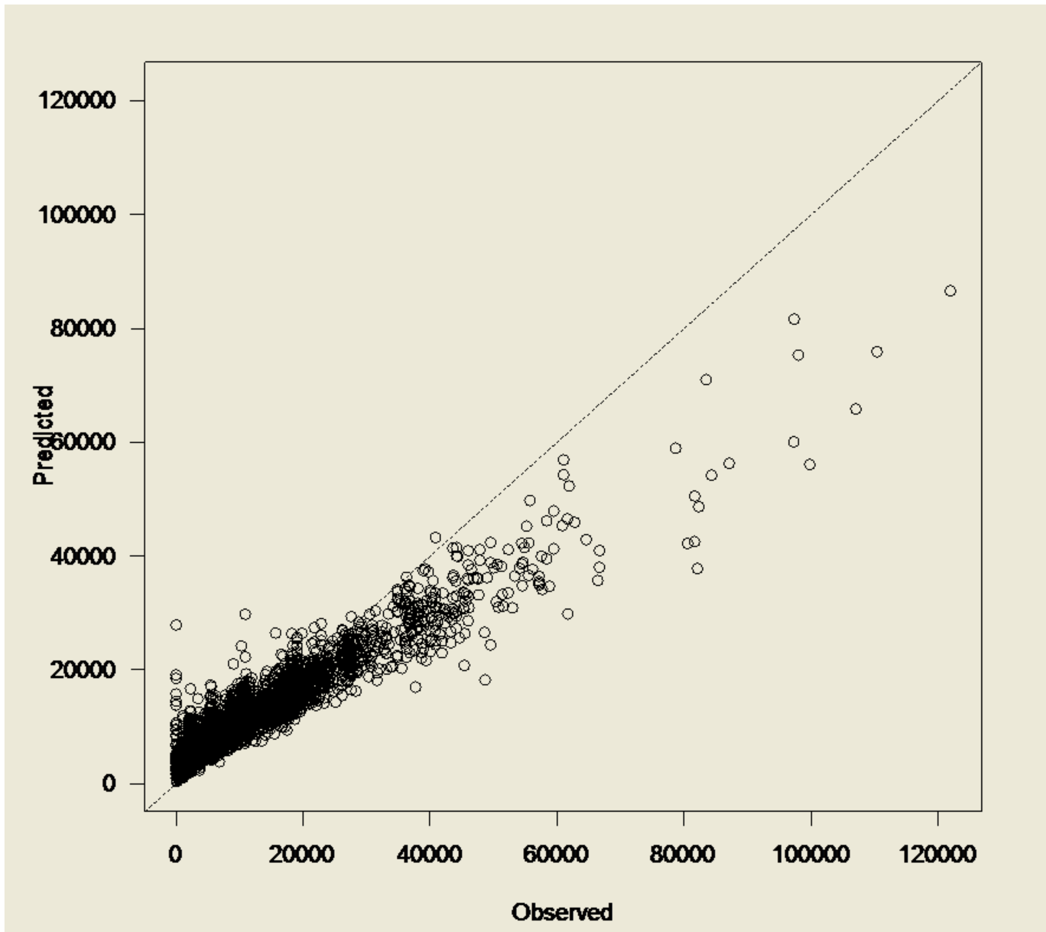
1m-21m Observed and predicted medical payment (medical only claims) comparison



