Appendix

Additional Tables and Figures

Pearson and Concordance Statistics Tables

	Test Statistic	Degrees of Freedom	P-Value
2011	89356	1	0
2012	83571	1	0
2013	78207	1	0
2014	63406	1	0
2015	54827	1	0
All Years	365920	1	0

Table A1: The Pearson chi-squared test-statistics for the general association between insurance status and having a personal doctor. The large test-statistics and p-values of virtually zero indicate that the association between insurance status and whether or not an individual has a personal doctor is highly statistically significant.

		95% Lower Bound	Estimate	95% Upper Bound	Test Statistic	P-value
2011	Tau-b	0.42219	0.42404	0.42589	298.92892	0
	Gamma	0.79771	0.80343	0.80914	275.70095	0
2012	Tau-b	0.42115	0.42306	0.42497	289.0915	0
	Gamma	0.79589	0.8018	0.8077	266.1061	0
2013	Tau-b	0.40084	0.40272	0.4046	279.65897	0
	Gamma	0.78043	0.78652	0.79261	253.04405	0
2014	Tau-b	0.37136	0.3733	0.37523	251.8112	0
	Gamma	0.7756	0.7825	0.7894	222.30481	0
2015	Tau-b	0.35421	0.3562	0.35819	234.15651	0
	Gamma	0.77406	0.78161	0.78917	202.73369	0
All	Tau-b	0.39516	0.39601	0.39687	604.91504	0
Years	Gamma	0.78665	0.7895	0.79236	541.99784	0

Table A2: The rank correlation statistics for the relationship between insurance status and having a personal doctor. The positive tau-b and gamma statistics indicate that having insurance is correlated with having a personal doctor. While the large test-statistics and p-values of virtually zero indicate that this correlation is highly statistically significant, the magnitude of the tau-b statistic indicates that the strength of the correlation is moderate.

	Test Statistic	Degrees of Freedom	P-Value
2011	49161	4	0
2012	46768	4	0
2013	46920	4	0
2014	36317	4	0
2015	29349	4	0
All Years	208477	4	0

Table A3: The Pearson chi-squared test-statistics for the general association between insurance status and time since last general health checkup (doctor utilization). The large test-statistics and p-values of virtually zero indicate that the association between insurance status and doctor utilization is highly statistically significant for every year.

	Test Statistic	Degrees of Freedom	P-Value
2011	12568	4	0
2012	12135	4	0
2013	11731	4	0
2014	8918	4	0
2015	6115	4	0
All Years	51461	4	0

Table A4: The Pearson chi-squared test-statistics for the general association between insurance status and doctor utilization when only considering individuals who reported having a personal doctor. Although the test-statistics are smaller than those observed when looking at all individuals, they are still very large and the p-values are still virtually zero, indicating that the association between insurance status and doctor utilization is still highly statistically significant.

		95% Lower Bound	Estimate	95% Upper Bound	Test Statistic	P-value
2011	Tau-b	0.28825	0.2901	0.29195	215.78958	0
	Gamma	0.56567	0.57291	0.58015	155.10423	0
2012	Tau-b	0.29033	0.29225	0.29416	210.38504	0
	Gamma	0.56962	0.57709	0.58455	151.45463	0
2013	Tau-b	0.28574	0.28762	0.2895	210.07944	0
	Gamma	0.57214	0.57967	0.58721	150.75636	0
2014	Tau-b	0.25465	0.25659	0.25852	181.67087	0

		95% Lower Bound	Estimate	95% Upper Bound	Test Statistic	P-value
	Gamma	0.55586	0.56452	0.57317	127.84276	0
2015	Tau-b	0.23259	0.23458	0.23657	161.80014	0
	Gamma	0.53851	0.54806	0.55761	112.43889	0
All	Tau-b	0.27349	0.27434	0.2752	440.79052	0
Years	Gamma	0.56693	0.5705	0.57407	313.61055	0

Table A5: The rank correlation statistics for the relationship between insurance status and time since last general health checkup (doctor utilizationt). The positive tau-b and gamma statistics indicate that having insurance is correlated with having a checkup more recently. While the large test-statistics and p-values of virtually zero indicate that this correlation is highly statistically significant, the magnitude of the tau-b statistic indicates that the strength of the correlation is not very strong.

