TwinLife Technical Report Series

No. 07, March 2021

Documentation *TwinLife* Data:

Filtering inconsistencies in discrimination items

v1.1.0

by Alexandra Starr, Amelie Nikstat, Sophia Mattheus, Merit Kaempfert, Anke Hufer

alexandra.starr@uni-bielefeld.de

https://www.twin-life.de/twinlife-series

Alexandra Starr, Amelie Nikstat, Sophia Mattheus, Merit Kaempfert, Anke Hufer
Documentation *TwinLife* Data: Filtering inconsistencies in discrimination items – v.1.1.0

TwinLife Technical Report Series No. 07
Project TwinLife "Genetic and social causes of life chances"
Bielefeld, March 2021

TwinLife Technical Report Series
General Editors: Martin Diewald, Rainer Riemann and Frank M. Spinath
ISSN 2512-403X

This publication has been funded by the German Research Foundation (DFG).

TwinLife Technical Reports are refereed scholarly papers. Submissions are reviewed by the general editors before a final decision on publication is made.

The Technical Report Series is a forum for presenting technical works (e.g., data documentation, field reports etc.) in progress. Readers should communicate comments on the manuscript directly to the author(s).

The papers can be downloaded from the project website:
https://www.twin-life.de/twinlife-series

TwinLife "Genetic and social causes of life chances"
University of Bielefeld
Faculty of Sociology
PO Box 100131
D-33501 Bielefeld
Germany

Phone: +49 (0)521 106-4309
Email: martin.diewald@uni-bielefeld.de
Web: https://www.twin-life.de

UNIVERSITÄT BIELEFELD

UNIVERSITÄT DES SAARLANDES

Table of Contents

Introduction

This technical report deals with filtering inconsistencies in items regarding experienced discrimination in the first survey wave of the TwinLife study. These inconsistencies occurred at the transition in between items in the paper-pencil mode of data collection and led to participants skipping items they should have answered or answering items they should have not. We describe inconsistencies in detail and suggest options to recode the items in order to make them available for further analyses.

Please note:

⇒ This technical report *cannot* be applied on data releases v2-1-0 to 4-0-0 because from data release v2-1-0 onwards, all TwinLife data were already adjusted for filtering inconsistencies, in the sense that entries or answers not meeting the filter conditions were deleted. This procedure assumes that the filter entry question was answered correctly whereas the following questions were answered incorrectly (see TwinLife ShortGuide for further details about the data adjustment).

⇒ From data release v4-1-0 onwards, you can apply the recommendations and syntax of this technical report because unadjusted data of all variables collected in the paper-pencil survey mode (including the discrimination items) are additionally provided in one data file for each data collection. Compared to our relatively strict filtering practice, the procedures described in this technical report can prevent the loss of in fact valid data.

Change log v1.1.0

Compared to the previous version (Starr et al., 2018), the following changes have been made:

- Information on the data releases for which this technical report is valid was added in the introduction
- SPSS-syntax for recoding the unadjusted discrimination items is now provided in Appendix A
- Improved order and description of filtering inconsistencies (p. 6 ff)
- A table of content was added
- Minor changes concerning formatting and wording

Elicitation of information on experienced discrimination

- The first survey wave was performed in a period of three years from 2014 to 2016 and 4097 families were interviewed. The survey covered a wide range of topics relevant to the development of social inequality. Among these were questions about experiences with discrimination belonging to module 5 of the survey. Module 5 covers the topics health, life satisfaction, social/political participation and media use.
- All relevant participants above the age of 10 were asked to assess their experience with discrimination. Relevant participants include twins, siblings and parents and, if available, partners of twins and step-parents.
- The questionnaires were either completed as a computer-assisted version on a laptop (CASI), as an online-computer-assisted version without the presence of an interviewer (CAWI) or as a paper-pencil version (PAPI; only possible if the participant was at least 16 years old). Family members not living in the same household as the twins always got the paper-pencil version whereas persons present at the time of the interview could decide for themselves between computer-assisted or paper-pencil version. Paper-pencil was often preferred by parents to save time during the household interview[1] as they could start completing the questionnaire while the children were interviewed using a CASI module. The selected survey mode did not significantly correlate with the discrimination variable ("I experienced discrimination within the last 12 months." answered with "yes" or "no"; $\chi^2(2) = 3.25$, $p > .05$, $\varphi = .016$).
  In total, 12,224 individual persons participated in answering the questions on discrimination, which are 64% of all persons relevant to the survey. Table 1 gives an overview of the frequencies of relevant participants completing those questions in CASI, PAPI or CAWI mode.

