Chapter 6: Evaluating models

William Revelle
Northwestern University
Prepared as part of course on latent variable analysis (Psychology 454)
and as a supplement to the Short Guide to R for psychologists

February 28, 2007

Evaluating model fit in SEM may be done by examining the various indices of fit and misfit supplied by the programs used. The R**sem** package provides 8 indices which we will discuss in this chapter. LISREL, EQS, and Mplus provide even more.

Before considering the various indices of fit, let us consider the reasons that models do not fit.

1. Errors in theory
   (a) Failure to include the appropriate variables
   (b) Failure to to model the appropriate relationships
2. Errors in assumptions
   (a) Problems in distributions
   (b) Methods or correlated error factors

6.1 Model misspecification: failure to include variables

A classic problem in statements of causal structure is the failure to include appropriate variables. Such model misspecification is the bane of using correlations to infer anything about causality, for there is always the lurking third variable that could explain the relationship.

In an attempt to demonstrate this effect, consider the correlation between three variables at time 1 as predictors of an important outcome at time 2. The measured variables at time 1 are Yellow Fingers, Yellow Teeth and Bad Breath. The outcome variable is probability of Lung Cancer (rescored with a logistic trasformation to be a continous variable ranging from -3 to 3.)

For the purposes of this demonstration, we create the correlation matrix of these four variables by defining a latent variable, θ, with factor loadings theta. The product of $\theta\theta^T$ is the observed correlation matrix:

```
> theta <- matrix(c(0.8, 0.7, 0.6, 0.5), nrow = 4)
> observed <- theta %*% t(theta)
> diag(observed) <- 1
> rownames(observed) <- colnames(observed) <- c("breath", "teeth",
+     "fingers", "cancer")
> observed
```

```
        breath teeth fingers cancer
breath    1.00  0.56    0.48   0.40
teeth     0.56  1.00    0.42   0.35
fingers   0.48  0.42    1.00   0.30
cancer    0.40  0.35    0.30   1.00
```

6.1.1 Misspecified Linear Regression

Using classical linear regression, we can predict cancer risk given 1, 2, or 3 predictors. To do this from the observed correlation matrix, we can use the **solve** function, or alternatively the **mat.regress** funcion in the **psych** package. This latter function will take a correlation matrix and then find the beta weights for a set of X predictors of Y variables. We do this multiple times, first to predict regress smoking on yellow fingers, then upon yellow teeth and yellow fingers, and then finally, on breath, yellow teeth and yellow fingers.

Remember to load the **psych** package before running this analyis.

```
> library(psych)
> mat.regress(observed, 3, 4)
```

```
$beta
fingers
    0.3

$R2
cancer
  0.09
```

```
> mat.regress(observed, c(2, 3), 4)

$beta
  teeth fingers
   0.27    0.19

$R2
cancer
  0.15

> mat.regress(observed, c(1:3), 4)

$beta
 breath   teeth fingers
   0.26    0.16    0.11

$R2
cancer
  0.19
```

Note how the beta weight for yellow fingers decreases as we add more variables into the model.

We can restate the θ term in the generating model (6.1.1) as "smoking" and generate the correlation matrix again, as well as the regressions. This time we add the "smoking" variable.

```
> theta <- matrix(c(1, 0.8, 0.7, 0.6, 0.5), nrow = 5)
> observed <- theta %*% t(theta)
> diag(observed) <- 1
> rownames(observed) <- colnames(observed) <- c("smoking", "breath",
+     "teeth", "fingers", "cancer")
> observed

        smoking breath teeth fingers cancer
smoking     1.0   0.80  0.70    0.60   0.50
breath      0.8   1.00  0.56    0.48   0.40
teeth       0.7   0.56  1.00    0.42   0.35
fingers     0.6   0.48  0.42    1.00   0.30
cancer      0.5   0.40  0.35    0.30   1.00

> mat.regress(observed, 4, 5)

$beta
fingers
    0.3

$R2
cancer
  0.09

> mat.regress(observed, c(3, 4), 5)

$beta
  teeth fingers
   0.27    0.19
```

```
$R2
cancer
  0.15
```

```
> mat.regress(observed, c(2:4), 5)
```

```
$beta
 breath   teeth fingers
   0.26    0.16    0.11

$R2
cancer
  0.19
```

```
> mat.regress(observed, c(1:4), 5)
```

```
$beta
smoking  breath   teeth fingers
    0.5     0.0     0.0     0.0

$R2
cancer
  0.25
```

Notice how if the model is correctly specified (i.e., the causal variable, smoking, is introduced), the beta weights for the non-causal variables go to zero. This is understandable if we consider the beta weights in the two predictor case:

$$\left\{ \begin{array}{l} \beta_1 = (r_{x1y}r_{x2x2} - r_{x1x2}r_{x2y})/(r_{x1x1}r_{x2x2} - r^2_{x1x2}) \\ \beta_2 = (r_{x2y}r_{x1x1} - r_{x1x2}r_{x1y})/(r_{x1x1}r_{x2x2} - r^2_{x1x2}) \end{array} \right\} \tag{6.1}$$

In the more general case,

$$\beta R = r_{xy} \tag{6.2}$$

and we can solve 6.2 for β by multiplying both sides by the inverse of R.

$$\beta = \beta R R^{-1} = r_{xy}R^{-1} \tag{6.3}$$

In the two variable case (see Appendix 2), finding the inverse of a two by two matrix is discussed and is shown to be

$$R^{-1} = \begin{pmatrix} \frac{r_{22}}{r_{11}r_{22}-r^2_{12}} & -\frac{r_{12}}{r_{11}r_{22}-r^2_{12}} \\ -\frac{r_{12}}{r_{11}r_{22}-r^2_{12}} & \frac{r_{11}}{r_{11}r_{22}-r^2_{12}} \end{pmatrix} \tag{6.4}$$

6.1.2 Misspecified Structural Equation Models

In parallel with the misspecication of the linear regression, compare a series of structural equation models. The first one is fully saturated (has no degrees of freedom), and models the effect of yellow fingers as leading to cancer. Note how we are using a subset of the correlation matrix. Remember to load the **sem** package before running this analyis.

one predictor

```
> library(sem)
> model.1 <- matrix(c("fingers -> cancer", 1, NA, "fingers <-> fingers",
+     5, NA, "cancer <-> cancer", 8, NA), byrow = TRUE, ncol = 3)
> sem.1 <- sem(model.1, observed[4:5, 4:5], 100)
> summary(sem.1, digits = 2)

 Model Chisquare =  -9.6e-15   Df =  0 Pr(>Chisq) = NA
 Chisquare (null model) =  9.3   Df =  1
 Goodness-of-fit index =  1
 BIC =  -9.6e-15

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0       0       0       0       0       0

 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1 0.30     0.096     3.1     1.8e-03  cancer <--- fingers
5 1.00     0.142     7.0     2.0e-12  fingers <--> fingers
8 0.91     0.129     7.0     2.0e-12  cancer <--> cancer

 Iterations =  0
```

Note how the path coeffient for fingers -> cancer is identical to the beta weight found in the regression model.