		95% Lower Bound	Estimate	95% Upper Bound	Test Statistic	P-value
2011	Tau-b	0.15673	0.15874	0.16074	107.4396	0
	Gamma	0.44249	0.45501	0.46752	71.26172	0
2012	Tau-b	0.1556	0.15767	0.15974	103.11534	0
	Gamma	0.44197	0.45501	0.46805	68.40036	0
2013	Tau-b	0.15475	0.1568	0.15886	103.04601	0
	Gamma	0.4555	0.46899	0.48248	68.15365	0
2014	Tau-b	0.1318	0.13391	0.13601	85.87215	0
	Gamma	0.4378	0.4536	0.4694	56.27613	0
2015	Tau-b	0.11116	0.1133	0.11545	71.16148	0
	Gamma	0.39925	0.41688	0.43451	46.33606	0
All	Tau-b	0.14521	0.14613	0.14706	212.92215	0
Years	Gamma	0.44824	0.45459	0.46093	140.33422	0

Table A6: The rank correlation statistics for the relationship between insurance status and doctor utilization when considering only individuals who reported having a personal doctor. Again, the positive tau-b and gamma statistics (and the p-values of approximately zero) indicate that having insurance is correlated with having a checkup more recently. However, the magnitude of the tau-b statistics are less than those observed when considering all individuals. Thus the strength of the correlation appears weaker when cosidering only individuals who reported having a personal doctor.

Proportional Differences

	Lower Bound	Estimate	Upper Bound
2011	0.45151	0.45489	0.45827
2012	0.44907	0.45255	0.45603
2013	0.44806	0.45156	0.45507
2014	0.44735	0.45133	0.4553
2015	0.44191	0.44627	0.45062
All	0.44837	0.45003	0.45169

Table A7: 95% Confidence intervals for the absolute differences between the proportion of insured individuals with a doctor and uninsured individuals with a doctor. Note that although the absolute differences appear to decrease from 2011 to 2015, the confidence intervals for consecutive years overlap.

	Lower Bound	Estimate	Upper Bound
2011	2.095321	2.112033	2.128879
2012	2.082682	2.099672	2.116801
2013	2.133082	2.151443	2.169962
2014	2.146895	2.168344	2.190007
2015	2.102963	2.125541	2.148361
All Years	2.116665	2.125125	2.13362

Table A8: 95% Confidence intervals for the relative differences between the proportion of insured individuals with a doctor and uninsured individuals with a doctor.

	Lower Bound	Estimate	Upper Bound
2011	0.34289	0.34635	0.34981
2012	0.34252	0.34608	0.34965
2013	0.34171	0.3453	0.3489
2014	0.32521	0.3293	0.33339
2015	0.31361	0.31807	0.32254
All Years	0.33719	0.33889	0.34058

Table A9: 95% Confidence intervals for the absolute differences between the proportion of insured individuals who have had a checkup within a year and the proportion of uninsured individuals that have. Note that although the differences are very close from year to year they do generally decrease from 2011 to 2015.

	Lower Bound	Estimate	Upper Bound
2011	0.22187	0.22689	0.23191
2012	0.21393	0.21905	0.22418
2013	0.21907	0.22435	0.22963
2014	0.20193	0.20797	0.214
2015	0.18163	0.18814	0.19465
All Years	0.21387	0.21635	0.21882

Table A10: 95% Confidence intervals for the absolute differences between the proportion of insured and uninsured individuals who have had a checkup within a year, only considering individuals who reported having personal doctor. Note that although the differences are very close from year to year they do generally decrease from 2011 to 2015.

