Supplementary materials for the article

Koukounari, A., Copas, A., Pickles, A.

A latent variable modelling approach for the pooled analysis of individual participant data on the association between depression and Chlamydia infection in adolescence and young adulthood in the UK

Appendix Part A: Additional results supporting analysis appearing in the main article and path diagrams

A1: LCA for Chlamydia infection

Table A1: Bayesian Information criteria and number of parameters for 4 different LCA models

Model	*R*	*BIC*
1◊	15	31443.459
2°	19	31429.606
3•	23	38738.948
4□	21	34530.312

◊*Model 1 considered prefixed item response probabilities for the Chlamydia test and allowed for estimation of item response probabilities for the following items: number of heterosexual/ homosexual partners without a condom last year, any overlap between partners and number of partners in lifetime. Model 1 assumed measurement invariance for all items.*

°*Model 2 considered prefixed item response probabilities for the Chlamydia test and allowed for estimation of item response probabilities for the same three items as Model 1. It also allowed for the item response probabilities to vary by gender for the following 2 items: number of heterosexual/ homosexual partners without a condom last year and any overlap between partners*

• *Model 3 considered prefixed item response probabilities for the Chlamydia test and allowed for estimation of item response probabilities for the same three items as Models 1 and 2 as well as for any STI symptoms, and attendance of an STI clinic. It also allowed for the item response probabilities to vary by gender for the following 2 items: number of heterosexual/ homosexual partners without a condom last year and any overlap between partners.*

□ *Model 4 is as Model 3 but excluding estimation of item response probabilities for item: attendance of an STI clinic.*

A2: Logit values – generated during LCA and used at the last stage of analysis

During the LCA estimation and using the latent class posterior distribution, the most likely class variable *N* is created as a nominal variable. This *N* variable as derived during Step 1 is later on specified as a nominal indicator of the latent class variable *C* (i.e. during Step 4) with uncertainty rates prefixed at the probabilities obtained during the LCA estimation. More specifically, during Step 4 we will regress the *C* on the growth parameters of depression–as described in the manuscript and at logits as indicated below in Table A2. In this way the measurement error in the most likely class variable *N* is taken into account during Step 4. This table contains results from the Mplus output if B1 code below is run.

Table A2. Mplus output, from LCA, which will be used later

Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column) by Latent Class (Row)		
	1	*2*
1	5.974	0.000
2	-0.676	0.000

A3: Sensitivity Analyses

A3.1 Evaluation of the quality of obtaining precalibration parameters from the selected IRT model described in Section 3.2 of the actual paper

To evaluate the quality of obtaining precalibration parameters from the selected IRT model as described in Section 3.2, we cross -validated those with a new randomly selected calibration sample from the 2 studies -the age distribution of which follows in the Table below

Table A.3 Number of participants in a new randomly selected calibration sample per study and age

	AGE												
STUDY	***10***	***12***	***13***	***16***	***17***	***18***	***19***	***20***	***21***	***22***	***23***	***24***	***Total***
ALSPAC, N (%)	1092 (100)	891 (100)	688 (100)	506 (56.98)	386 (47.19)	236 (34.15)	0	0	0	0	0	0	**3799**
NATSAL-3, N (%)	0	0	0	382 (43.02)	432 (52.81)	455 (65.85)	390 (100)	391 (100)	370 (100)	404 (100)	410 (100)	373 (100)	**3607**
Total	**1092**	**891**	**688**	**888**	**818**	**691**	**390**	**391**	**370**	**404**	**410**	**373**	**7406**

From this new calibration sample and the same model as described in section 3.2, results follow in the table below. Such results should be compared with results displayed in Table 4 of the actual paper. There are not big differences in the displayed coefficients from the same IRT model fitted to the different pre-calibration samples. Such coefficients would then be fixed for common item parameters in the longitudinal IRT equating models linked to second order Latent Growth Curve models as described in Section 3.3 of the actual paper. Thus, the harmonized depression scores deriving from these longitudinal IRT equating models, would not change dramatically either, which strengthens the validity of our findings from these models.

Table A.4 IRT model pre-calibration parameters of equating depression scores from both studies -n=7406

Item Description	***Factor Loading (SE)***	***Standardized Factor Loading (SE)***	***Threshold (SE)*** ***1st*** ***2nd***
1. *Young person has not enjoyed anything in the last two weeks*	1.000 (0.000)	0.765 (0.010)	7.924 (0.569) 10.711 (0.587)
2. *Young person has felt unhappy/miserable in the last two weeks* *Study covariate effect (DIF)*	1.090 (0.049)	0.803 (0.011)	2.742 (0.597) 6.524 (0.613)

