Name: Chan Ka Tsun
SID: 53563363

Q1 It is expected that increasing compensation will lead to an increase in injury duration. This is because the subject will be incentivized to stay at home for recovery for a longer period of time if the compensation is sufficient to cover his living, or high enough for him to not to worry about staying out of work. It is not advised to perform a simple regression, which can show correlation between variables, but correlation does not mean causation, therefore, cannot prove that the conjecture of increased compensation leads to an increase in duration, instead, difference in difference or DDD should be applied.

Q2. (See Figure 1) As there are 2 different states, different compensation policies may be enforced, that the results are listed according to states and whether they are from high earnings group. It is postulated that the increase in benefit will affect the high earning group in a larger magnitude compared to non-high earnings group. One reason of that can be that high earnings group will be able to enjoy the lifted maximum compensation while the low earnings group is unaffected by the maximum ceiling. This leads to a decrease in opportunity cost for the high earners to stay out of work, meaning a higher incentive to have a longer duration of benefit. While the low earners are not affected by the compensation ceiling, their opportunity cost of remains the same either way. DID is performed that there is a 0.188 overall increase in duration in high earn group after the compensation lift, comparing to the non high earn group. T-test

result shows that there is significant evidence for the increase in duration after compensation lift in both states, while p-value in low earn group is too high that null hypothesis of no difference in duration is not rejected.

Q3. (See Figure 2) The result shows that there is significant increase in duration for high earning group in Kentucky and overall. The interaction effect of afchnge and highearn shows significance with p-value close to 0, that high-earners tends to have a longer duration after the benefit lift. The difference in results in Kentucky and Michigan may be due to a different benefit policy in terms of amount that one can claim or other unknown rules that must be met before the compensation. The unbalanced sample size (smaller in Michigan) may also be a reason that result is insignificant in Michigan comparing to that in Kentucky.

Q4. (See Figure 3) When there are omitted variables, there might be endogeneity caused by unobserved heterogeneity. The observed association between outcome variable duration and explanatory variable Variables such as **marital status**: married person may want a longer duration to recover and spend time with families, or may have shorter duration as the subject may be the breadwinner of the family, **gender**: male may recover faster**, type of injury**: injury in important parts such as head may require a longer recovery time, **type of industry**: construction type may be more dangerous and lead to more severe injury compared to

manufacturing, **previous earnings:** a higher previous earning may lead to longer recovery time as the subject may be financially supported by previous earnings to enable a longer duration being out of work. **Comparing the results:** A more significant (smaller p-values) result is observed after more variables are included. It is also coherent to the conjecture that male correlates to a shorter duration and head injury correlates to a longer duration.

Q5. **Endogeneity:** There might be endogenous variables in this natural experiment. For examples, a higher medical costs can mean a longer hospital stay; an older individual might need a longer stay for recovery, or a more detailed health check that incurs higher medical cost; lower metabolism rate in women in general might lead to a longer recovery time or duration. **Heterogeneity:** There might be heterogeneity that we can observe but we do not have the data about such as education level: education might affect the values towards health thus affecting the duration the subject is willing to take to recover from the injury. There might be heterogeneity that we cannot observe inherently such as previous life experiences affecting the subject's decision on taking a longer duration leave. **Time trend assumption:** There might not be a common time trend for high-earners and low-earners: For examples, high-earners are older in age that the changes before and after the benefit lift follows a different time trend compared to low-earners, instead of purely due to their differ in earnings.

Name: Chan Ka Tsun
SID: 53563363

Appendix I

	Treatment Group (High earn)		Control Group (not high earn)		DID	t-test (T) p-value	t-test (C) p-value
	Before	After	Before	After			
Kentucky							
Mean ln(duration)	1.38	1.58	1.126	1.133	0.191	0.00019***	0.862
Sample size	1233	1161	1705	1527			
Michigan							
Mean ln(duration)	1.58	1.87	1.41	1.51	0.192	0.03***	0.24
Sample size	239	219	589	477			
Overall							
Mean ln(duration)	1.41	1.62	1.20	1.22	0.188	0.00002***	0.547
Sample size	1472	1380	2294	2004			

Figure 1 Q2 Results - Difference in high-earners vs low-earners

Name: Chan Ka Tsun
SID: 53563363

	Dependent variable:		
	Kentucky (1)	Michigan (2)	Overall (3)
afchnge	0.008	0.097	0.024
	p = 0.865	p = 0.251	p = 0.552
highearn	0.256***	0.169	0.215***
	p = 0.00000	p = 0.110	p = 0.00000
afchnge:highearn	0.191***	0.192	0.188***
	p = 0.006	p = 0.214	p = 0.003
Constant	1.126***	1.413***	1.199***
	p = 0.000	p = 0.000	p = 0.000
Observations	5,626	1,524	7,150
R2	0.021	0.012	0.016
Adjusted R2	0.020	0.010	0.015
Residual Std. Error	1.269 (df = 5622)	1.376 (df = 1520)	1.298 (df = 7146)
F Statistic	39.540*** (df = 3; 5622)	6.049*** (df = 3; 1520)	38.342*** (df = 3; 7146)
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 2 Q3 Interaction Effect