Two predictors, don't model the correlation

A slightly more complicated model adds the effects of having yellow teeth.

```
> model.2 <- matrix(c("fingers -> cancer", 1, NA, "teeth -> cancer",
+     2, NA, "fingers <-> fingers", 5, NA, "teeth <-> teeth", 6,
+     NA, "cancer <-> cancer", 8, NA), byrow = TRUE, ncol = 3)
> sem.2 <- sem(model.2, observed[3:5, 3:5], 100)
> summary(sem.2, digits = 2)

 Model Chisquare =  19   Df =  1 Pr(>Chisq) = 1.2e-05
 Chisquare (null model) =  35   Df =  3
 Goodness-of-fit index =  0.9
 Adjusted goodness-of-fit index =  0.37
 RMSEA index =  0.43   90% CI: (0.28, 0.6)
 Bentler-Bonnett NFI =  0.46
 Tucker-Lewis NNFI =  -0.69
 Bentler CFI =  0.44
 BIC =  15

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.31    0.76    1.39    1.14    4.18
```

```
 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1 0.19     0.10      1.8     6.9e-02  cancer <--- fingers
2 0.27     0.10      2.7     7.7e-03  cancer <--- teeth
5 1.00     0.14      7.0     2.0e-12  fingers <--> fingers
6 1.00     0.14      7.0     2.0e-12  teeth <--> teeth
8 0.85     0.12      7.0     2.0e-12  cancer <--> cancer

 Iterations =  0
```

```
> print(standardized.residuals(sem.2), digits = 2)
```

```
        teeth fingers cancer
teeth   0.000    0.42  0.078
fingers 0.420    0.00  0.114
cancer  0.078    0.11  0.042
```

Even with 100 subjects, the model does not fit in terms of χ^2 or any of the conventional fit statistics. Although the path coefficients predicting cancer exactly match the regression betas, the failure to fit is due to the failure to model the correlations between the predictor variables. That is, our measurement model is faulty (because we are not actually trying to measure it.)

Two predictors, model the correlation

```
> model.3 <- matrix(c("fingers -> cancer", 1, NA, "teeth -> cancer",
+     2, NA, "fingers <-> fingers", 5, NA, "teeth <-> teeth", 6,
+     NA, "fingers <-> teeth", 7, NA, "cancer <-> cancer", 8, NA),
+     byrow = TRUE, ncol = 3)
> sem.3 <- sem(model.3, observed[3:5, 3:5], 100)
> summary(sem.3, digits = 2)
```

```
 Model Chisquare =  5.5e-15   Df =  0 Pr(>Chisq) = NA
 Chisquare (null model) =  35   Df =  3
 Goodness-of-fit index =  1
 BIC =  5.5e-15

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0       0       0       0       0       0

 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1 0.19     0.10      1.8     6.9e-02  cancer <--- fingers
2 0.27     0.10      2.7     7.7e-03  cancer <--- teeth
5 1.00     0.14      7.0     2.0e-12  fingers <--> fingers
6 1.00     0.14      7.0     2.0e-12  teeth <--> teeth
7 0.42     0.11      3.9     1.2e-04  teeth <--> fingers
8 0.85     0.12      7.0     2.0e-12  cancer <--> cancer

 Iterations =  0
```

```
> print(standardized.residuals(sem.3), digits = 2)
```

```
        teeth fingers cancer
teeth       0       0      0
fingers     0       0      0
cancer      0       0      0
```

Fitting the correlation between fingers and teeth produces a fully saturated model (with no degrees of freedom). The paths are the correct beta weights.

Three predictors, don't model the correlations

```
> model.4 <- matrix(c("fingers -> cancer", 1, NA, "teeth -> cancer",
+     2, NA, "breath -> cancer", 3, NA, "fingers <-> fingers", 5,
+     NA, "teeth <-> teeth", 6, NA, "breath <-> breath", 7, NA,
+     "cancer <-> cancer", 8, NA), byrow = TRUE, ncol = 3)
> sem.4 <- sem(model.4, observed[2:5, 2:5], 100)
> summary(sem.4, digits = 2)
```

```
 Model Chisquare =  68   Df =  3 Pr(>Chisq) = 1.4e-14
 Chisquare (null model) =  89   Df =  6
 Goodness-of-fit index =  0.74
 Adjusted goodness-of-fit index =  0.12
 RMSEA index =  0.47   90% CI: (0.37, 0.57)
 Bentler-Bonnett NFI =  0.24
 Tucker-Lewis NNFI =  -0.56
 Bentler CFI =  0.22
 BIC =  54

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0     1.2     2.0     2.5     4.3     5.6

 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1 0.11     0.11      1.0     3.0e-01  cancer <--- fingers
2 0.16     0.11      1.4     1.5e-01  cancer <--- teeth
3 0.26     0.12      2.2     2.5e-02  cancer <--- breath
5 1.00     0.14      7.0     2.0e-12  fingers <--> fingers
6 1.00     0.14      7.0     2.0e-12  teeth <--> teeth
7 1.00     0.14      7.0     2.0e-12  breath <--> breath
8 0.81     0.11      7.0     2.0e-12  cancer <--> cancer

 Iterations =  0
```

```
> print(standardized.residuals(sem.4), digits = 2)
```

```
        breath teeth fingers cancer
breath    0.00  0.56    0.48  0.142
teeth     0.56  0.00    0.42  0.190
fingers   0.48  0.42    0.00  0.191
cancer    0.14  0.19    0.19  0.088
```

Once again, although the prediction paths from the predictors to the criterion match the beta weights, the model does not fit, because this model fails to model the correlation between the predictors. Once again, our failure to have a measurement model is at fault.

We can fix the variance of the predictors to increase the degrees of freedom, but we are still not modeling the covariances. (Is this a failure of the sem package, or is this a general problem?)

```
> model.4a <- matrix(c("fingers -> cancer", 1, NA, "teeth -> cancer",
+     2, NA, "breath -> cancer", 3, NA, "fingers <-> fingers", NA,
+     1, "teeth <-> teeth", NA, 1, "breath <-> breath", NA, 1, "cancer <-> cancer",
+     8, NA), byrow = TRUE, ncol = 3)
> sem.4a <- sem(model.4a, observed[2:5, 2:5], 100)
> summary(sem.4a, digits = 2)

 Model Chisquare =  68   Df =  6 Pr(>Chisq) = 1.3e-12
 Chisquare (null model) =  89   Df =  6
 Goodness-of-fit index =  0.74
 Adjusted goodness-of-fit index =  0.56
 RMSEA index =  0.32   90% CI: (0.26, 0.39)
 Bentler-Bonnett NFI =  0.24
 Tucker-Lewis NNFI =  0.26
 Bentler CFI =  0.26
 BIC =  40

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0     1.2     2.0     2.5     4.3     5.6

 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1 0.11     0.11      1.0     3.0e-01  cancer <--- fingers
2 0.16     0.11      1.4     1.5e-01  cancer <--- teeth
3 0.26     0.12      2.2     2.5e-02  cancer <--- breath
8 0.81     0.11      7.0     2.0e-12  cancer <--> cancer

 Iterations =  0

> print(standardized.residuals(sem.4a), digits = 2)

        breath teeth fingers cancer
breath    0.00  0.56    0.48  0.142
teeth     0.56  0.00    0.42  0.190
fingers   0.48  0.42    0.00  0.191
cancer    0.14  0.19    0.19  0.088
```