			-2.660 (0.113)
3. *Young person has felt so tired they sat around and did nothing in the last two weeks*	0.498 (0.027)	0.509 (0.018)	2.793 (0.299) 5.444 (0.314)
4. *Young person has felt very restless in the last two weeks*	0.445 (0.026)	0.467 (0.018)	2.434 (0.270) 5.062 (0.284)
5. *Young person felt they were no good anymore in the last two weeks*	1.808 (0.094)	0.907 (0.007)	13.368 (1.107) 16.708 (1.156)
6. *Young person has cried a lot in the last two weeks*	0.963 (0.049)	0.753 (0.014)	7.389 (0.579) 9.662 (0.599)
7. *Young person has found it hard to think properly/concentrate in the last two weeks*	0.683 (0.034)	0.630 (0.015)	3.493 (0.394) 6.626 (0.411)
8. *Young person has hated themselves in the last two weeks*	1.732 (0.093)	0.899 (0.008)	13.279 (1.074) 16.226 (1.119)
9. *Young person has felt they were a bad person in the last two weeks*	1.017 (0.053)	0.770 (0.014)	8.118 (0.623) 10.849 (0.654)
10. *Young person has felt lonely in the last two weeks*	1.198 (0.057)	0.818 (0.010)	7.803 (0.697) 10.997 (0.724)
11. *Young person has felt nobody really loved them in the last two weeks*	1.421 (0.074)	0.860 (0.010)	11.126 (0.874) 13.886 (0.910)
12. *Young person thought they could never be as good as other kids in the last two weeks*	1.155 (0.056)	0.808 (0.011)	8.176 (0.678) 11.169 (0.705)
13. *Young person has felt they did everything wrong in the last two weeks*	1.416 (0.072)	0.860 (0.009)	10.605 (0.856) 13.935 (0.898)
14. *Young person has been having fun in the last two weeks*	-0.535 (0.034)	-0.536 (0.022)	-8.262 (0.405) -4.617 (0.350)
15. *Young person has felt happy in the last two weeks*	-0.550 (0.033)	-0.547 (0.020)	-7.137 (0.370) -4.332 (0.346)
16. *Young person has enjoyed doing lots of things in the last two weeks*	-0.495 (0.031)	-0.507 (0.021)	-6.670 (0.342)

			-3.901 (0.318)

Factor mean/Covariate effect	*Estimate (SE)*	*Standardized Estimate (SE)*
Study membership (reference category: ALSPAC)	5.940 (1.107)	1.378 (0.255)
Age	0.118 (0.078)	0.238 (0.157)
Age^2	0.012 (0.005)	0.830 (0.322)
AgexStudy	-0.304 (0.065)	-1.981 (0.419)
Gender (reference category: males)	-0.037 (0.056)	-0.009 (0.013)

A3.2 Validity for an association between trajectories before age 16 and Chlamydia infection

To test the validity for an association between trajectories before age 16 and Chlamydia infection, from the harmonization effort among the two studies, we explored this relationship with ALSPAC data only. More precisely, we have fitted a latent growth model (LGM) with intercept being at 10 years old and slope representing changes in MFQ scores between ages 10-16. We have subsequently saved these growth factor scores for each person (i.e. the intercept and slope) predicted from this model. We then fit a logistic regression model where the dependent variable is just the binary outcome of the original ALSPAC urine test (with 0: uninfected and 1: infected) and covariates are the growth factor scores as described above. We have also tried to fit an LGM where the intercept was age 16 and slope was representing the changes between 10-16 but the result was that the model covariance matrix was not positive definite and factor scores could not be computed from this model. Despite this computational challenge the table below shows still a positive association between the changes of MFQ scores during 10-16 years and an increased likelihood to being infected with Chlamydia at the approximate age of 17 years old (OR=2.054).

Table A.5 Logits and Odds Ratios for the Effects of Growth Factors of depression MFQ scores (i.e. Intercept at 10 years old and Slope representing changes between ages 10 – 16 years old derived from a Latent Growth Model) on Chlamydia infection as determined by the urine test in the ALSPAC data (n=2776)

Effect	*Logit (SE)*	*p-value*	*OR*
Intercept *(at 10 years old)*	-2.353 (0.169)	<0.001	0.095
Slope *(changes during ages 10-16)*	0.720 (0.063)	<0.001	2.054

OR: Odds Ratio

Appendix Part B-Mplus scripts

B1: Step 1- Mplus script of LCA for Chlamydia infection

Data: File is ALLCHL_OCTN.DAT;

Variable: Names are

TEST STISYM NOCOND AGEN SEX ID OVRPN NNUMPT STICLN RACE CONDN NNEWP NNUMPLY STUDY EARSEXN;

IDVARIABLE IS ID;

USEVARIABLES

TEST NOCOND OVRPN NNUMPT SEX;

CATEGORICAL = TEST NOCOND OVRPN NNUMPT;

! TEST: is the biological Chlamydia test; NOCOND is the number of heterosexual/ homosexual partners without a condom last year, OVRPN is any overlap between partners and NNUMPT is the number of partners in lifetime

MISSING ARE ALL (-9); CLASSES = c (2);

ANALYSIS: TYPE = MIXTURE MISSING; STARTS=100 10; PROCESS=2(STARTS); COVERAGE=0.05;

MODEL:

%OVERALL%

!Partial measurement invariance for males and females and the questions corresponding to number of partners without a condom in the last year and any overlap between partners

NOCOND OVRPN ON SEX;

%C#1%

! Below we fix the first threshold of the biological Chlamydia test to be 15. This will yield a probability of 1 for observed negatives and thus through this way we fix the specificity of this test to be 1 or 100 %

[TEST$1@15];