	Lower Bound	Estimate	Upper Bound
2011	1.8806	1.8965	1.9126
2012	1.8585	1.8744	1.8904
2013	1.8417	1.8574	1.8732
2014	1.7719	1.7886	1.8055
2015	1.7357	1.7535	1.7715
All Years	1.8333	1.8406	1.848

Table A11: 95% Confidence intervals for the relative differences between the proportion of insured individuals who have had a checkup within a year and the proportion of uninsured individuals that have. Note that although the relative differences are close from year to year they do generally decrease from 2011 to 2015.

	Lower Bound	Estimate	Upper Bound
2011	1.3957	1.4082	1.4207
2012	1.3706	1.3828	1.395
2013	1.3754	1.3878	1.4004
2014	1.335	1.3484	1.362
2015	1.2939	1.3076	1.3215
All Years	1.3672	1.3729	1.3787

Table A12: 95% Confidence intervals for the relative differences between the proportion of individuals who have had a checkup within a year, only considering individuals who reported having a personal doctor. Note that although the relative differences are close from year to year they do generally decrease from 2011 to 2015.

	Lower Bound	Estimate	Upper Bound
2011	0.18391	0.18685	0.1898
2012	0.17839	0.18138	0.18437
2013	0.17978	0.18279	0.1858
2014	0.17494	0.17836	0.18178
2015	0.17474	0.17849	0.18225
All Years	0.18068	0.1821	0.18353

Table A13: 95% Confidence intervals for the absolute differences between the proportion of uninsured individuals who have not had a checkup within five years and the proportion of insured individuals that have not.

	Lower Bound	Estimate	Upper Bound
2011	0.083867	0.087207	0.090547
2012	0.084034	0.087436	0.090838
2013	0.082483	0.085923	0.089363
2014	0.083365	0.087346	0.091327
2015	0.076596	0.080836	0.085077
All Years	0.084842	0.086473	0.088103

Table A14: 95% Confidence intervals for the absolute differences between the proportion of individuals who have not had a checkup within five years, only considering individuals who reported having a personal doctor.

	Lower Bound	Estimate	Upper Bound
2011	3.7078	3.7675	3.8282
2012	3.7553	3.8192	3.8841
2013	3.9498	4.018	4.0874
2014	3.8264	3.8979	3.9707
2015	3.7548	3.8289	3.9044
All Years	3.8444	3.8745	3.9048

Table A15: 95% Confidence intervals for the relative differences between the proportion of uninsured individuals who have not had a checkup within five years and the proportion of insured individuals that have not.

	Lower Bound	Estimate	Upper Bound
2011	2.9539	3.0419	3.1324
2012	3.07	3.1654	3.2638
2013	3.3084	3.4179	3.5311
2014	3.367	3.4919	3.6215
2015	3.1359	3.2637	3.3966
All Years	3.239	3.2875	3.3367

Table A16: 95% Confidence intervals for the relative differences between the proportion of individuals who have not had a checkup within five years, only considering individuals who reported having a personal doctor.

Data-Set Cleaning

Data Set Year	Total Before Cleaning	Total After Cleaning	Number Removed	Percentage Removed
2011	506467	503325	3142	0.6204
2012	475687	472780	2907	0.6111
2013	491773	488240	3533	0.7184
2014	464664	461057	3607	0.7763
2015	441456	437959	3497	0.7922

Table A17: Values removed from the data sets during cleaning.

Technical Section: Example Equations and Code

In order to demonstrate how the nation level analysis was done, example calculations for the personal doctor measure are shown below.

Pearson Chi-Squared Test

The Pearson chi-squared test determines whether or not the distribution of the response variable (in this case whether or not an individual has a personal doctor) is the same at both levels of the explanatory variable (insurance status). The null hypothesis is that there is no association between the explanatory and response variable and thus the distribution of the

response variable is the same at both levels of the explanatory variable. The alternative hypothesis is that the distribution of the response variable is not the same at both levels of the explanatory variable and thus there is an association between the explanatory and response variables.