Three predictors, model the correlations, case 1

Revise the previous model to include a "yellow" latent variable. That is, we notice from the residuals that yellow teeth and fingers seem to go together. Perhaps, with a bit of creativity, we can explain this as due to the influence of yellowing agents which need to be controlled.

```
> model.5 <- matrix(c("fingers -> cancer", 1, NA, "teeth -> cancer",
+     2, NA, "breath -> cancer", 3, NA, "fingers <-> fingers", 5,
+     NA, "teeth <-> teeth", 6, NA, "breath <-> breath", 7, NA,
+     "cancer <-> cancer", 8, NA, "yellow <-> yellow", NA, 1, "yellow -> fingers",
+     10, NA, "yellow -> teeth", NA, 1), byrow = TRUE, ncol = 3)
> model.5
      [,1]                  [,2] [,3]
 [1,] "fingers -> cancer"   "1"  NA
 [2,] "teeth -> cancer"     "2"  NA
 [3,] "breath -> cancer"    "3"  NA
 [4,] "fingers <-> fingers" "5"  NA
 [5,] "teeth <-> teeth"     "6"  NA
 [6,] "breath <-> breath"   "7"  NA
 [7,] "cancer <-> cancer"   "8"  NA
 [8,] "yellow <-> yellow"   NA   "1"
 [9,] "yellow -> fingers"   "10" NA
[10,] "yellow -> teeth"     NA   "1"

> sem.5 <- sem(model.5, observed[2:5, 2:5], 100)
> summary(sem.5, digits = 2)

 Model Chisquare =  48   Df =  2 Pr(>Chisq) = 3.2e-11
 Chisquare (null model) =  89   Df =  6
 Goodness-of-fit index =  0.84
 Adjusted goodness-of-fit index =  0.19
 RMSEA index =  0.48   90% CI: (0.37, 0.61)
 Bentler-Bonnett NFI =  0.45
 Tucker-Lewis NNFI =  -0.68
 Bentler CFI =  0.44
 BIC =  39

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0e+00 7.1e-06 1.3e+00 1.9e+00 2.3e+00 5.6e+00

 Parameter Estimates
   Estimate Std Error z value  Pr(>|z|)
1   1.1e-01 0.11       1.0e+00 3.0e-01  cancer <--- fingers
2   1.6e-01 0.11       1.4e+00 1.5e-01  cancer <--- teeth
3   2.6e-01 0.12       2.2e+00 2.5e-02  cancer <--- breath
5   8.2e-01 0.12       6.9e+00 6.1e-12  fingers <--> fingers
6  -6.5e-07 0.14      -4.6e-06 1.0e+00  teeth <--> teeth
7   1.0e+00 0.14       7.0e+00 2.0e-12  breath <--> breath
8   8.1e-01 0.11       7.0e+00 2.0e-12  cancer <--> cancer
10  4.2e-01 0.11       3.9e+00 1.2e-04  fingers <--- yellow

 Iterations =  14

> print(standardized.residuals(sem.5), digits = 2)

        breath   teeth fingers cancer
breath    0.00 5.6e-01 4.8e-01  0.142
teeth     0.56 6.5e-07 7.8e-07  0.145
```

```
fingers   0.48 7.8e-07 4.0e-07  0.124
cancer    0.14 1.4e-01 1.2e-01  0.073
```

This model is significant improvement over the previous model, (examine the change in χ^2 for the one degree of freedom used), but still does not fit very well.

Three predictors, model the correlations, case 2

Looking at the residuals suggests perhaps we should model a latent mouth variable as well. Perhaps the yellowing of the teeth have an additional component related to being in the mouth.

```
> model.6 <- matrix(c("fingers -> cancer", 1, NA, "teeth -> cancer",
+     2, NA, "breath -> cancer", 3, NA, "fingers <-> fingers", 5,
+     NA, "teeth <-> teeth", 6, NA, "breath <-> breath", 7, NA,
+     "cancer <-> cancer", 8, NA, "yellow <-> yellow", NA, 1, "yellow -> fingers",
+     10, NA, "yellow -> teeth", NA, 1, "mouth -> teeth", NA, 1,
+     "mouth -> breath", 11, NA, "mouth <-> mouth", NA, 1), byrow = TRUE,
+     ncol = 3)
> model.6
```

```
      [,1]                  [,2] [,3]
 [1,] "fingers -> cancer"   "1"  NA
 [2,] "teeth -> cancer"     "2"  NA
 [3,] "breath -> cancer"    "3"  NA
 [4,] "fingers <-> fingers" "5"  NA
 [5,] "teeth <-> teeth"     "6"  NA
 [6,] "breath <-> breath"   "7"  NA
 [7,] "cancer <-> cancer"   "8"  NA
 [8,] "yellow <-> yellow"   NA   "1"
 [9,] "yellow -> fingers"   "10" NA
[10,] "yellow -> teeth"     NA   "1"
[11,] "mouth -> teeth"      NA   "1"
[12,] "mouth -> breath"     "11" NA
[13,] "mouth <-> mouth"     NA   "1"
```

```
> sem.6 <- sem(model.6, observed[2:5, 2:5], 100)
> summary(sem.6, digits = 2)
```

```
 Model Chisquare =  26   Df =  1 Pr(>Chisq) = 3.5e-07
 Chisquare (null model) =  89   Df =  6
 Goodness-of-fit index =  0.9
 Adjusted goodness-of-fit index =  -0.033
 RMSEA index =  0.5   90% CI: (0.35, 0.68)
 Bentler-Bonnett NFI =  0.71
 Tucker-Lewis NNFI =  -0.81
 Bentler CFI =  0.7
 BIC =  21

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-8.9e-07  6.4e-01  7.8e-01  1.4e+00  1.8e+00  4.8e+00
```

```
 Parameter Estimates
   Estimate Std Error z value Pr(>|z|)
1   0.11    0.105      1.0    3.0e-01  cancer <--- fingers
2   0.16    0.111      1.4    1.5e-01  cancer <--- teeth
3   0.26    0.115      2.2    2.5e-02  cancer <--- breath
5   0.96    0.136      7.1    1.7e-12  fingers <--> fingers
6  -1.09    0.124     -8.7    0.0e+00  teeth <--> teeth
7   0.78    0.118      6.6    3.4e-11  breath <--> breath
8   0.81    0.115      7.0    2.0e-12  cancer <--> cancer
10  0.20    0.097      2.0    4.3e-02  fingers <--- yellow
11  0.47    0.114      4.1    4.4e-05  breath <--- mouth

 Iterations =  21
```

```
> print(standardized.residuals(sem.6), digits = 2)
```

```
         breath teeth  fingers cancer
breath  1.3e-07 0.094  4.8e-01  0.067
teeth   9.4e-02 0.088  2.2e-01  0.063
fingers 4.8e-01 0.224 -1.3e-07  0.160
cancer  6.7e-02 0.063  1.6e-01  0.045
```

This is a great improvement (once again, look at the change in χ^2 for the 1 degree of freedom more complex model), but the model still does not fit at all well.

Three predictors, model the correlations, case 3

Alternatively, we could just allow all the predictors to correlate:

```
> model.7 <- matrix(c("fingers -> cancer", 1, NA, "teeth -> cancer",
+     2, NA, "breath -> cancer", 3, NA, "fingers <-> fingers", 5,
+     NA, "teeth <-> teeth", 6, NA, "breath <-> breath", 7, NA,
+     "cancer <-> cancer", 8, NA, "teeth <-> breath", 9, NA, "teeth <-> fingers",
+     10, NA, "fingers <-> breath", 11, NA), byrow = TRUE, ncol = 3)
> sem.7 <- sem(model.7, observed[2:5, 2:5], 100)
> summary(sem.7, digits = 2)
```

```
 Model Chisquare =  2.2e-14   Df =  0 Pr(>Chisq) = NA
 Chisquare (null model) =  89   Df =  6
 Goodness-of-fit index =  1
 BIC =  2.2e-14

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0e+00 0.0e+00 0.0e+00 1.3e-16 1.3e-16 5.3e-16

 Parameter Estimates
   Estimate Std Error z value Pr(>|z|)
1  0.11     0.11      1.0     3.0e-01  cancer <--- fingers
2  0.16     0.11      1.4     1.5e-01  cancer <--- teeth
3  0.26     0.12      2.2     2.5e-02  cancer <--- breath
```

```
5  1.00     0.14      7.0     2.0e-12  fingers <--> fingers
6  1.00     0.14      7.0     2.0e-12  teeth <--> teeth
7  1.00     0.14      7.0     2.0e-12  breath <--> breath
8  0.81     0.11      7.0     2.0e-12  cancer <--> cancer
9  0.56     0.12      4.9     1.2e-06  breath <--> teeth
10 0.42     0.11      3.9     1.2e-04  fingers <--> teeth
11 0.48     0.11      4.3     1.7e-05  breath <--> fingers

 Iterations =  0
```

```
> print(standardized.residuals(sem.7), digits = 2)
```

```
         breath teeth fingers  cancer
breath  0.0e+00     0 0.0e+00 5.6e-17
teeth   0.0e+00     0 0.0e+00 0.0e+00
fingers 0.0e+00     0 0.0e+00 5.6e-17
cancer  5.6e-17     0 5.6e-17 0.0e+00
```

This model is fully saturated, and thus the χ^2 statistic is meaningless. The β weights match the regression model, and the modeled correlations match the data.