!Partial measurement invariance for males and females and the questions corresponding to number of partners without a condom in the last year and any overlap between partners

NOCOND OVRPN ON SEX;

%C#2%

! Below we fix the first threshold of the biological Chlamydia test to be -15. This will yield a probability of 0 for observed negatives and its complementary probability of 1 for observed positives; through this approach we fix the sensitivity of this test to be 1 or 100 %

[TEST$1@-15];

! Partial measurement invariance for males and females and the questions corresponding to number of partners without a condom in the last year and any overlap between partners

NOCOND OVRPN ON SEX;

SAVEDATA: FILE IS ALLCHLPP_OCTN.TXT;

!Saving class probabilities and most likely latent class membership information (this is the nominal variable *N* in Asparouhov and Muthén 2014 paper and as it appears in the corresponding Mplus code of Appendix B4, later on)

SAVE=CPROB;

B2: Mplus script of IRT graded response model

```
DATA: FILE IS DEPMNLFA2.DAT;

VARIABLE: NAMES ARE


STUD AGE SEX NOTEN MISUN TIRED RESTLS NOGOOD

CRIED CONCEN HATED BAD LONEL NOLOV NVRGOD

WRONG FUN HAPPY DOING AGE2 AGESTUD;


CATEGORICAL ARE NOTEN MISUN TIRED RESTLS NOGOOD

CRIED CONCEN HATED BAD LONEL NOLOV NVRGOD

WRONG FUN HAPPY DOING;

! NOTEN: Young person has not enjoyed anything in the last two weeks

MISUN: Young person has felt unhappy/miserable in the last two weeks

TIRED:  Young person has felt so tired they sat around and did nothing in the last two weeks

RESTLS: Young person has felt very restless in the last two weeks

NOGOOD: Young person felt they were no good anymore in the last two weeks

CRIED: Young person has cried a lot in the last two weeks

CONCEN: Young person has found it hard to think properly/concentrate in the last two weeks

HATED: Young person has hated themselves in the last two weeks

BAD: Young person has felt they were a bad person in the last two weeks

LONEL: Young person has felt lonely in the last two weeks

NOLOV: Young person has felt nobody really loved them in the last two weeks

NVRGOD: Young person thought they could never be as good as other kids in the last two weeks

WRONG: Young person has felt they did everything wrong in the last two weeks

FUN: Young person has been having fun in the last two weeks

HAPPY: Young person has felt happy in the last two weeks

DOING: Young person has enjoyed doing lots of things in the last two weeks

MISSING ARE ALL (-9);

ANALYSIS:

ESTIMATOR=ml;
```

```
ALGORITHM=integration;
MODEL: DEP BY NOTEN MISUN TIRED RESTLS NOGOOD
CRIED CONCEN HATED BAD LONEL NOLOV NVRGOD
WRONG FUN HAPPY DOING;
! DEP is the factor DEPRESSION
DEP ON AGE AGE2 STUD AGESTUD SEX;
! AGE2 IS THE SQUARED TERM FOR AGE
!AGESTUD IS THE INTERACTION TERM BETWEEN AGE AND STUDY
[DEP@0];
NOTEN ON STUD@0;
MISUN ON STUD;
NOTEN-DOING ON AGE @0;
NOTEN-DOING ON AGE2@0;
NOTEN-DOING ON AGESTUD@;
NOTEN-DOING ON SEX @0;
!NOTENJ16 ON STUD;
OUTPUT: standardized TECH2;
SAVEDATA: FILE IS STEP2FSJUL2.DAT;SAVE = FSCORES;
```