[1] Brix J, Pupeter M, Rysina A, et al. A longitudinal twin family study of the life course and individual development (TWINLIFE): Data collection and instruments of wave 1 face-to-face interviews. TwinLife Technical Report Series. Vol 05. Bielefeld: Project TwinLife "Genetic and social causes of life chances" (Universität Bielefeld / Universität des Saarlandes); 2017.

Table 1. Frequencies of participants completing the discrimination items in CASI.

		Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
CASI	Twin 1			248	300	548
	Twin 2			238	301	539
	Partner Twin 1				25	25
	Partner Twin 2				18	18
	Sibling	50	239	124	71	484
	Mother	309	111	107	177	704
	Father	205	113	93	84	495
	Step-father	5	10	4	11	30
	Step-mother		1	1	1	3
	Total	569	474	815	988	2846

Table 2. Frequencies of participants completing the discrimination items in PAPI.

		Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
PAPI	Twin 1			812	680	1492
	Twin 2			822	680	1502
	Partner Twin 1				125	125
	Partner Twin 2				110	110
	Sibling	21	127	260	268	676
	Mother	660	889	883	741	3173
	Father	503	597	556	424	2080
	Step-father	19	25	54	21	119
	Step-mother	1	1	7	4	13
	Total	1204	1641	3394	3053	9292

Table 3. Frequencies of participants completing the discrimination items in CAWI.

		Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
CAWI	Twin1					
	Twin 2					
	Partner Twin 1				13	13
	Partner Twin 2				17	17
	Sibling	1	2	12	22	37
	Mother			1		1
	Father	2	5	9	2	18
	Step-father					
	Step-mother					
	Total	3	7	22	54	86

Filtering

- The items in the CASI, CAWI and PAPI version did not differ regarding content but while the filtering of questions in CASI and CAWI was performed automatically by the program, it was a potential source of inconsistencies in PAPI as participants did not always follow the filtering instructions given in the questionnaire.
- Overall, filtering was necessary for the question on discrimination as it consisted of three parts:
  First, we asked in general whether the participant had experienced discrimination within the last 12 months (variable dis0100 in the final dataset). The question could be answered with 'yes' (= 1) or 'no' (= 2). In case of answering 'no', in CASI further questions on discrimination were skipped and the program continued with the next topic. In PAPI participants were instructed to skip the next questions and continue with the new topic (instruction was 'If *no*, continue with question 64').
  If general discrimination was answered with 'yes', the participant was secondly asked for the cause of their discrimination by listing six possible causes (dis0200 – dis0700 and dis0799 for 'prefer not to say') with the option to multiple answers. For each of the selected causes the participant was thirdly asked to rate how heavily the discrimination weighed on them (dis0201 – dis0701) on a 4-point-rating scale from 'not at all' (= 1) to 'very strongly' (= 4).
- At every transition from one of the three items to the next, filtering inconsistencies in PAPI could occur. In the following we list the most common mistakes made and explain how to deal with the data. The occurred filtering inconsistencies lead to the data not being correctly analyzable. Therefore, we composed a SPSS syntax to recode the discrimination items in order to use the data for further analyses.

Common filtering inconsistencies and recommended recoding of variables

1) a. The participant did **not** answer the question on experienced discrimination 'yes', rated all perceived burden as 1 = 'not at all' and did not specify any cause. (We assumed participants wanted to emphasize that they did not experience discrimination and therefore are not burdened at all.)

Syntax: In the recoded version, dis0X01 (perceived burden) is set to a missing value if all items on perceived burden are rated as 1 = 'not at all' and all causes are specified as 0 = 'not mentioned' or missing.

b. The participant answered the question on experienced discrimination 'yes' and rated one or more items on perceived burden as 1 = 'not at all', but did **not** specify the corresponding cause itself, whereas they rated the perceived burden as greater than 1 = 'not at all' in other cases (with or without specifying the cause itself).
(We assumed participants did not forget to state the cause by accident, but presumably, wanted to emphasize that they are not stressed because of this kind of discrimination by stating a burden of 1 ('not at all') for the cause. The recoding to a missing value provides correct information for cases where participants rated the perceived burden as greater than 'not stressful at all' for some causes but rated all of the remaining causes as 'not stressful at all' to underline they did not experience discrimination based on these causes.)

Syntax: In the recoded version, dis0X01 (perceived burden) is set to a missing value if the question on experienced discrimination is answered 1 = 'yes' the burden is specified as 1 = 'not at all' and the corresponding cause is specified as 0 = 'not mentioned' or missing, whereas the perceived burden for one or more other causes is rated as greater than 1 = 'not at all'.