However, if we fix the variances of the three predictors to be 1, then we have gained 3 degrees of freedom and now the model looks great!

```
> model.7a <- matrix(c("fingers -> cancer", 1, NA, "teeth -> cancer",
+     2, NA, "breath -> cancer", 3, NA, "fingers <-> fingers", NA,
+     1, "teeth <-> teeth", NA, 1, "breath <-> breath", NA, 1, "cancer <-> cancer",
+     8, NA, "teeth <-> breath", 9, NA, "teeth <-> fingers", 10,
+     NA, "fingers <-> breath", 11, NA), byrow = TRUE, ncol = 3)
> sem.7a <- sem(model.7a, observed[2:5, 2:5], 100)
> summary(sem.7a, digits = 2)
```

```
 Model Chisquare =  2.2e-14   Df =  3 Pr(>Chisq) = 1
 Chisquare (null model) =  89   Df =  6
 Goodness-of-fit index =  1
 Adjusted goodness-of-fit index =  1
 RMSEA index =  0   90% CI: (NA, NA)
 Bentler-Bonnett NFI =  1
 Tucker-Lewis NNFI =  1.1
 Bentler CFI =  1
 BIC =  -14

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0e+00 0.0e+00 0.0e+00 1.3e-16 1.3e-16 5.3e-16

 Parameter Estimates
   Estimate Std Error z value Pr(>|z|)
1  0.11     0.105     1.0     3.0e-01  cancer <--- fingers
2  0.16     0.111     1.4     1.5e-01  cancer <--- teeth
3  0.26     0.115     2.2     2.5e-02  cancer <--- breath
8  0.81     0.115     7.0     2.0e-12  cancer <--> cancer
9  0.56     0.060     9.3     0.0e+00  breath <--> teeth
10 0.42     0.075     5.6     2.6e-08  fingers <--> teeth
```

```
11 0.48     0.069     6.9     4.6e-12  breath <--> fingers

 Iterations =  0
```

```
> print(standardized.residuals(sem.7a), digits = 2)
```

```
         breath teeth fingers  cancer
breath  0.0e+00     0 0.0e+00 5.6e-17
teeth   0.0e+00     0 0.0e+00 0.0e+00
fingers 0.0e+00     0 0.0e+00 5.6e-17
cancer  5.6e-17     0 5.6e-17 0.0e+00
```

6.1.3 Three predictors, model the correlations with one latent variable

An alternative model is to note that the three predictors correlate and to consider that perhaps they reflect an unknown, latent variable. Perhaps it is this latent variable which leads to cancer.

```
> model.8 <- matrix(c("latent -> cancer", 1, NA, "latent -> breath",
+     2, NA, "latent -> fingers", 3, NA, "latent -> teeth", 4, NA,
+     "fingers <-> fingers", 5, NA, "teeth <-> teeth", 6, NA, "breath <-> breath",
+     7, NA, "cancer <-> cancer", 8, NA, "latent <-> latent", NA,
+     1), byrow = TRUE, ncol = 3)
> model.8
```

```
      [,1]                  [,2] [,3]
 [1,] "latent -> cancer"    "1"  NA
 [2,] "latent -> breath"    "2"  NA
 [3,] "latent -> fingers"   "3"  NA
 [4,] "latent -> teeth"     "4"  NA
 [5,] "fingers <-> fingers" "5"  NA
 [6,] "teeth <-> teeth"     "6"  NA
 [7,] "breath <-> breath"   "7"  NA
 [8,] "cancer <-> cancer"   "8"  NA
 [9,] "latent <-> latent"   NA   "1"
```

```
> sem.8 <- sem(model.8, observed[2:5, 2:5], 100)
> summary(sem.8, digits = 2)
```

```
 Model Chisquare =  1.9e-10   Df =  2 Pr(>Chisq) = 1
 Chisquare (null model) =  89   Df =  6
 Goodness-of-fit index =  1
 Adjusted goodness-of-fit index =  1
 RMSEA index =  0   90% CI: (NA, NA)
 Bentler-Bonnett NFI =  1
 Tucker-Lewis NNFI =  1.1
 Bentler CFI =  1
 BIC =  -9.2

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
9.4e-07 1.7e-06 3.0e-06 3.5e-06 4.7e-06 1.2e-05
```

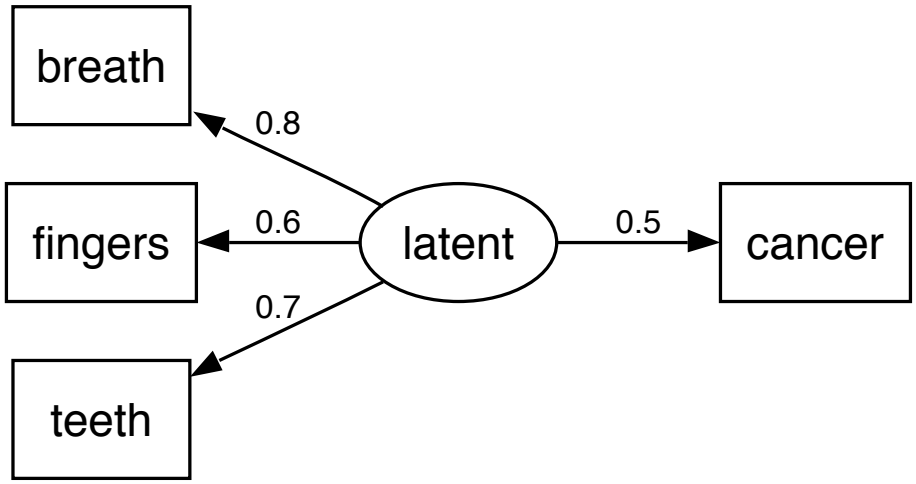

Figure 6.1: Whatever is common to bad breadth, yellow teeth, and yellow hands seems to lead to lung cancer. Thus, one should use mouth freshners, visit your dentist, and wear latex gloves.

```
 Parameter Estimates
  Estimate Std Error z value Pr(>|z|)
1 0.50     0.11      4.7     2.8e-06  cancer <--- latent
2 0.80     0.10      7.8     5.1e-15  breath <--- latent
3 0.60     0.10      5.8     8.2e-09  fingers <--- latent
4 0.70     0.10      6.8     9.8e-12  teeth <--- latent
5 0.64     0.11      5.9     4.8e-09  fingers <--> fingers
6 0.51     0.10      4.9     1.2e-06  teeth <--> teeth
7 0.36     0.11      3.3     9.1e-04  breath <--> breath
8 0.75     0.12      6.4     2.0e-10  cancer <--> cancer

 Iterations =  13
```

```
> print(standardized.residuals(sem.8), digits = 2)
```

```
         breath   teeth fingers  cancer
breath  6.1e-07 5.4e-07 3.5e-07 5.2e-07
teeth   5.4e-07 2.8e-07 1.1e-07 2.9e-07
fingers 3.5e-07 1.1e-07 1.3e-07 1.8e-07
cancer  5.2e-07 2.9e-07 1.8e-07 1.7e-06
```

Ah, that did it. We now understand the "causal" structure (although our inference about what is common between bad breath, yellow teeth and yellow fingers will probably ignore the real cause). The secret to solving lung cancer is to use mouth freshners, visit your dentist, and wear latex gloves! (See Figure 6.2

6.2 Including the correct variables, but misspecifying the models

Based upon the previous model fitting in section 6.1.1 we have concluded that there is some latent variable that ties our four variables together. We now examine what happens when we add yet another variable to the mix.