B3: Mplus script of IRT growth model of longitudinal and cross-sectional depression scores

```
TITLE: IRT FOR ALL

 DATA: FILE IS

 IRTALLCOVS.dat;

 VARIABLE: NAMES ARE ID MISUN18 NOTENJ18 TIRED18 RESTLS18 NOGOOD18 CRIED18

             CONCEN18 HATED18 BAD18 LONEL18 NOLOV18 NVRGOD18 WRONG18

             MISUN16 FUN16 NOTENJ16 TIRED16 RESTLS16 NOGOOD16 CRIED16

             HAPPY16 CONCEN16 HATED16 DOING16 BAD16 LONEL16 NOLOV16

             NVRGOD16 WRONG16 MISUN13 FUN13 NOTENJ13 TIRED13 RESTLS13

             NOGOOD13 CRIED13 HAPPY13 CONCEN13 HATED13 DOING13 BAD13

             LONEL13 NOLOV13 NVRGOD13 WRONG13 MISUN12 FUN12 NOTENJ12

             TIRED12 RESTLS12 NOGOOD12 CRIED12 HAPPY12 CONCEN12 HATED12

             DOING12 BAD12 LONEL12 NOLOV12 NVRGOD12 WRONG12 SEX

             MISUN17 FUN17 NOTENJ17 TIRED17 RESTLS17 NOGOOD17 CRIED17

             HAPPY17 CONCEN17 HATED17 DOING17 BAD17 LONEL17 NOLOV17

             NVRGOD17 WRONG17 MISUN10 FUN10 NOTENJ10 TIRED10 RESTLS10

             NOGOOD10 CRIED10 HAPPY10 CONCEN10 HATED10 DOING10 BAD10

             LONEL10 NOLOV10 NVRGOD10 WRONG10 STUD NOTENJ19 NOTENJ20

             NOTENJ21 NOTENJ22 NOTENJ23 NOTENJ24 MISUN19 MISUN20

             MISUN21 MISUN22 MISUN23 MISUN24;


 !USEVARIABLES    ;


 CATEGORICAL ARE MISUN18 NOTENJ18 TIRED18 RESTLS18 NOGOOD18

         CRIED18 CONCEN18 HATED18 BAD18 LONEL18 NOLOV18

         NVRGOD18 WRONG18

         MISUN16 FUN16 NOTENJ16 TIRED16

         RESTLS16 NOGOOD16 CRIED16 HAPPY16 CONCEN16
```

```
            HATED16 DOING16 BAD16 LONEL16 NOLOV16

            NVRGOD16 WRONG16

            MISUN13 FUN13 NOTENJ13 TIRED13

            RESTLS13 NOGOOD13 CRIED13 HAPPY13 CONCEN13 HATED13

            DOING13 BAD13 LONEL13 NOLOV13 NVRGOD13 WRONG13

            MISUN12 FUN12 NOTENJ12 TIRED12 RESTLS12 NOGOOD12 CRIED12 HAPPY12

            CONCEN12 HATED12 DOING12 BAD12 LONEL12 NOLOV12 NVRGOD12 WRONG12

            MISUN17 FUN17 NOTENJ17 TIRED17 RESTLS17 NOGOOD17

            CRIED17 HAPPY17 CONCEN17 HATED17 DOING17 BAD17 LONEL17

            NOLOV17 NVRGOD17 WRONG17 MISUN10 FUN10 NOTENJ10 TIRED10 RESTLS10

            NOGOOD10 CRIED10 HAPPY10 CONCEN10 HATED10 DOING10 BAD10 LONEL10

            NOLOV10 NVRGOD10 WRONG10 NOTENJ19 NOTENJ20 NOTENJ21 NOTENJ22 NOTENJ23

            NOTENJ24 MISUN19 MISUN20 MISUN21 MISUN22 MISUN23 MISUN24;

! the variable names are the same as in B2, the numbers at the end represent the ages for which these measurements were taken

 IDVARIABLE IS ID;

 MISSING ARE ALL (-9);


 ANALYSIS: ESTIMATOR IS MLF;

 process=8;

 LINK IS LOGIT;

 INTEGRATION = MONTECARLO;

 MITERATIONS = 700;

!The values at which loadings and thresholds are fixed below can be found at Table 3 in the actual paper

 MODEL:

 F10 BY NOTENJ10@1 MISUN10@1.114 FUN10@-0.516 TIRED10@0.535

      RESTLS10@0.459 NOGOOD10@1.818 CRIED10@0.944 HAPPY10@-0.566
```

```
CONCEN10@0.635 HATED10@1.679 DOING10@-0.525 BAD10@1.026
LONEL10@1.277 NOLOV10@1.407 NVRGOD10@1.145 WRONG10@1.392;

F12 BY NOTENJ12@1 MISUN12@1.114 FUN12@-0.516 TIRED12@0.535
RESTLS12@0.459 NOGOOD12@1.818 CRIED12@0.944 HAPPY12@-0.566
CONCEN12@0.635 HATED12@1.679 DOING12@-0.525 BAD12@1.026
LONEL12@1.277 NOLOV12@1.407 NVRGOD12@1.145 WRONG12@1.392;

F13 BY NOTENJ13@1 MISUN13@1.114 FUN13@-0.516 TIRED13@0.535
RESTLS13@0.459 NOGOOD13@1.818 CRIED13@0.944 HAPPY13@-0.566
CONCEN13@0.635 HATED13@1.679 DOING13@-0.525 BAD13@1.026
LONEL13@1.277 NOLOV13@1.407 NVRGOD13@1.145 WRONG13@1.392;

F16 BY NOTENJ16@1 MISUN16@1.114 FUN16@-0.516 TIRED16@0.535 RESTLS16@0.459
NOGOOD16@1.818 CRIED16@0.944 HAPPY16@-0.566 CONCEN16@0.635
HATED16@1.679 DOING16@-0.525 BAD16@1.026 LONEL16@1.277
NOLOV16@1.407 NVRGOD16@1.145 WRONG16@1.392;

F17 BY NOTENJ17@1 MISUN17@1.114 FUN17@-0.516 TIRED17@0.535
RESTLS17@0.459 NOGOOD17@1.818 CRIED17@0.944
HAPPY17@-0.566 CONCEN17@0.635 HATED17@1.679
DOING17@-0.525 BAD17@1.026 LONEL17@1.277
NOLOV17@1.407 NVRGOD17@1.145 WRONG17@1.392;

F18 BY NOTENJ18@1 MISUN18@1.114 TIRED18@0.535 RESTLS18@0.459
NOGOOD18@1.818 CRIED18@0.944 CONCEN18@0.635 HATED18@1.679
BAD18@1.026 LONEL18@1.277 NOLOV18@1.407 NVRGOD18@1.145
```

```
      WRONG18@1.392;

 F19 BY NOTENJ19@1
      MISUN19@1.114;

 F20 BY NOTENJ20@1
      MISUN20@1.114;

 F21 BY NOTENJ21@1
      MISUN21@1.114;

 F22 BY NOTENJ22@1
      MISUN22@1.114;

 F23 BY NOTENJ23@1
      MISUN23@1.114;

 F24 BY NOTENJ24@1
      MISUN24@1.114;
! F stand for the factors at the different ages
!Spacings of time, intercept at 16, 1st slope during ages 10-16, 2nd slope during ages 16-24
 int slp1|F10@-1 F12@-1.2 F13@-1.3 F16@0
          F17@0 F18@0 F19@0 F20@0
          F21@0 F22@0 F23@0 F24@0;
 int slp2|F10@0 F12@0 F13@0 F16@1 F17@2
        F18@3 F19@4 F20@5 F21@6 F22@7
        F23@8 F24@9;
```

```
[int*0];

[slp1*0.25];

[slp2*0.1];


int*.1 (vint);

slp1*.1 (vslp1);

slp2*.1 (vslp2);


F10*4.223 (vth)

F12*4.223 (vth)

F13*4.223 (vth)

F16*4.223 (vth)

F17-F24*4.223 (vth);


int WITH slp1;

int WITH slp2;

slp1 WITH slp2;

!Covariate effects on growth factors: intercept, slope1 and slope2

int on SEX;

int on STUD;

slp1 on SEX;

!slp1 on STUD;

slp2 on SEX;

slp2 on STUD;
```