2) The participant answered the question for experienced discrimination with 'yes' and rated the perceived burden without specifying the correspondent cause. (We assumed that they overlooked the box for specifying the cause.)

Syntax: In the recoded version, dis0X00 (cause of discrimination) is set to 1 = 'mentioned' if the item on experienced discrimination (dis0100) was answered 1 = 'yes' and the corresponding burden (recoded version dis0X01_rec) was rated (i.e., not missing).

3) The participant did **not** answer/skipped the question on experienced discrimination (dis0100) but nevertheless specified at least one cause and/or rated the corresponding burden. (We assumed that all remaining cases after recoding overlooked the question on experienced discrimination by accident.)

Syntax: In the recoded version, dis0100 is set to 1 = 'yes' if at least one cause (recoded version dis0X00_rec) was specified as 1 = 'mentioned' or at least one burden (recoded version dis0X01_rec) was rated greater than 1 = 'not at all'.

4) The participant answered the question for experienced discrimination with 'yes' but did **not** fill in any of the following items on cause and/or burden.

Syntax: In the recoded version dis0799 (prefer not to say) was set to 1 = 'mentioned' if the question on experienced discrimination (recoded version dis0100_rec) was answered 1 = "yes" but no cause (recoded version dis0X00_rec) or burden (recoded version dis0X01) was specified.

Please be aware that recoding the discrimination variables according to our suggestions does not necessarily remove all possible mistakes and in individual cases can even falsify the data (e.g., participants in fact wanted to state experienced discrimination that they perceived as not stressful at all and forgot to specify dis0100 = 'yes'). These errors are kept to a minimum but cannot be excluded completely, since in some cases we cannot know without fail what participants wanted to express with their pattern of answers.
In addition, it is important to note that the missing data-codes specified in TwinLife datasets are not transferred automatically to the newly generated variables by applying the syntax enclosed in the Appendix. If the specific codes are important for your analyses, this issue should be taken care of beforehand. The missing data-codes are included in the lower part of the syntax.