We use the correlation matrix from section 6.1.1. Note that the correlation matrix is identical for the previous four variables, and that the smoking variable is equivalent to the latent factor that generated the data.

6.2.1 Including the correct variables in linear regression

Remember that if we include smoking into the linear regression, the effect of the other variables vanishes:

```
> mat.regress(observed, c(2:4), 5)
```

```
$beta
 breath   teeth fingers
   0.26    0.16    0.11

$R2
cancer
  0.19
```

```
> mat.regress(observed, c(1:4), 5)
```

```
$beta
smoking  breath   teeth fingers
    0.5     0.0     0.0     0.0

$R2
cancer
  0.25
```

If, however, we were to make smoking an unreliable measure and thus not perfectly correlated with the latent factor, the other variables still seem to have an effect:

```
> theta <- matrix(c(0.9, 0.8, 0.7, 0.6, 0.5), nrow = 5)
> observed1 <- theta %*% t(theta)
> diag(observed1) <- 1
> rownames(observed1) <- colnames(observed1) <- c("smoking", "breath",
+     "teeth", "fingers", "cancer")
> observed1
```

```
        smoking breath teeth fingers cancer
smoking    1.00   0.72  0.63    0.54   0.45
breath     0.72   1.00  0.56    0.48   0.40
teeth      0.63   0.56  1.00    0.42   0.35
fingers    0.54   0.48  0.42    1.00   0.30
cancer     0.45   0.40  0.35    0.30   1.00
```

```
> mat.regress(observed1, c(2:4), 5)

$beta
 breath   teeth fingers
   0.26    0.16    0.11

$R2
cancer
  0.19

> mat.regress(observed1, c(1:4), 5)

$beta
smoking  breath   teeth fingers
   0.28    0.13    0.08    0.05

$R2
cancer
  0.22
```

Compare the regression weights for the two data sets (observed and observed1). Note how the other variables still contribute to the regression unless smoking is measured perfectly reliably.

Do this one more time, with a very unreliable measure of smoking:

```
> theta <- matrix(c(0.5, 0.8, 0.7, 0.6, 0.5), nrow = 5)
> observed2 <- theta %*% t(theta)
> diag(observed2) <- 1
> rownames(observed2) <- colnames(observed2) <- c("smoking", "breath",
+     "teeth", "fingers", "cancer")
> observed2

        smoking breath teeth fingers cancer
smoking    1.00   0.40  0.35    0.30   0.25
breath     0.40   1.00  0.56    0.48   0.40
teeth      0.35   0.56  1.00    0.42   0.35
fingers    0.30   0.48  0.42    1.00   0.30
cancer     0.25   0.40  0.35    0.30   1.00

> mat.regress(observed2, c(2:4), 5)

$beta
 breath   teeth fingers
   0.26    0.16    0.11

$R2
cancer
  0.19

> mat.regress(observed2, c(1:4), 5)

$beta
smoking  breath   teeth fingers
   0.07    0.24    0.15    0.10
```

```
$R2
cancer
   0.2
```

Note that in this case, we completely over estimate the contribution of the other variables and underestimate the contribution of smoking.

6.2.2 Including the correct variables in the Structural Equation

```
> model.9 <- matrix(c("latent -> cancer", 1, NA, "latent -> breath",
+     2, NA, "latent -> fingers", 3, NA, "latent -> teeth", 4, NA,
+     "latent -> smoking", 9, NA, "fingers <-> fingers", 5, NA,
+     "teeth <-> teeth", 6, NA, "breath <-> breath", 7, NA, "cancer <-> cancer",
+     8, NA, "smoking <-> smoking", 10, NA, "latent <-> latent",
+     NA, 1), byrow = TRUE, ncol = 3)
> model.9
      [,1]                  [,2] [,3]
 [1,] "latent -> cancer"    "1"  NA
 [2,] "latent -> breath"    "2"  NA
 [3,] "latent -> fingers"   "3"  NA
 [4,] "latent -> teeth"     "4"  NA
 [5,] "latent -> smoking"   "9"  NA
 [6,] "fingers <-> fingers" "5"  NA
 [7,] "teeth <-> teeth"     "6"  NA
 [8,] "breath <-> breath"   "7"  NA
 [9,] "cancer <-> cancer"   "8"  NA
[10,] "smoking <-> smoking" "10" NA
[11,] "latent <-> latent"   NA   "1"
> sem.9 <- sem(model.9, observed, 100)
> summary(sem.9, digits = 2)

 Model Chisquare =  1.8e-11   Df =  5 Pr(>Chisq) = 1
 Chisquare (null model) =  240   Df =  10
 Goodness-of-fit index =  1
 Adjusted goodness-of-fit index =  1
 RMSEA index =  0   90% CI: (NA, NA)
 Bentler-Bonnett NFI =  1
 Tucker-Lewis NNFI =  1.0
 Bentler CFI =  1
 BIC =  -23

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-8.2e-07 -3.6e-07  5.1e-08  2.3e-07  8.5e-07  1.2e-06

 Parameter Estimates
   Estimate Std Error z value  Pr(>|z|)
1   5.0e-01 0.095      5.3e+00 1.3e-07  cancer <--- latent
2   8.0e-01 0.085      9.4e+00 0.0e+00  breath <--- latent
3   6.0e-01 0.092      6.5e+00 7.5e-11  fingers <--- latent
```

```
4   7.0e-01 0.089      7.9e+00 3.6e-15  teeth <--- latent
9   1.0e+00 0.075      1.3e+01 0.0e+00  smoking <--- latent
5   6.4e-01 0.093      6.9e+00 4.9e-12  fingers <--> fingers
6   5.1e-01 0.076      6.7e+00 2.3e-11  teeth <--> teeth
7   3.6e-01 0.060      6.0e+00 1.7e-09  breath <--> breath
8   7.5e-01 0.107      7.0e+00 2.8e-12  cancer <--> cancer
10 -7.1e-08 0.048     -1.5e-06 1.0e+00  smoking <--> smoking

 Iterations =  15
```

```
> print(standardized.residuals(sem.9), digits = 2)
```

```
         smoking   breath    teeth  fingers  cancer
smoking -1.9e-08 -3.1e-08 -1.0e-07 -2.4e-08 1.2e-07
breath  -3.1e-08  9.0e-08 -5.1e-08  5.7e-09 1.2e-07
teeth   -1.0e-07 -5.1e-08 -6.0e-08 -3.9e-08 6.7e-08
fingers -2.4e-08  5.7e-09 -3.9e-08  1.7e-07 8.9e-08
cancer   1.2e-07  1.2e-07  6.7e-08  8.9e-08 1.6e-07
```

Note that with the perfect data set, the estimate for the error variance of smoking is appropriately very small.