[MISUN10$1@2.179]; [MISUN12$1@2.179]; [MISUN13$1@2.179];[MISUN16$1@2.179]; [MISUN17$1@2.179]; [MISUN18$1@2.179]; [MISUN19$1*](17); [MISUN20$1*](17); [MISUN21$1*](17); [MISUN22$1*](17);[MISUN23$1*](17);[MISUN24$1*](17);

[MISUN10$2@5.867];[MISUN12$2@5.867];[MISUN13$2@5.867]; [MISUN16$2@5.867]; [MISUN17$2@5.867];[MISUN18$2@5.867];[MISUN19$2*](18);[MISUN20$2*](18);

[MISUN21$2*](18);[MISUN22$2*](18);[MISUN23$2*](18);[MISUN24$2*](18);

[NOTENJ10$1@7.434];[NOTENJ12$1@7.434]; [NOTENJ13$1@7.434];[NOTENJ16$1@7.434]; [NOTENJ17$1@7.434];[NOTENJ18$1@7.434];[NOTENJ19$1*](19);[NOTENJ20$1*](19);

[NOTENJ21$1*](19);[NOTENJ22$1*](19);[NOTENJ23$1*](19);[NOTENJ24$1*](19);

[NOTENJ10$2@10.136]; [NOTENJ12$2@10.136];[NOTENJ13$2@10.136];[NOTENJ16$2@10.136];

[NOTENJ17$2@10.136];[NOTENJ18$2@10.136];[NOTENJ19$2*](20);[NOTENJ20$2*](20);

[NOTENJ21$2*](20);[NOTENJ22$2*](20);[NOTENJ23$2*](20);[NOTENJ24$2*](20);

[TIRED10$1@2.606];[TIRED12$1@2.606];[TIRED13$1@2.606];[TIRED16$1@2.606];

[TIRED17$1@2.606];[TIRED18$1@2.606];

[TIRED10$2@5.241];[TIRED12$2@5.241];[TIRED13$2@5.241];[TIRED16$2@5.241];

[TIRED17$2@5.241];[TIRED18$2@5.241];

[RESTLS10$1@2.306];[RESTLS12$1@2.306];[RESTLS13$1@2.306];[RESTLS16$1@2.306];

[RESTLS17$1@2.306];[RESTLS18$1@2.306];

[RESTLS10$2@4.977];[RESTLS12$2@4.977];[RESTLS13$2@4.977];[RESTLS16$2@4.977];

[RESTLS17$2@4.977];[RESTLS18$2@4.977];

[NOGOOD10$1@12.447];[NOGOOD12$1@12.447];[NOGOOD13$1@12.447];

[NOGOOD16$1@12.447]; [NOGOOD17$1@12.447];[NOGOOD18$1@12.447];

[NOGOOD10$2@15.948];[NOGOOD12$2@15.948];[NOGOOD13$2@15.948];

[NOGOOD16$2@15.948];[NOGOOD17$2@15.948];[NOGOOD18$2@15.948];

[CRIED10$1@6.741];[CRIED12$1@6.741];[CRIED13$1@6.741];[CRIED16$1@6.741];

[CRIED17$1@6.741];[CRIED18$1@6.741];

[CRIED10$2@8.898];
[CRIED12$2@8.898];[CRIED13$2@8.898];[CRIED16$2@8.898];[CRIED17$2@8.898];

[CRIED18$2@8.898];

[CONCEN10$1@2.935];[CONCEN12$1@2.935];[CONCEN13$1@2.935];[CONCEN16$1@2.935];
[CONCEN17$1@2.935];[CONCEN18$1@2.935];

[CONCEN10$2@5.968];[CONCEN12$2@5.968];[CONCEN13$2@5.968];[CONCEN16$2@5.968];