Appendix A

```
*Encoding: UTF-8.
Recoding discrimination variables
Referring to variables dis0100 - dis0701.

*1a.
*We set the new variable for burden of discrimination to missing (dis0X01_rec = -95)
if burden of discrimination was specified 'not at all' for ALL causes (dis0X01 = 1)
and ALL causes of discrimination were either not mentioned or missing (dis0X00 < 1).
COMPUTE dis0201_rec = dis0201.
IF ((dis0100 ne 1)
and dis0200 < 1
and dis0300 < 1
and dis0400 < 1
and dis0500 < 1
and dis0600 < 1
and dis0700 < 1
and dis0201 = 1
and dis0301 = 1
and dis0401 = 1
and dis0501 = 1
and dis0601 = 1
and dis0701 = 1) dis0201_rec = -95.
execute.

COMPUTE dis0301_rec = dis0301.
IF ((dis0100 ne 1)
and dis0200 < 1
and dis0300 < 1
and dis0400 < 1
and dis0500 < 1
and dis0600 < 1
and dis0700 < 1
and dis0201 = 1
and dis0301 = 1
and dis0401 = 1
and dis0501 = 1
and dis0601 = 1
and dis0701 = 1) dis0301_rec = -95.
execute.

COMPUTE dis0401_rec = dis0401.
IF ((dis0100 ne 1)
and dis0200 < 1
and dis0300 < 1
and dis0400 < 1
and dis0500 < 1
and dis0600 < 1
and dis0700 < 1
and dis0201 = 1
and dis0301 = 1
and dis0401 = 1
and dis0501 = 1
and dis0601 = 1
and dis0701 = 1) dis0401_rec = -95.
execute.

COMPUTE dis0501_rec = dis0501.
IF ((dis0100 ne 1)
and dis0200 < 1
and dis0300 < 1
and dis0400 < 1
and dis0500 < 1
and dis0600 < 1
```

```
and dis0700 < 1
and dis0201 = 1
and dis0301 = 1
and dis0401 = 1
and dis0501 = 1
and dis0601 = 1
and dis0701 = 1) dis0501_rec = -95.
execute.

COMPUTE dis0601_rec = dis0601.
IF ((dis0100 ne 1)
and dis0200 < 1
and dis0300 < 1
and dis0400 < 1
and dis0500 < 1
and dis0600 < 1
and dis0700 < 1
and dis0201 = 1
and dis0301 = 1
and dis0401 = 1
and dis0501 = 1
and dis0601 = 1
and dis0701 = 1) dis0601_rec = -95.
execute.

COMPUTE dis0701_rec = dis0701.
IF ((dis0100 ne 1)
and dis0200 < 1
and dis0300 < 1
and dis0400 < 1
and dis0500 < 1
and dis0600 < 1
and dis0700 < 1
and dis0201 = 1
and dis0301 = 1
and dis0401 = 1
and dis0501 = 1
and dis0601 = 1
and dis0701 = 1) dis0701_rec = -95.
execute.

*1b.
*We set the new variable for burden of discrimination to missing (dis0X01_rec = -95)
if experienced discrimination was answered 'yes' (dis0100 = 1),
and burden of discrimination was answered 'not at all' (dis0X01 = 1)
and the corresponding cause of discrimination was not mentioned or missing (dis0X00 < 1)
for one or more causes
while for other causes the burden of discrimination was answered greater than 'not at all'
(dis0X01 > 1).
IF ((dis0100 = 1
and dis0201 = 1 and dis0200 < 1)
and (dis0301 > 1
or dis0401 > 1
or dis0501 > 1
or dis0601 > 1
or dis0701 > 1)) dis0201_rec = -95.
execute.

IF ((dis0100 = 1
and dis0301 = 1 and dis0300 < 1)
and (dis0201 > 1
or dis0401 > 1
or dis0501 > 1
or dis0601 > 1
or dis0701 > 1)) dis0301_rec = -95.
execute.
```

```
IF ((dis0100 = 1
and dis0401 = 1 and dis0400 < 1)
and (dis0201 > 1
or dis0301 > 1
or dis0501 > 1
or dis0601 > 1
or dis0701 > 1)) dis0401_rec = -95.
execute.

IF ((dis0100 = 1
and dis0501 = 1 and dis0500 < 1)
and (dis0201 > 1
or dis0301 > 1
or dis0401 > 1
or dis0601 > 1
or dis0701 > 1)) dis0501_rec = -95.
execute.

IF ((dis0100 = 1
and dis0601 = 1 and dis0600 < 1)
and (dis0201 > 1
or dis0301 > 1
or dis0401 > 1
or dis0501 > 1
or dis0701 > 1)) dis0601_rec = -95.
execute.

IF ((dis0100 = 1
and dis0701 = 1 and dis0700 < 1)
and (dis0201 > 1
or dis0301 > 1
or dis0401 > 1
or dis0501 > 1
or dis0601 > 1)) dis0701_rec = -95.
execute.

*2.
*We set the new variable for cause of discrimination to 'mentioned' (dis0X00_rec = 1)
if experienced discrimination was answered 'yes' (dis0100 = 1)
and the corresponding burden of discrimination was specified (dis0X01_rec > 0).
*discrimination on grounds of ethnic or cultural background.
COMPUTE dis0200_rec = dis0200.
IF (dis0100 = 1 and dis0201_rec > 0) dis0200_rec = 1.

*discrimination on grounds of sex.
COMPUTE dis0300_rec = dis0300.
IF (dis0100 = 1 and dis0301_rec > 0) dis0300_rec = 1.

*discrimination on grounds of religion.
COMPUTE dis0400_rec = dis0400.
IF (dis0100 = 1 and dis0401_rec > 0) dis0400_rec = 1.

*discrimination on grounds of age.
COMPUTE dis0500_rec = dis0500.
IF (dis0100 = 1 and dis0501_rec > 0) dis0500_rec = 1.

*discrimination on grounds of sexual identity.
COMPUTE dis0600_rec = dis0600.
IF (dis0100 = 1 and dis0601_rec > 0) dis0600_rec = 1.

*discrimination on grounds of disability.
COMPUTE dis0700_rec = dis0700.
IF (dis0100 = 1 and dis0701_rec >0) dis0700_rec = 1.
execute.

*3.
*We set the new variable for experienced discrimination to 'yes' (dis0100_rec = 1)