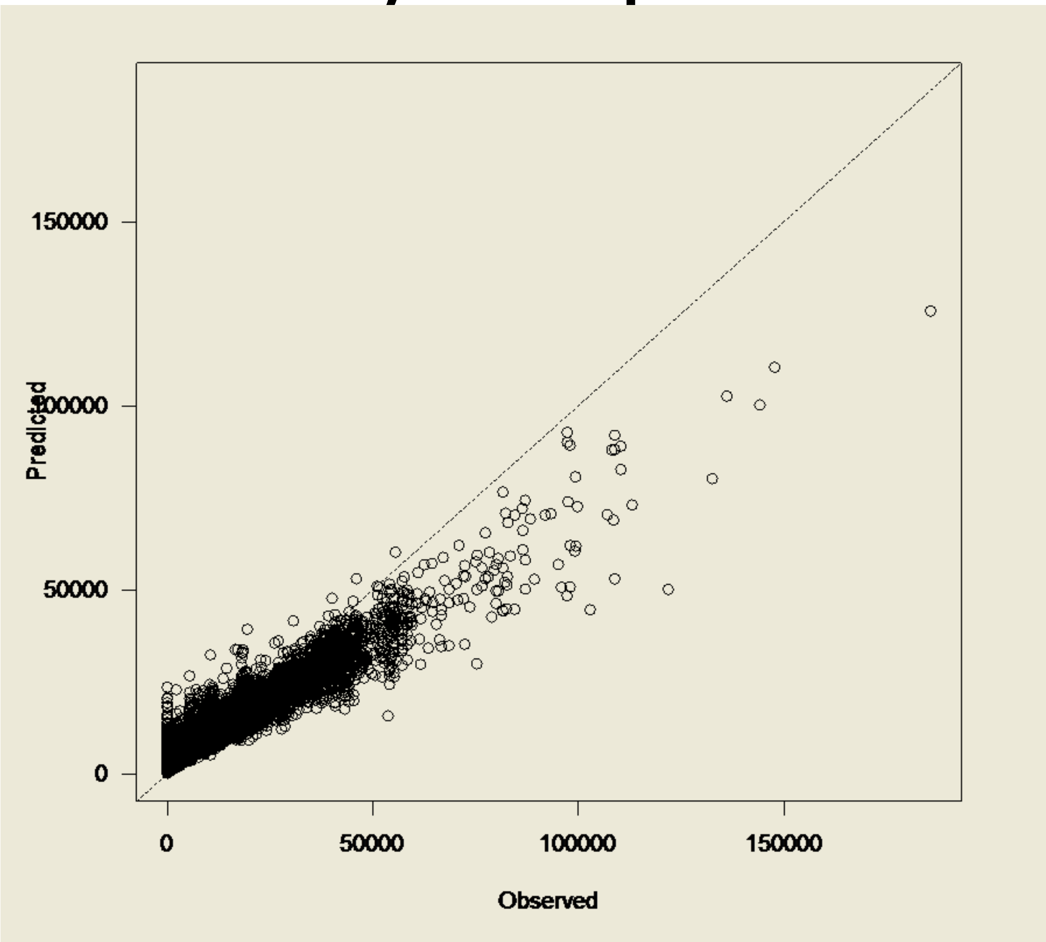
1m-21m Observed and predicted timeloss payment (timeloss claims only) comparison



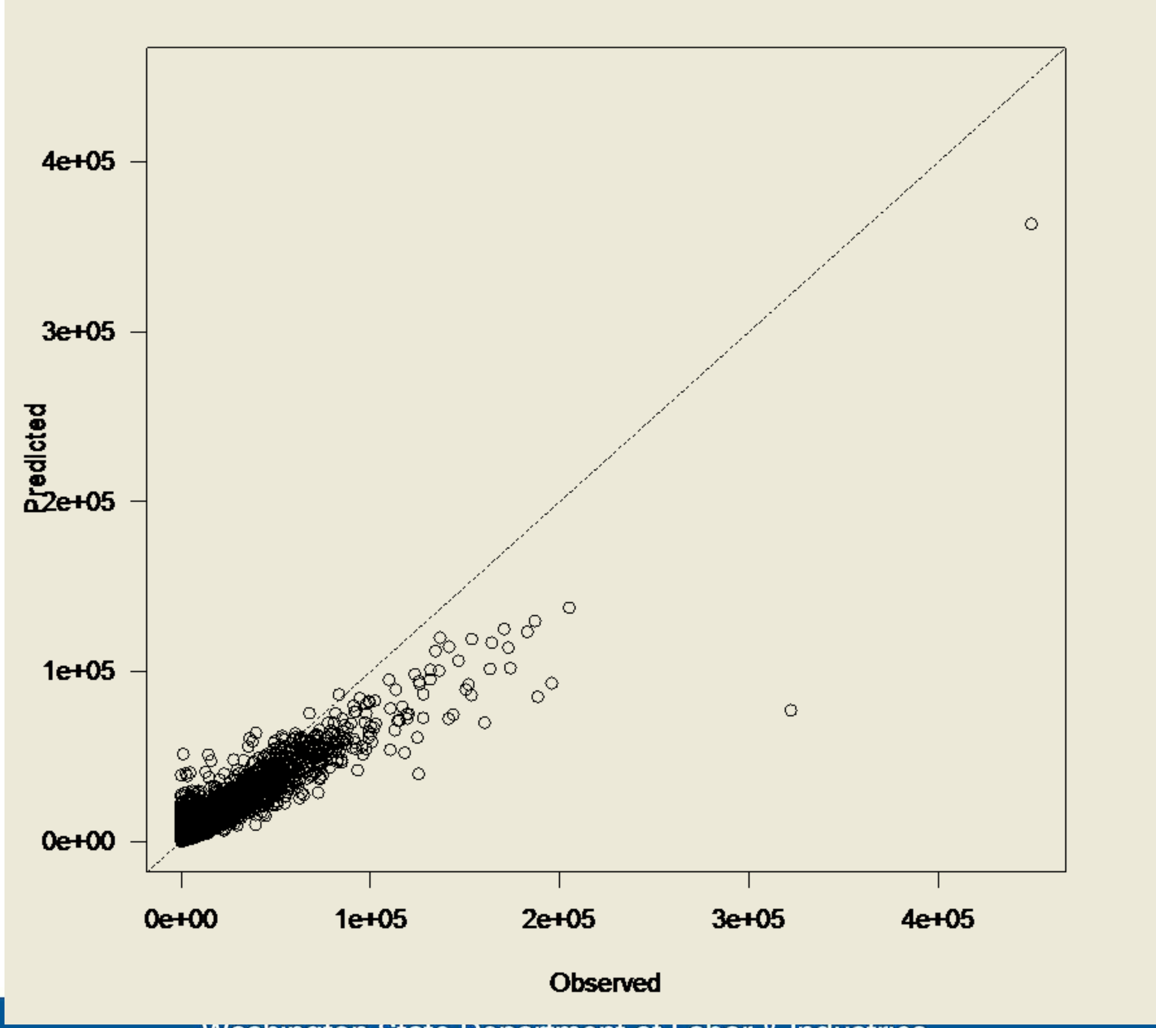
1m-21m Observed and predicted PPD payment comparison