[CONCEN17$2@5.968];[CONCEN18$2@5.968];

[HATED10$1@12.190];[HATED12$1@12.190];[HATED13$1@12.190];[HATED16$1@12.190];

[HATED17$1@12.190];[HATED18$1@12.190];

[HATED10$2@14.936];[HATED12$2@14.936];[HATED13$2@14.936];[HATED16$2@14.936];

[HATED17$2@14.936];[HATED18$2@14.936];

[BAD10$1@7.619];[BAD12$1@7.619];[BAD13$1@7.619];[BAD16$1@7.619];[BAD17$1@7.619];

[BAD18$1@7.619];

[BAD10$2@10.281];[BAD12$2@10.281];[BAD13$2@10.281];[BAD16$2@10.281];

[BAD17$2@10.281];[BAD18$2@10.281];

[LONEL10$1@7.654];[LONEL12$1@7.654];[LONEL13$1@7.654];[LONEL16$1@7.654];
[LONEL17$1@7.654];[LONEL18$1@7.654];

[LONEL10$2@10.932];[LONEL12$2@10.932];[LONEL13$2@10.932];[LONEL16$2@10.932];

[LONEL17$2@10.932];[LONEL18$2@10.932];

[NOLOV10$1@10.312];[NOLOV12$1@10.312];[NOLOV13$1@10.312];[NOLOV16$1@10.312]

;[NOLOV17$1@10.312];[NOLOV18$1@10.312];

[NOLOV10$2@12.972];[NOLOV12$2@12.972];[NOLOV13$2@12.972];[NOLOV16$2@12.972];

[NOLOV17$2@12.972];[NOLOV18$2@12.972];

[NVRGOD10$1@7.644];[NVRGOD12$1@7.644];[NVRGOD13$1@7.644];[NVRGOD16$1@7.644];

[NVRGOD17$1@7.644];[NVRGOD18$1@7.644];

[NVRGOD10$2@10.338];[NVRGOD12$2@10.338];[NVRGOD13$2@10.338];

[NVRGOD16$2@10.338];[NVRGOD17$2@10.338];[NVRGOD18$2@10.338];

[WRONG10$1@9.800];[WRONG12$1@9.800];[WRONG13$1@9.800];[WRONG16$1@9.800];

[WRONG17$1@9.800];[WRONG18$1@9.800];

[WRONG10$2@12.961];[WRONG12$2@12.961];[WRONG13$2@12.961];[WRONG16$2@12.961];

[WRONG17$2@12.961];[WRONG18$2@12.961];

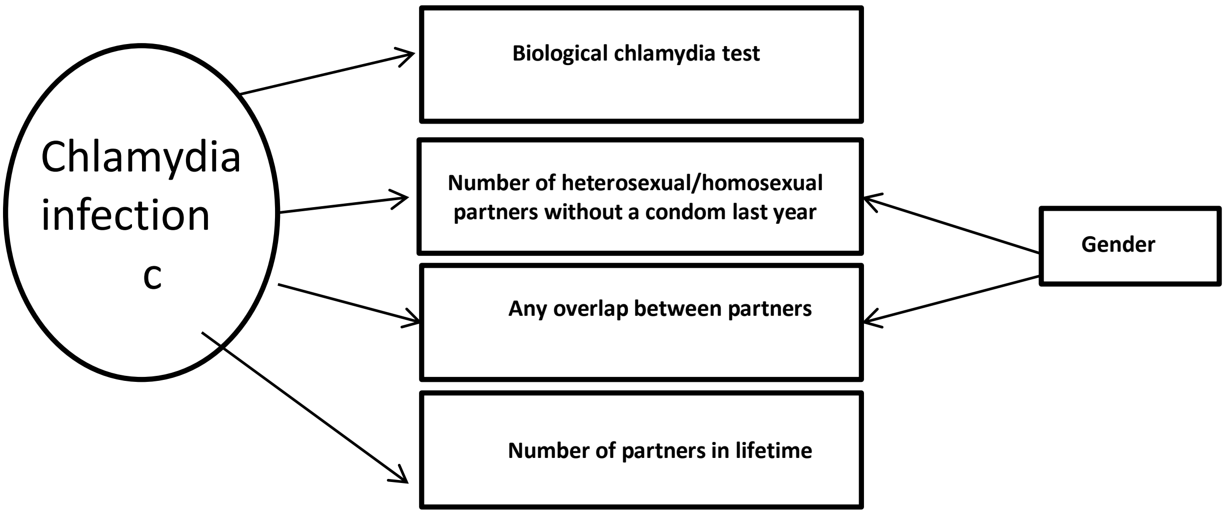
```
[FUN10$1@-7.386];[FUN12$1@-7.386];[FUN13$1@-7.386];[FUN16$1@-7.386];[FUN17$1@-7.386];

[FUN10$2@-4.141];[FUN12$2@-4.141];[FUN13$2@-4.141];[FUN16$2@-4.141];[FUN17$2@-4.141];

[HAPPY10$1@-6.738];[HAPPY12$1@-6.738];[HAPPY13$1@-6.738]; [HAPPY16$1@-6.738];

[HAPPY17$1@-6.738];

[HAPPY10$2@-4.064];[HAPPY12$2@-4.064];[HAPPY13$2@-4.064];

[HAPPY16$2@-4.064];[HAPPY17$2@-4.064];

[DOING10$1@-6.253];[DOING12$1@-6.253];[DOING13$1@-6.253];[DOING16$1@-6.253];

[DOING17$1@-6.253];

[DOING10$2@-3.706];[DOING12$2@-3.706];[DOING13$2@-3.706];[DOING16$2@-3.706];

[DOING17$2@-3.706];


MODEL CONSTRAINT:

vth > 0; ! residual variance of theta

vslp1 > 0; ! variance of the slope

vslp2 > 0;

vint > 0; ! variance of the intercept

OUTPUT: TECH1, TECH4, TECH10;

NOCHISQUARE;

SAVEDATA: FILE IS

FS_IRTALL16_WITHSTUDJUL.DAT;

SAVE = FSCORES;
```