This test is performed with a frequency table. The test-statistic is the the sum of the squared difference between the observed and expected values for each cell divided by the expected values of each cell. Under the null hypothesis, this test-statistic has a chi-square distribution with degrees of freedom equal to the number of rows minus one times the number of columns minus one:

$$X^2 = \sum_{i=1}^{n} \frac{(observed - expected)^2}{expected} \sim \chi^2_{(r-1)(c-1)}$$

We can easily calculate this statistic in R using the "chisq.test()" function, which takes as an argument the contingency table we are interested in. The function returns the Pearson test-statistic, the degrees of freedom, and the p-value. It can also return the observed and expected tables if requested.

```
# ST_PD_t11[,2:3] is the frequency table for insurance status and whether or not an individual has a per
sonal doctor.

chisq.test(ST_PD_t11[,2:3])
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  ST_PD_t11[, 2:3]
## X-squared = 89356, df = 1, p-value < 2.2e-16
```

```
pcs <- chisq.test(ST_PD_t11[,2:3])

pcs$statistic
```

```
## X-squared
##  89356.01
```

```
pcs$parameter
```

```
## df
##  1
```

```
pcs$p.value
```

```
## [1] 0
```

```
pcs$observed
```

```
##               1        3
## [1,] 351558.16 55356.99
## [2,]  36834.62 53211.23
```

```
pcs$expected
```

```
##                1        3
## [1,] 318018.73 88896.42
## [2,]  70374.05 19671.80
```

We used this function to perform the Pearson chi-squared tests for all of our measures of interest.

Rank Correlation Statistics

The rank correlation statistics, or concordance statistics, indicate the direction of association between two ordinal variables. The Kendall's tau-b and Goodman and Kruskal's gamma statistics range from -1 to 1, with negative values indicating negative correlation between the two variables and positive values indicating positive correlation between the two variables. A negative 1 indicates perfect negative correlation, a positive 1 indicates perfect positive correlation, and a 0 indicates no correlation. These statistics are calculated using concordant, discordant and (for the tau-b statistic) tied pairs of observations. A concordant pair of observations is a pair for which one observation ranks higher in both the variables than the other observation. A discordant pair of observations is when one observation ranks higher in one variable but lower in the other variable compared to the other observation. A tied pair of observations is when the pair of observations tie for rank in one or both variables. For instance, if we were looking at height and weight, a pair of individuals with one individual being 6 feet tall and 180 pounds and the other individual being 5.8 feet tall and 160 pounds would be a concordant pair. On the other hand, a pair of individuals with one individual being 6 feet tall and 150 pounds and the other being 5.8 feet tall and 160 pounds would be a discordant pair. Finally, a pair of individuals with one individual being 6 feet tall and 100 pounds and one being 6 feet tall and 200 pounds would be a tied pair (as long as they are tied in at least one variable they are considered a tied pair). For a data set of *n* observations, there are $ n*(n-1)/2 $ total pairs. For hypothesis testing with both Kendall's tau-b and Goodman and Kruskal's gamma, we are testing the null hypothesis that there is no correlation between the two variables against the alternative hypothesis that there is a correlation.

Goodman and Kruskall's gamma is the simpler of the two rank correlation statistics that we examined, but potentially less informative. Gamma only takes into account the number of discordant and concordant pairs. Thus, a data set with 100 concordant, 10 discordant, and 5 tied pairs would have the same gamma statistic as a data set with 100 concordant, 10 discordant, and 10000 tied pairs. Gamma is calculated by taking the difference between the number of concordant and discordant pairs and dividing it by the number of concordant and discordant pairs combined.

$$\widehat{\gamma} = \frac{C - D}{C + D}$$

- *C* is the number of concordant pairs
- *D* is the number of discordant pairs

An approximate standard error for gamma for large samples is given by:

$$\widehat{SE}[\gamma] = \sqrt{\frac{n(1 - \widehat{\gamma}^2)}{C + D}}$$

We used this standard error to calculate approximate confidence intervals for gamma.