Repeat this analysis with the less than perfect reliability of smoking of the observed1 data set:

```
> sem.10 <- sem(model.9, observed1, 100)
> summary(sem.10, digits = 2)
```

```
 Model Chisquare =  1.1e-10   Df =  5 Pr(>Chisq) = 1
 Chisquare (null model) =  188   Df =  10
 Goodness-of-fit index =  1
 Adjusted goodness-of-fit index =  1
 RMSEA index =  0   90% CI: (NA, NA)
 Bentler-Bonnett NFI =  1
 Tucker-Lewis NNFI =  1.1
 Bentler CFI =  1
 BIC =  -23

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-3.8e-06 -2.5e-06  3.1e-07  2.7e-07  2.7e-06  5.7e-06

 Parameter Estimates
   Estimate Std Error z value Pr(>|z|)
1  0.50     0.099      5.0    5.0e-07  cancer <--- latent
2  0.80     0.088      9.1    0.0e+00  breath <--- latent
3  0.60     0.096      6.2    4.5e-10  fingers <--- latent
4  0.70     0.092      7.6    3.3e-14  teeth <--- latent
9  0.90     0.084     10.7    0.0e+00  smoking <--- latent
5  0.64     0.098      6.5    7.7e-11  fingers <--> fingers
6  0.51     0.084      6.1    1.1e-09  teeth <--> teeth
7  0.36     0.070      5.1    3.0e-07  breath <--> breath
8  0.75     0.111      6.7    1.7e-11  cancer <--> cancer
10 0.19     0.064      3.0    2.8e-03  smoking <--> smoking
```

```
 Iterations =  12
```

```
> print(standardized.residuals(sem.10), digits = 2)
```

```
         smoking   breath    teeth fingers   cancer
smoking  1.5e-07  8.2e-08 -1.8e-07 6.5e-07 -2.7e-07
breath   8.2e-08  4.4e-08 -2.2e-07 5.3e-07 -2.8e-07
teeth   -1.8e-07 -2.2e-07  2.9e-09 3.0e-07 -3.9e-07
fingers  6.5e-07  5.3e-07  3.0e-07 5.1e-07  1.2e-07
cancer  -2.7e-07 -2.8e-07 -3.9e-07 1.2e-07 -5.4e-07
```

Repeat this analysis with the even less reliability of smoking of the observed2 data set:

```
> sem.11 <- sem(model.9, observed2, 100)
> summary(sem.11, digits = 2)
```

```
 Model Chisquare =  4.2e-10   Df =  5 Pr(>Chisq) = 1
 Chisquare (null model) =  110   Df =  10
 Goodness-of-fit index =  1
 Adjusted goodness-of-fit index =  1
 RMSEA index =  0   90% CI: (NA, NA)
 Bentler-Bonnett NFI =  1
 Tucker-Lewis NNFI =  1.1
 Bentler CFI =  1
 BIC =  -23

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-1.0e-05 -2.4e-06 -1.1e-06 -5.4e-07  2.2e-06  5.5e-06

 Parameter Estimates
   Estimate Std Error z value Pr(>|z|)
1  0.50     0.105     4.7     2.1e-06  cancer <--- latent
2  0.80     0.098     8.2     2.2e-16  breath <--- latent
3  0.60     0.102     5.9     4.6e-09  fingers <--- latent
4  0.70     0.100     7.0     2.3e-12  teeth <--- latent
9  0.50     0.105     4.7     2.1e-06  smoking <--- latent
5  0.64     0.107     6.0     2.2e-09  fingers <--> fingers
6  0.51     0.099     5.1     2.7e-07  teeth <--> teeth
7  0.36     0.097     3.7     2.1e-04  breath <--> breath
8  0.75     0.117     6.4     1.3e-10  cancer <--> cancer
10 0.75     0.117     6.4     1.3e-10  smoking <--> smoking

 Iterations =  11
```

```
> print(standardized.residuals(sem.11), digits = 2)
```

```
         smoking   breath    teeth  fingers   cancer
smoking  7.7e-07 -2.4e-07  5.2e-07 -2.6e-07  2.3e-07
breath  -2.4e-07 -6.2e-07 -1.4e-08 -1.1e-06 -2.4e-07
teeth    5.2e-07 -1.4e-08 -2.7e-07 -1.2e-07  5.2e-07
fingers -2.6e-07 -1.1e-06 -1.2e-07 -1.1e-07 -2.6e-07
cancer   2.3e-07 -2.4e-07  5.2e-07 -2.6e-07  7.7e-07
```

We now see the real power of the SEM approach. For by modeling the correlations between the X predictor set, we are able to correct for unreliability and see the structure of the data. But, the conclusion is still wrong, because now we are forced to interpret that whatever it is that is common to smoking, bad breath, yellow fingers and yellow teeth lead to cancer. Although our latent modeling approach has helped and is able to reproduce the data perfectly, it has not led to the correct conclusion as to causality. (See Figure **??**).

6.2.3 Direct the causal path

What happens if we make smoking a causal variable that leads to the latent variable?

```
> model.12 <- matrix(c("latent -> cancer", 1, NA, "latent -> breath",
+     2, NA, "latent -> fingers", 3, NA, "latent -> teeth", 4, NA,
+     "smoking -> latent", NA, 1, "fingers <-> fingers", 5, NA,
+     "teeth <-> teeth", 6, NA, "breath <-> breath", 7, NA, "cancer <-> cancer",
+     8, NA, "smoking <-> smoking", NA, 1, "latent <-> latent",
+     12, NA), byrow = TRUE, ncol = 3)
> sem.12 <- sem(model.12, observed, 100)
> summary(sem.12, digits = 2)

 Model Chisquare =  4.9e-12   Df =  6 Pr(>Chisq) = 1
 Chisquare (null model) =  240   Df =  10
 Goodness-of-fit index =  1
 Adjusted goodness-of-fit index =  1
 RMSEA index =  0   90% CI: (NA, NA)
 Bentler-Bonnett NFI =  1
 Tucker-Lewis NNFI =  1.0
 Bentler CFI =  1
 BIC =  -28

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-5.8e-07 -1.0e-07  1.7e-07  3.1e-07  4.4e-07  1.2e-06

 Parameter Estimates
   Estimate Std Error z value Pr(>|z|)
1  5.0e-01  0.087     5.7e+00 9.2e-09  cancer <--- latent
2  8.0e-01  0.060     1.3e+01 0.0e+00  breath <--- latent
3  6.0e-01  0.080     7.5e+00 8.5e-14  fingers <--- latent
4  7.0e-01  0.072     9.8e+00 0.0e+00  teeth <--- latent
5  6.4e-01  0.093     6.9e+00 4.9e-12  fingers <--> fingers
6  5.1e-01  0.076     6.7e+00 2.3e-11  teeth <--> teeth
7  3.6e-01  0.060     6.0e+00 1.7e-09  breath <--> breath
8  7.5e-01  0.107     7.0e+00 2.8e-12  cancer <--> cancer
12 9.7e-09  0.048     2.0e-07 1.0e+00  latent <--> latent

 Iterations =  15

> print(standardized.residuals(sem.12), digits = 2)
```