if at least one cause was mentioned (dis0X00_rec = 1)
```

```
or the burden of discrimination was specified as 'somewhat onerous' or greater
(dis0X01_rec > 1).
COMPUTE dis0100_rec = dis0100.
IF (dis0200_rec = 1 or
dis0300_rec = 1 or
dis0400_rec = 1 or
dis0500_rec = 1 or
dis0600_rec = 1 or
dis0700_rec = 1) dis0100_rec = 1.

IF (dis0201_rec > 1 or
dis0301_rec > 1 or
dis0401_rec > 1 or
dis0501_rec > 1 or
dis0601_rec > 1 or
dis0701_rec > 1) dis0100_rec = 1.
execute.

*4.
*We set the new variable for 'prefer not to say' to 'mentioned' (dis0799_rec = 1)
if experienced discrimination was answered 'yes' (dis0100_rec = 1)
but no cause (dis0X00_rec) or burden of discrimination (dis0X01_rec) were specified.
COMPUTE dis0799_rec = dis0799.
IF (dis0100_rec = 1 and
dis0200_rec < 1 and
dis0300_rec < 1 and
dis0400_rec < 1 and
dis0500_rec < 1 and
dis0600_rec < 1 and
dis0700_rec < 1 and
dis0201_rec < 1 and
dis0301_rec < 1 and
dis0401_rec < 1 and
dis0501_rec < 1 and
dis0601_rec < 1 and
dis0701_rec < 1) dis0799_rec = 1.
execute.

*5.
*Labels for all new variables.
VARIABLE LABELS
dis0100_rec 'recoded - experience with discrimination'
dis0799_rec 'recoded - discrimination because of: not specified (refused to answer)'
dis0200_rec 'recoded - discrimination on grounds of ethnic or cultural background'
dis0300_rec 'recoded - discrimination on grounds of sex'
dis0400_rec 'recoded - discrimination on grounds of religion'
dis0500_rec 'recoded - discrimination on grounds of age'
dis0600_rec 'recoded - discrimination on grounds of sexual identity'
dis0700_rec 'recoded - discrimination on grounds of disability'
dis0201_rec 'recoded - burden of discrimination on grounds of ethnic or cultural
background'
dis0301_rec 'recoded - burden of discrimination on grounds of sex'
dis0401_rec 'recoded - burden of discrimination on grounds of religion'
dis0501_rec 'recoded - burden of discrimination on grounds of age'
dis0601_rec 'recoded - burden of discrimination on grounds of sexual identity'
dis0701_rec 'recoded - burden of discrimination on grounds of disability'.
execute.

*6.
*Labels for values of the new variables.
VALUE LABELS
dis0100_rec
-99 'not specified (refused to answer)'
-98 'do not know'
-97 'does not apply (specification of other category)'
-96 'I cannot remember'
-95 'does not apply (screened out)'
```

```
-94 'technical error/faulty insertion'
-93 'unclear classification of system missing'
-92 'no participation in survey module'
-91 'reply on request'
-90 'no participation in survey wave'
-89 'unreadable'
-88 'subject legible but grade illegible'
-87 'multiple answers'
-86 'not available/empty/not codable'
1 '1: yes'
2 '2: no'.
execute.

VALUE LABELS
dis0200_rec to dis0700_rec
-99 'not specified (refused to answer)'
-98 'do not know'
-97 'does not apply (specification of other category)'
-96 'I cannot remember'
-95 'does not apply (screened out)'
-94 'technical error/faulty insertion'
-93 'unclear classification of system missing'
-92 'no participation in survey module'
-91 'reply on request'
-90 'no participation in survey wave'
-89 'unreadable'
-88 'subject legible but grade illegible'
-87 'multiple answers'
-86 'not available/empty/not codable'
0 '0: not mentioned'
1 '1: mentioned'.
execute.

VALUE LABELS
dis0201_rec to dis0701_rec
-99 'not specified (refused to answer)'
-98 'do not know'
-97 'does not apply (specification of other category)'
-96 'I cannot remember'
-95 'does not apply (screened out)'
-94 'technical error/faulty insertion'
-93 'unclear classification of system missing'
-92 'no participation in survey module'
-91 'reply on request'
-90 'no participation in survey wave'
-89 'unreadable'
-88 'subject legible but grade illegible'
-87 'multiple answers'
-86 'not available/empty/not codable'
1 '1: not at all'
4 '4: very strongly'.
execute.

VALUE LABELS
dis0799_rec
-99 'not specified (refused to answer)'
-98 'do not know'
-97 'does not apply (specification of other category)'
-96 'I cannot remember'
-95 'does not apply (screened out)'
-94 'technical error/faulty insertion'
-93 'unclear classification of system missing'
-92 'no participation in survey module'
-91 'reply on request'
-90 'no participation in survey wave'
-89 'unreadable'
-88 'subject legible but grade illegible'
```

```
-87 'multiple answers'
-86 'not available/empty/not codable'
0 '0: not mentioned'
1 '1: mentioned'.
execute.

*7.
*Specifying missing values.
MISSING VALUES
dis0201_rec to dis0799_rec (-99 THRU -86).
execute.

*end of syntax.
```