For large samples, gamma divided by its standard error has an approximatel standard normal distribution. This is what we used as our test-statistic for gamma, in order to test if there is significant correlation between the two variables of interest (in this case, insurance status and whether or not an individual has a personal doctor).

$$\frac{\widehat{\gamma}}{\widehat{SE}} \sim AN(0,1)$$

In order to calculate the concodance statistics (both gamma and tau), their confidence intervals, their test-statistics, and the associated p-values we created an R function called "Rank.Correlation()". This function will be described after Kendall's tau-b is described next.

Kendall's tau-b is a little more complicated to compute and interpret than gamma, but it takes ties into account. Our data contained many ties so the tau-b statistic is perhaps a more useful measure for our situation than gamma. Kendall's tau-b looks at the number of discordant pairs and concordant pairs relative to the total number of pairs (including ties) in the data set. The formula is thus more complex.

$$\widehat{\tau_b} = \frac{C - D}{\sqrt{(n_0 - n_1)(n_0 - n_2)}}$$

- n_0 is the total number of pairs in the data set: $n(n-1)/2$
- n_1 is the sum of total number of *tied* pairs at each level of the first variable: $n_1 = \sum T_1$
- n_2 is the sum of total number of *tied* pairs at each level of the second variable: $n_2 = \sum T_2$

An approximate standard error for the tau-b statistic with large samples is:

$$\widehat{SE}[\tau_b] = \frac{2(2n+5)}{9n(n-1)}$$

We used this standard error to caculate approximate confidence intervals for tau-b, but we did not use it to calculate our test-statistic. Instead we used a more conservative measure.

The test-statistic we used for hypothesis testing with tau-b is given below. For large samples this test-statistic has an approximately standard normal distribution.

$$\widehat{z_{\tau_b}} = \frac{C - D}{\sqrt{\left(\frac{v_0 - v_t - v_u}{18}\right) + v_1 + v_2}} \sim AN(0,1)$$

- $v_0 = n(n-1)(2n+5)$
- $v_t = \sum_i t_i(t_i - 1)(2t_i + 5)$
    - t_i is the total number of tied *values* (not pairs) for the ith level of the first variable. *Tied values* for each level of the first variable are the same as row totals (if the row variable is considered the first variable) or column totals (if the the column variable is considered the first variable).
- $v_u = \sum_j u_j(u_j - 1)(2u_j + 5)$
    - u_i is the total number of tied *values* (not pairs) for the jth level of the second variable. *Tied values* for each level of the second variable are the same as row totals (if the row variable is considered the second variable) or column totals (if the the column variable is considered the second variable).
- $v_1 = \frac{\sum_i t_i(t_i-1)\sum_j u_j(u_j-1)}{2n(n-1)}$
- $v_2 = \frac{\sum_i t_i(t_i-1)(t_i-2)\sum_j u_j(u_j-1)(u_j-2)}{9n(n-1)(n-2)}$