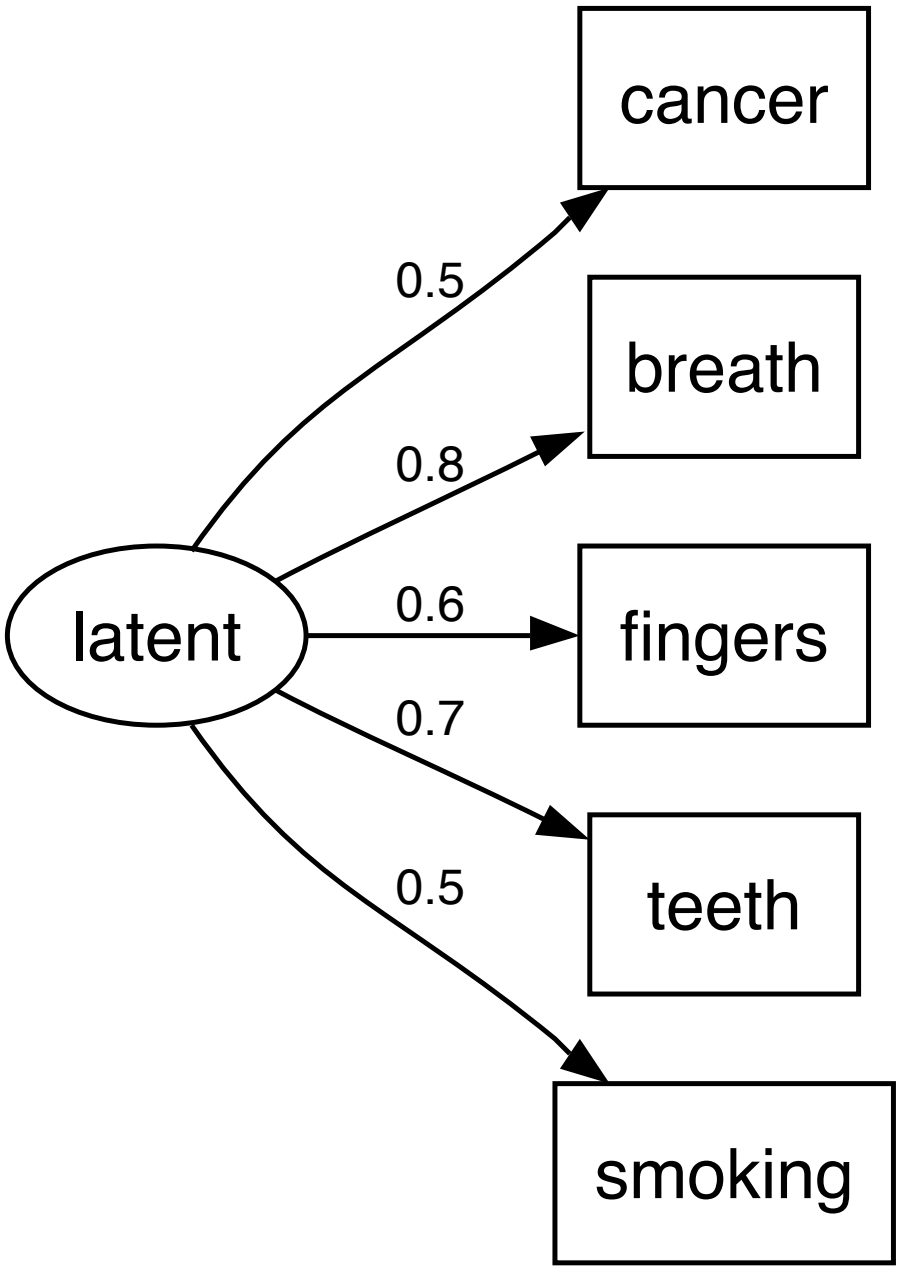

Figure 6.2: Whatever is common to smoking, bad breadth, yellow teeth, and yellow hands seems to lead to lung cancer. Thus, one should use mouth freshners, visit your dentist, and wear latex gloves. It is unclear why the latent variable leads to smoking

```
         smoking  breath   teeth  fingers   cancer
smoking  0.0e+00 5.5e-08 1.0e-07 -1.4e-08 -1.2e-08
breath   5.5e-08 1.4e-07 1.2e-07  1.7e-08  1.5e-08
teeth    1.0e-07 1.2e-07 1.7e-07  4.8e-08  4.0e-08
fingers -1.4e-08 1.7e-08 4.8e-08  2.5e-08 -1.7e-08
cancer  -1.2e-08 1.5e-08 4.0e-08 -1.7e-08 -8.2e-08
```

Repeat this analysis with noisy data from observed2. (Remember that in this case, smoking is not measured reliably).

```
> sem.13 <- sem(model.12, observed2, 100)
> summary(sem.13, digits = 2)

 Model Chisquare =  7.4e-11   Df =  6 Pr(>Chisq) = 1
 Chisquare (null model) =  110   Df =  10
 Goodness-of-fit index =  1
 Adjusted goodness-of-fit index =  1
 RMSEA index =  0   90% CI: (NA, NA)
 Bentler-Bonnett NFI =  1
 Tucker-Lewis NNFI =  1.1
 Bentler CFI =  1
 BIC =  -28

 Normalized Residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-5.1e-06 -2.7e-06 -2.4e-06 -2.3e-06 -1.1e-06  6.4e-07

 Parameter Estimates
   Estimate Std Error z value Pr(>|z|)
1  0.25     0.070     3.6     3.6e-04  cancer <--- latent
2  0.40     0.087     4.6     4.0e-06  breath <--- latent
3  0.30     0.075     4.0     7.1e-05  fingers <--- latent
4  0.35     0.082     4.3     1.8e-05  teeth <--- latent
5  0.64     0.107     6.0     2.2e-09  fingers <--> fingers
6  0.51     0.099     5.1     2.7e-07  teeth <--> teeth
7  0.36     0.097     3.7     2.1e-04  breath <--> breath
8  0.75     0.117     6.4     1.3e-10  cancer <--> cancer
12 3.00     1.364     2.2     2.8e-02  latent <--> latent

 Iterations =  18

> print(standardized.residuals(sem.13), digits = 2)

         smoking   breath    teeth  fingers   cancer
smoking  0.0e+00 -3.0e-07 -1.2e-07 -1.2e-07 -2.4e-07
breath  -3.0e-07 -7.3e-07 -3.1e-07 -2.9e-07 -4.8e-07
teeth   -1.2e-07 -3.1e-07  9.1e-08 -8.8e-08 -2.8e-07
fingers -1.2e-07 -2.9e-07 -8.8e-08 -7.1e-07 -2.5e-07
cancer  -2.4e-07 -4.8e-07 -2.8e-07 -2.5e-07 -2.8e-07
```