Name: Chan Ka Tsun
SID: 53563363

	Dependent variable:		
	Kentucky	Michigan	Overall
	(1)	(2)	(3)
afchnge	-0.006	0.003	-0.011
	p = 0.866	p = 0.969	p = 0.741
highearn	-1.657**	2.830	-0.904
	p = 0.042	p = 0.301	p = 0.111
highlpre	0.269**	-0.450	0.139
	p = 0.043	p = 0.276	p = 0.128
male	-0.054	-0.328***	-0.088**
	p = 0.164	p = 0.0002	p = 0.013
married	0.062*	-0.064	0.031
	p = 0.073	p = 0.361	p = 0.320
lage	0.249***	0.482***	0.309***
	p = 0.00000	p = 0.00001	p = 0.000
ltotmed	0.361***	0.316***	0.351***
	p = 0.000	p = 0.000	p = 0.000
hosp	0.252***	0.244***	0.249***
	p = 0.000	p = 0.004	p = 0.000
manuf	-0.138***	-0.125*	-0.131***
	p = 0.0001	p = 0.065	p = 0.00002
construc	0.055	0.348***	0.128***
	p = 0.214	p = 0.0001	p = 0.002
head	-0.426***	-0.831***	-0.555***
	p = 0.0001	p = 0.0005	p = 0.00000
neck	0.370***	-0.283	0.219*
	p = 0.007	p = 0.338	p = 0.079
upextr	0.132	-0.221	0.024
	p = 0.125	p = 0.175	p = 0.756
trunk	0.147	-0.135	0.063
	p = 0.110	p = 0.443	p = 0.438
lowback	0.192**	-0.339**	0.050
	p = 0.026	p = 0.041	p = 0.510
lowextr	0.189**	-0.307*	0.053
	p = 0.030	p = 0.065	p = 0.489
occdis	0.534***	0.358	0.533***
	p = 0.003	p = 0.206	p = 0.0004
afchnge:highearn	0.164***	0.203	0.181***
	p = 0.006	p = 0.126	p = 0.001
Constant	-1.922***	-1.609***	-1.868***
	p = 0.000	p = 0.0001	p = 0.000
Observations	5,347	1,475	6,822
R2	0.318	0.304	0.311
Adjusted R2	0.316	0.295	0.309
Residual Std. Error	1.055 (df = 5328)	1.155 (df = 1456)	1.082 (df = 6803)
F Statistic	138.124*** (df = 18; 5328)	35.332*** (df = 18; 1456)	170.736*** (df = 18; 6803)
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 2 Q4 Including other variables

Name: Chan Ka Tsun
SID: 53563363

Appendix II

```r
library(foreign)
library(stargazer)

df <- read.dta("~/Downloads/INJURY.dta")
ky <- subset(df, ky == 1)
mi <- subset(df, mi == 1)

#Q2: mean,sample size and t test
#ky, mi, and whole(df)
#sample size
t_b = sapply(subset(df, afchnge == 0 & highearn == 1, select = ldurat),length)
t_a = sapply(subset(df, afchnge == 1 & highearn == 1, select = ldurat),length)
c_b = sapply(subset(df, afchnge == 0 & highearn == 0, select = ldurat),length)
c_a = sapply(subset(df, afchnge == 1 & highearn == 0, select = ldurat),length)
#mean
t_b = sapply(subset(df, afchnge == 0 & highearn == 1, select = ldurat),mean)
t_a = sapply(subset(df, afchnge == 1 & highearn == 1, select = ldurat),mean)
c_b = sapply(subset(df, afchnge == 0 & highearn == 0, select = ldurat),mean)
c_a = sapply(subset(df, afchnge == 1 & highearn == 0, select = ldurat),mean)
#t test (difference of before and after in control and treatment group)
t.test(subset(df, afchnge == 0 & highearn == 1, select = ldurat),subset(df, afchnge == 1 & highearn == 1, select = ldurat))
t.test(subset(df, afchnge == 0 & highearn == 0, select = ldurat),subset(df, afchnge == 1 & highearn == 0, select = ldurat))

#DID
whole.reg <- lm(ldurat~afchnge+highearn+afchnge*highearn,data=df)
ky.reg <- lm(ldurat~afchnge+highearn+afchnge*highearn,data=ky)
mi.reg <- lm(ldurat~afchnge+highearn+afchnge*highearn,data=mi)
summary(ky.reg)
summary(mi.reg)
summary(whole.reg)
#Formatting the table
stargazer(ky.reg, mi.reg, whole.reg, dep.var.labels="",
          column.labels=c("Kentucky","Michigan","Overall"),
          type="text", no.space=TRUE, report=('vc*p'))

#DID with more control variables
kyadv.reg <- lm(ldurat~afchnge+highearn+afchnge*highearn+highlpre+male+married+lage+ltotmed+hosp
                +manuf+construc+head+neck+upextr+trunk+lowback+lowextr+occdis,data=ky)
miadv.reg <- lm(ldurat~afchnge+highearn+afchnge*highearn+highlpre+male+married+lage+ltotmed+hosp
                +manuf+construc+head+neck+upextr+trunk+lowback+lowextr+occdis,data=mi)
wholeadv.reg <- lm(ldurat~afchnge+highearn+afchnge*highearn+highlpre+male+married+lage+ltotmed
                   +hosp+manuf+construc+head+neck+upextr+trunk+lowback+lowextr+occdis,data=df)
summary(kyadv.reg)
summary(miadv.reg)
summary(wholeadv.reg)
```