To calculate both the gamma and tau-b statistics, their confidence intervals, their test-statistics, and the corresponding p-values we created the R function "Rank.Correlation()" shown below. This function also calculates the tau-a statistic (a Kendall's tau statistic that does not account for ties) and another test-statistic for the tau-b statistic. However, we found the above statistic to be more conservative and thus chose it for our report. Both test-statistics give p-values of approximately zero for all tests performed in this report.

```r
Rank.Correlation <- function(x, alpha = 0.05){

  z <- qnorm(alpha/2, lower.tail = F)
  cols <- dim(x)[2]
  n <- sum(x[1,]) + sum(x[2,])
  np <- (n*(n-1))/2
  SE.LS.approx <- sqrt((2*(2*n+5))/(9*n*(n-1)))

  T.pairs <- sum(x[1,])*(sum(x[1,])-1)/2 + sum(x[2,])*(sum(x[2,])-1)/2
  Tied.r <- T.pairs

  C.pairs <- 0
  D.pairs <- 0
  Tied.c <- 0
  for(i in 1:cols){
    if(i < cols){
      for(j in (i + 1):cols){
        C.pairs <- C.pairs + x[1,i]*x[2,j]
        D.pairs <- D.pairs + x[2,i]*x[1,j]
      }
    }

    T.pairs <- T.pairs + x[1,i]*x[2,i]
    Tied.c <- Tied.c + sum(x[,i])*(sum(x[,i])-1)/2

  }

  vt <- 0
  v1.c <- 0
  v2.c <- 0
  for(i in 1:cols){
    s <- sum(x[,i])*(sum(x[,i])-1)
    vt <- vt + s*(2*sum(x[,i])+5)
    v1.c <- v1.c + s
    v2.c <- v2.c + s*(sum(x[,i])-2)
  }

  vu <- 0
  v1.r <- 0
  v2.r <- 0
  for(i in 1:2){
    s <- sum(x[i,])*(sum(x[i,])-1)
    vu <- vu + s*(2*sum(x[i,])+5)
    v1.r <- v1.r + s
    v2.r <- v2.r + s*(sum(x[i,])-2)
  }

  v1 <- v1.c*v1.r/(2*n*(n-1))
  v2 <- v2.c*v2.r/(9*n*(n-1)*(n-2))

  v0 <- n*(n-1)*(2*n+5)

  v <- (v0 - vt - vu) / (18) + v1 + v2


  Ktau.a <- (C.pairs - D.pairs) / np
  Ktau.b <- (C.pairs - D.pairs) / sqrt( (np - Tied.r)*(np - Tied.c) )
```

```
    Z.a <- as.numeric( 3*(C.pairs - D.pairs) / sqrt(n*(n-1)*(2*n+5)/2) )
    pva <- pnorm(abs(Z.a), lower.tail = F)
    Z.a.Alt <- as.numeric( Ktau.a/SE.LS.approx )
    pva.Alt <- pnorm(abs(Z.a.Alt), lower.tail = F)

    Ta.LB <- Ktau.a - z*SE.LS.approx
    Ta.UB <- Ktau.a + z*SE.LS.approx

    Z.b <- as.numeric( (C.pairs - D.pairs) / sqrt(v) )
    pvb <- pnorm(abs(Z.b), lower.tail = F)
    Z.b.Alt <- as.numeric( Ktau.b/SE.LS.approx )
    pvb.Alt <- pnorm(abs(Z.b.Alt), lower.tail = F)

    Tb.LB <- Ktau.b - z*SE.LS.approx
    Tb.UB <- Ktau.b + z*SE.LS.approx

    gamma <- (C.pairs - D.pairs)/(C.pairs + D.pairs)
    SE.gamma <- sqrt(n*(1-gamma^2)/(C.pairs+D.pairs))
    Z.gamma <- as.numeric( gamma/SE.gamma )
    pvg <- pnorm(abs(Z.gamma), lower.tail = F)

    G.LB <- gamma - z*SE.gamma
    G.UB <- gamma + z*SE.gamma

    out <- matrix(c(Ta.LB, Tb.LB, G.LB, Ktau.a, Ktau.b, gamma, Ta.UB, Tb.UB, G.UB, Z.a, Z.b, Z.gamma,
                    pva, pvb, pvg, Z.a.Alt, Z.b.Alt, NA, pva.Alt, pvb.Alt, NA),
                  nrow = 3, ncol = 7,
                  dimnames = list(c("Tau.a", "Tau.b", "Gamma"),
                                  c("Lower.Bound","Estimate", "Upper Bound", "TS", "p.val", "Alt.TS", "p.v
al")))

    out
}

roc.pd11 <- Rank.Correlation(ST_PD_t11[,2:3])[2:3,1:5]

pandoc.table(roc.pd11)
```