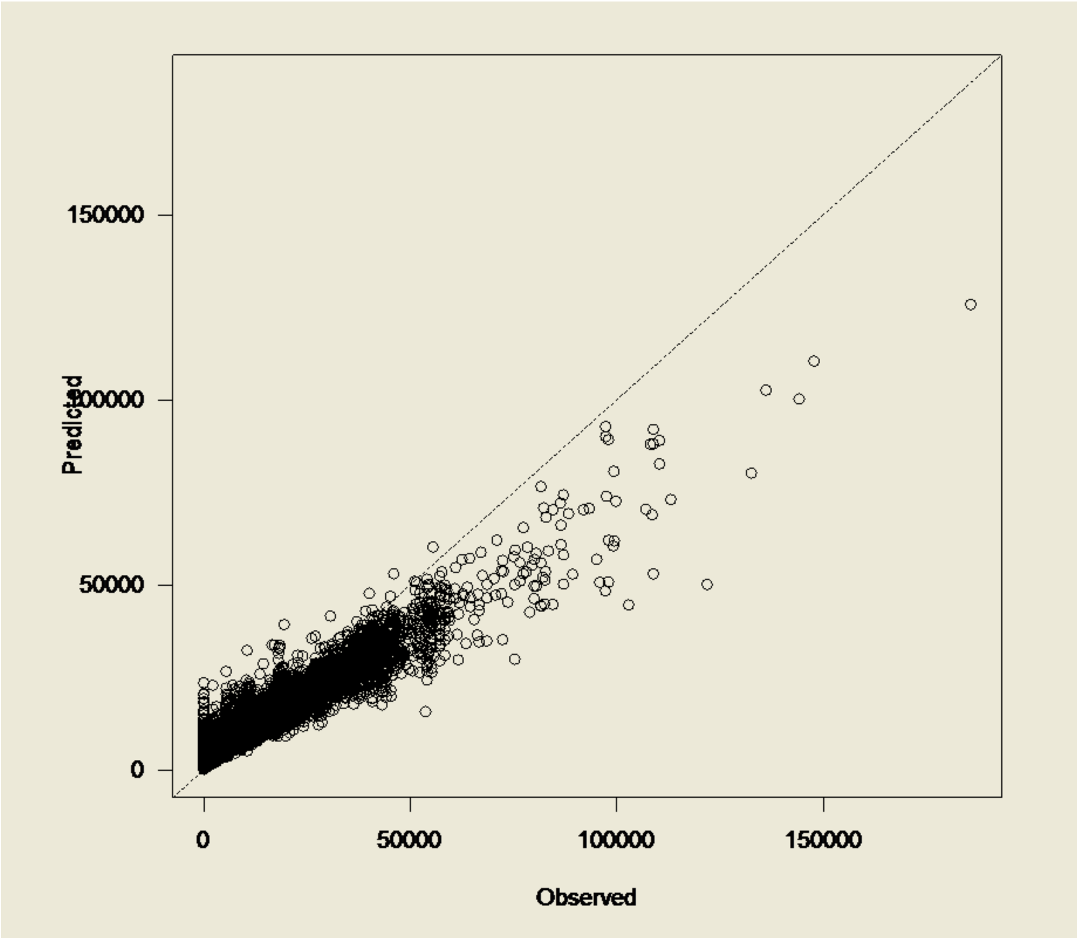
9m-21m Observed and predicted medical payment (medical only claims) comparison



9m-21m Observed and predicted timeloss payment (timeloss only claims) comparison



9m-21m Observed and predicted PPD payment comparison



Questions