We can also model smoking as a noisy variable, and then fix one path (in this case, the latent to cancer) to estimate the model for pure, moderate, and very noisy smoking.

```
> model.14 <- matrix(c("latent -> cancer", NA, 1, "latent -> breath",
+     2, NA, "latent -> fingers", 3, NA, "latent -> teeth", 4, NA,
+     "smoking -> latent", 11, NA, "fingers <-> fingers", 5, NA,
+     "teeth <-> teeth", 6, NA, "breath <-> breath", 7, NA, "cancer <-> cancer",
+     8, NA, "smoking <-> smoking", 10, NA, "latent <-> latent",
+     NA, 1), byrow = TRUE, ncol = 3)
> sem.14 <- sem(model.14, observed2, 100)
> summary(sem.14, digits = 2)

 Model Chisquare =  26   Df =  6 Pr(>Chisq) = 0.00021
 Chisquare (null model) =  110   Df =  10
 Goodness-of-fit index =  0.92
 Adjusted goodness-of-fit index =  0.8
 RMSEA index =  0.18   90% CI: (0.12, 0.26)
 Bentler-Bonnett NFI =  0.76
 Tucker-Lewis NNFI =  0.67
 Bentler CFI =  0.8
 BIC =  -1.5

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   -3.4    -2.8    -1.4    -1.4     0.0     0.4

 Parameter Estimates
   Estimate Std Error z value Pr(>|z|)
2  0.88     0.108     8.1     6.7e-16  breath <--- latent
3  0.67     0.114     5.9     4.4e-09  fingers <--- latent
4  0.78     0.110     7.1     1.6e-12  teeth <--- latent
11 0.40     0.113     3.6     3.7e-04  latent <--- smoking
5  0.64     0.107     6.0     1.9e-09  fingers <--> fingers
6  0.52     0.098     5.3     1.4e-07  teeth <--> teeth
7  0.39     0.096     4.1     4.5e-05  breath <--> breath
8  0.77     0.144     5.4     8.3e-08  cancer <--> cancer
10 1.00     0.142     7.0     2.0e-12  smoking <--> smoking

 Iterations =  14

> print(standardized.residuals(sem.14), digits = 2)

        smoking breath  teeth fingers cancer
smoking   0.000  0.048  0.037   0.031  -0.15
breath    0.048 -0.281 -0.232  -0.201  -0.62
teeth     0.037 -0.232 -0.222  -0.185  -0.55
fingers   0.031 -0.201 -0.185  -0.164  -0.48
cancer   -0.152 -0.617 -0.554  -0.477  -0.93

> sem.15 <- sem(model.14, observed1, 100)
> summary(sem.15, digits = 2)

 Model Chisquare =  71   Df =  6 Pr(>Chisq) = 2.1e-13
 Chisquare (null model) =  188   Df =  10
 Goodness-of-fit index =  0.81
 Adjusted goodness-of-fit index =  0.53
```

```
 RMSEA index =  0.33   90% CI: (0.27, 0.4)
 Bentler-Bonnett NFI =  0.62
 Tucker-Lewis NNFI =  0.39
 Bentler CFI =  0.63
 BIC =  44

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -4.31   -3.26   -1.73   -1.61    0.00    0.72

 Parameter Estimates
   Estimate Std Error z value Pr(>|z|)
2  0.77     0.088     8.7     0.0e+00  breath <--- latent
3  0.57     0.096     6.0     2.2e-09  fingers <--- latent
4  0.67     0.093     7.2     5.1e-13  teeth <--- latent
11 0.81     0.125     6.5     9.1e-11  latent <--- smoking
5  0.64     0.104     6.2     7.6e-10  fingers <--> fingers
6  0.51     0.093     5.5     4.6e-08  teeth <--> teeth
7  0.35     0.086     4.0     5.3e-05  breath <--> breath
8  0.92     0.181     5.1     3.6e-07  cancer <--> cancer
10 1.00     0.142     7.0     2.0e-12  smoking <--> smoking

 Iterations =  13
```

```
> print(standardized.residuals(sem.15), digits = 2)
```

```
        smoking breath  teeth fingers cancer
smoking   0.000  0.095  0.087   0.075  -0.36
breath    0.095 -0.330 -0.295  -0.252  -0.88
teeth     0.087 -0.295 -0.249  -0.216  -0.76
fingers   0.075 -0.252 -0.216  -0.183  -0.65
cancer   -0.362 -0.878 -0.760  -0.651  -1.58
```

```
> sem.16 <- sem(model.14, observed, 100)
> summary(sem.16, digits = 2)
```

```
 Model Chisquare =  111   Df =  6 Pr(>Chisq) = 0
 Chisquare (null model) =  240   Df =  10
 Goodness-of-fit index =  0.76
 Adjusted goodness-of-fit index =  0.39
 RMSEA index =  0.42   90% CI: (0.35, 0.49)
 Bentler-Bonnett NFI =  0.54
 Tucker-Lewis NNFI =  0.24
 Bentler CFI =  0.55
 BIC =  83

 Normalized Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -4.63   -3.38   -1.81   -1.68    0.00    0.81

 Parameter Estimates
   Estimate Std Error z value Pr(>|z|)
2  0.73     0.082     8.9     0.0e+00  breath <--- latent
3  0.54     0.090     6.0     2.2e-09  fingers <--- latent
```

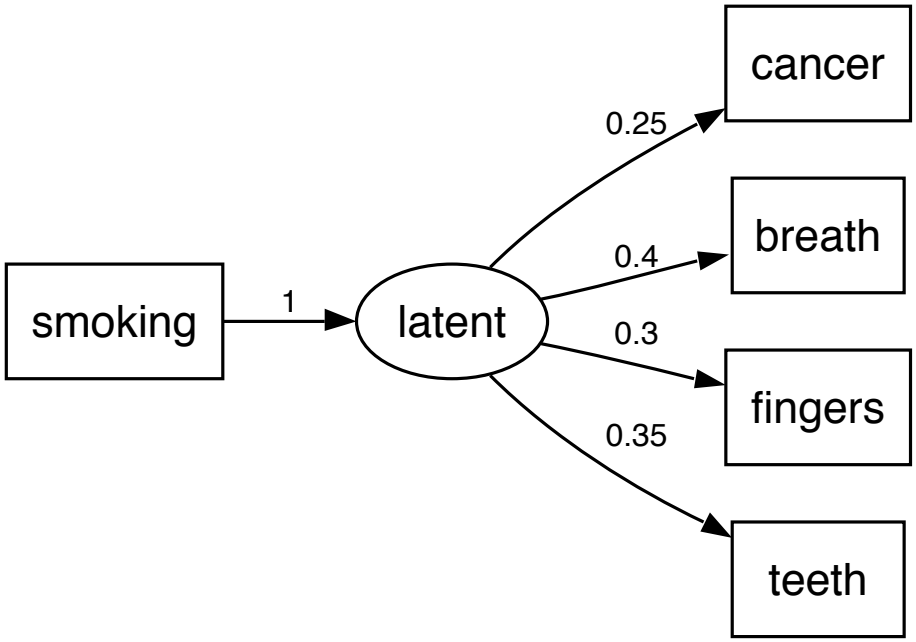

Figure 6.3: Smoking seems to affect something that leads to bad breadth, yellow teeth, and yellow hands as well as lung cancer.

```
4  0.63     0.087     7.2     6.0e-13  teeth <--- latent
11 0.95     0.133     7.2     8.5e-13  latent <--- smoking
5  0.63     0.102     6.2     5.4e-10  fingers <--> fingers
6  0.50     0.091     5.5     3.5e-08  teeth <--> teeth
7  0.32     0.082     3.9     8.5e-05  breath <--> breath
8  1.02     0.208     4.9     9.7e-07  cancer <--> cancer
10 1.00     0.142     7.0     2.0e-12  smoking <--> smoking

 Iterations =  13
```

6.3 Measures of fit

As has been seen in the previous sections, the use of fit statisics does not guarantee meaningful models. If we do not specify the model correctly, either because we do not include the correct variables or because we fail to use the appropriate measurement model, we will lead to incorrect conclusions.

Even if we have a very good fit, we are unable to determine causal structure from the model, even if we bother to add time into the model.

6.3.1 χ^2

As we saw in the previous chapter, χ^2 is very sensitive to many sources of error in our model specification. χ^2 is sensitive to failures of our distributional assumptions (continuous, multivariate normal) as well as to our failures to correctly specify the structure.

6.3.2 GFI, NFI, ...

6.3.3 RMSEA

6.4 What does it mean to fit a model

What should we do when the model does not fit? This is a recurring controversy, discussed, for instance in the March, 2007 issue of Personality and Individual Differences. It is also a continuing source of debate on the SEM-net list serve. There are those who treat fit statistics (particularly χ^2) as the definitive test and evidence for model adequacy. There are others who do not take such an all or none approach, and are concerned with comparisons of models to alternative models.

6.5 References