	Lower.Bound	Estimate	Upper Bound	TS	p.val
Tau.b	0.4221869	0.4240404	0.4258939	298.9289	0
Gamma	0.7977142	0.8034258	0.8091373	275.701	0

Absolute Difference of Proportions

The absolute difference (AD) in proportions is the proportion in one group minus the proportion in the other group. For example, for the personal doctor measure, the absolute difference in proportion is the difference between the proportion of insured individuals with a personal doctor and the proportion of uninsured individuals with a personal doctor.

	Personal Doctor	No Personal Doctor
Insured	351558.2	55356.99
Uninsured	36834.62	53211.23

2011 contingency table for personal doctor

$$\widehat{AD}_{pd} = \widehat{p_1} - \widehat{p_2} = \frac{351558.2}{351558.2 + 55356.99} - \frac{36834.62}{36834.62 + 53211.23} \approx 0.864 - 0.409 = 0.455$$

The standard error is calculated as if for a difference of two proportions test.

$$\widehat{SE}[AD] = \sqrt{\frac{\widehat{p_1}(1 - \widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1 - \widehat{p_2})}{n_2}}$$

$$\widehat{SE}[AD_{pd}] \approx \sqrt{\frac{0.864(1 - 0.864)}{351558.2 + 55356.99} + \frac{0.409(1 - 0.409)}{36834.62 + 53211.23}} = \sqrt{\frac{0.864(.136)}{406915.19} + \frac{0.409(.591)}{90045.85}} \approx 1.72 * 10^{-3}$$

In order to calculate the 95% confidence intervals we add and subtract $1.96 * \widehat{SE}$ to the estimated absolute difference:

$$95\% \text{ CI}[AD_{pd}] = \widehat{AD}_{pd} \pm 1.96 * \widehat{SE}[AD_{pd}] \approx 0.455 \pm 1.96(1.72 * 10^{-3}) \approx 0.455 \pm 0.00337 \approx (0.451, 0.458)$$

Relative Difference of Proportions

The relative difference of proportions is ratio of one proportion to another. For example, for the personal doctor measure, the relative difference of proportions is the ratio of the proportion of insured individuals that have a doctor to the proportion of uninsured individuals that have a doctor. So the relative difference for the personal doctor measure in 2011 is 2.11, meaning insured individuals are estimated to be about 2.11 times as likely to have a personal doctor than uninsured individuals.

	Personal Doctor	No Personal Doctor
Insured	351558.2	55356.99
Uninsured	36834.62	53211.23

2011 contingency table for personal doctor

$$\widehat{RD}_{pd} = \frac{\widehat{p_1}}{\widehat{p_2}} \approx \frac{.864}{.409} \approx 2.11$$

The confidence interval for relative difference is calculated as follows.

$$(1 - \alpha)\% \text{ CI}[RD] = \widehat{RD} * e^{\pm z_{\alpha/2} * \sqrt{v}} \quad \text{where: } v = \frac{1 - \widehat{p_1}}{n_{11}} + \frac{1 - \widehat{p_2}}{n_{12}}$$

So the 95% confidence interval for the relative difference between the insured and uninsured for the personal doctor measure for 2011 is:

$$\sqrt{v} = \sqrt{\frac{1 - \widehat{p_1}}{n_{11}} + \frac{1 - \widehat{p_2}}{n_{12}}} \approx \sqrt{\frac{.136}{351558.2} + \frac{.591}{36834.62}} \approx 4.053 * 10^{-3}$$

$$95\% \text{ CI}[RD_{pd}] = 2.11 * e^{\pm 1.96*0.004053} \approx (2.09, 2.13)$$