B4: Mplus script of LCA with covariates; associations between Chlamydia infection and depression trajectories

```
DATA: FILE IS ALLDEPCHLAUG.DAT;

VARIABLE: NAMES = INT SLP1 SLP2 SEX STUD ID N;

USEVARIABLES= INT SLP1 SLP2 STUD N;

CLASSES=C(2);

NOMINAL=N;

MISSING ARE ALL(-9);

ANALYSIS: TYPE  = MIXTURE MISSING; STARTS=0;

MODEL: %OVERALL%

      C ON INT SLP1 SLP2 STUD;

!INT SLP1 SLP2 have been saved in the previous step-these are the growth parameters from the piecewise linear model

!N is a variable indicating at which latent class each person had been assigned during the LCA fitting

!The values below were included in Table A2 in this Appendix

      %c#1%

      [n#1@5.974];

      %c#2%

      [n#1@-0.676];
```

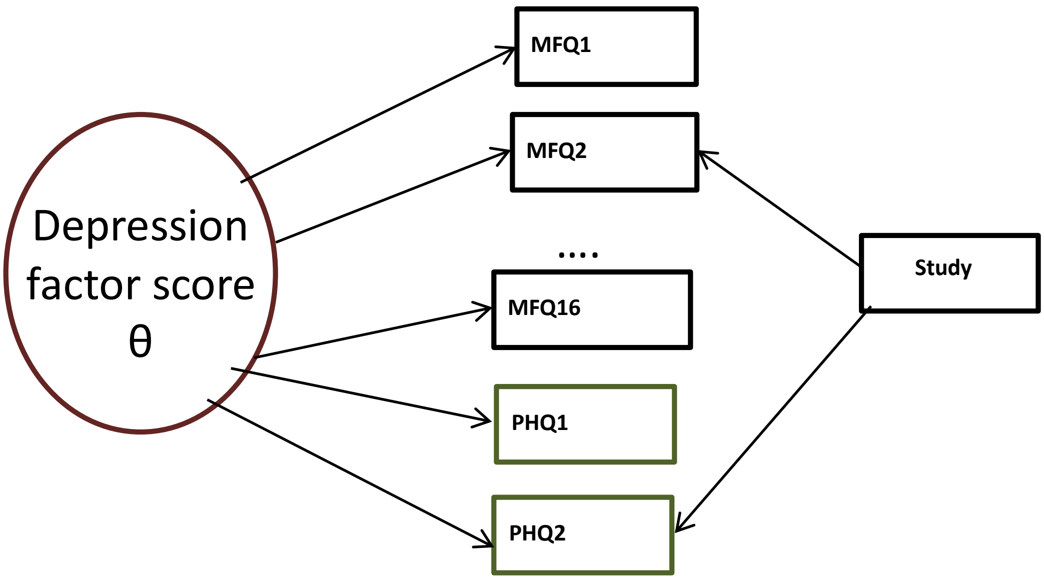
Appendix Part C-Path diagrams for fitted models

C1. Latent Class Analysis for Chlamydia Infection



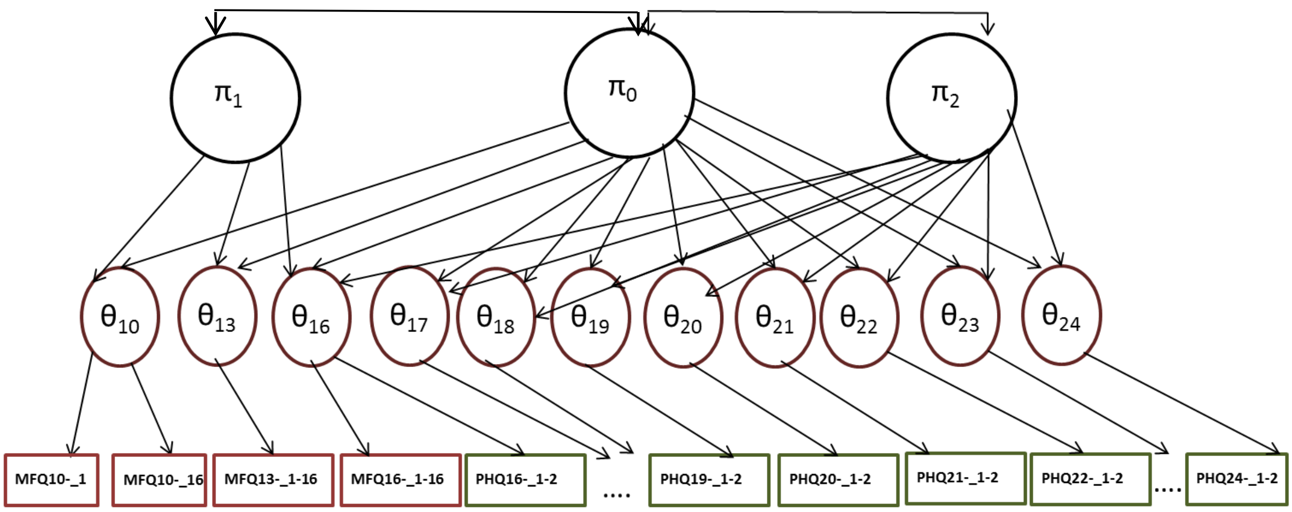
In this Figure and all figures representing fitted latent variable models, squares indicate observed variables while circles indicate latent variables. Gender arrows indicate partial measurement invariance for the questions included in this model. This model is Model 2 in Table A1 in this Appendix. C stands for a categorical latent variable as described in the paper.

C2. IRT model to obtain pre-calibration parameters for equating depression scores from both studies



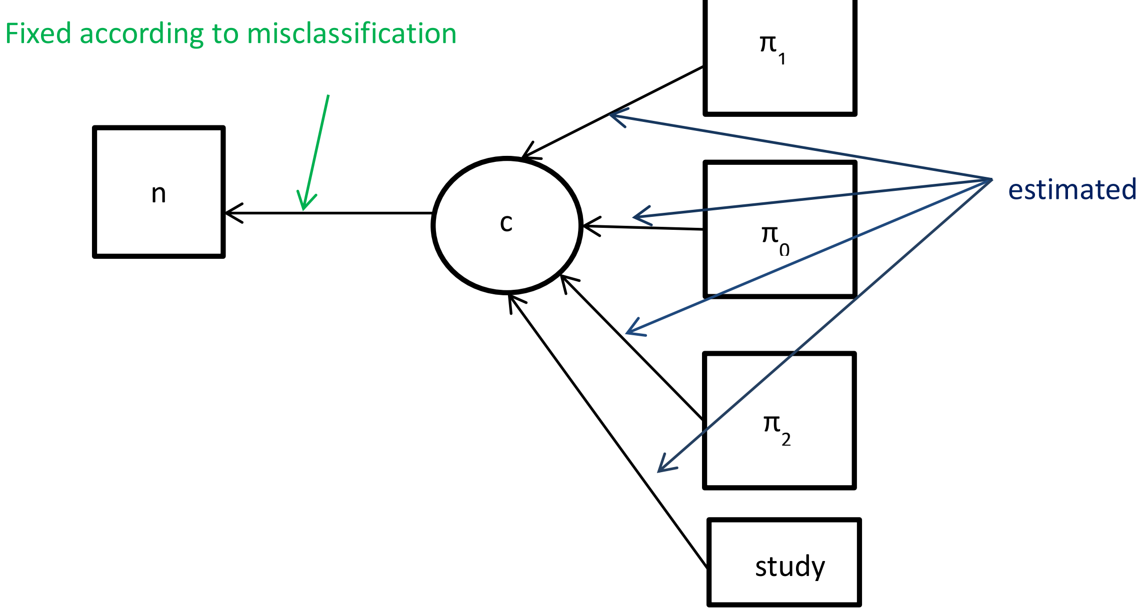
Study arrows indicate partial measurement invariance for the questions included in this model. MFQ2 and PHQ2 are the harmonized items as described in the paper. The depression factor scores here are estimated though 16 items from the Mood and Feelings Questionnaire from ALSPAC and 2 items from Natsal-3 fitted to a calibration sample as described in the paper. This model (and its obtained pre-calibration parameters used in Step 4 can be found in Table 3 in the paper. θ stands for the IRT factor score although we do not use this further, just the obtained pre-calibration parameters as also shown in code B3 here in the Appendix.

C3. Latent Growth Curve second-order IRT model of depression scores



In this second order IRT growth model a piecewise linear model is shown at the second level while IRT θ scores are estimated at the first level. π_0 denotes the intercept of the IRT θ scores defined at 16 years old; π_1 denotes the slope of the IRT θ scores covering 10-16 years old; π_2 denotes the slope of the IRT scores covering 16-24 years old. MFQ indicate the observed measures (16 items) at ages 10-18 from ALSPAC study and the Mood and Feelings Questionnaire. PHQ indicate the observed measures (2 items) at ages 16-24 from NATSAL-3 study and the Patient Health Questionnaire

C4. LCA with covariates; associations between Chlamydia infection and depression trajectories



π_0, π_1 and π_2 denote growth parameters of the piecewise linear model as described before. The *n* variable was derived during LCA estimation and was specified as a nominal indicator of the latent class variable *c* during the last stage of analysis with uncertainty rates prefixed at the probabilities obtained also during the LCA estimation.