TwinLife Technical Report Series

No. 07, April 2018

Documentation TwinLife Data:

Filtering inconsistencies in discrimination items

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

alexandra.starr@uni-bielefeld.de

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

TwinLife Technical Report Series No. 07
Project TwinLife “Genetic and social causes of life chances”
Bielefeld, April 2018

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:
http://www.twin-life.de/en/twinlife-technical-report-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: http://www.twin-life.de/en

Universität Bielefeld

UNIVERSITÄT DES SAARLANDES

Content

List of Figures and Tables

This technical report deals with filtering inconsistencies in items of 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. After describing the inconsistencies in detail a way to recode the items is suggested in order to make them available for further research.

1. Elicitation of information on experiences with discrimination

The first survey wave was performed in a period of three years from 2014 to 2016 and 4097 families were interviewed. The whole survey covered a wide range of topics relevant to the main focus to investigate 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, contentment/satisfaction, politics and media.

All persons above the age of 10 relevant for the survey were asked to assess their experience with discrimination. Relevant persons include twins, siblings and parents and if existent, 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[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 month." answered with "yes" or "no"; $\chi^2(2) = 3.25$, $p > .05$, $\phi = .016$).

In total, 12224 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

2. Filtering

The items in the CASI, CAWI and PAPI version did not differ with regard to 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 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 factors of their discrimination by listing six possible factors(dis0200 – dis0700 and dis0799 for 'prefer not to say') with the option to multiple answers. For each of the selected factors 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' to 'very strongly'.

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 syntax for SPSS to recode the discrimination items in order to use the data for further analyses.

3. Common filtering inconsistencies and recoding of variables

1) The participant answered the question for experienced discrimination with 'no' or did not answer it at all, but nonetheless did not skip the following two questions but reported a factor of discrimination and/or the perceived burden.

   Syntax: In the recoded version item dis0100_rec (asking about experienced discrimination in general) is set to 1 ('yes') automatically, if at least one factor was specified or the burden of discrimination for at least one factor was specified greater

than 1. We assumed that the affected participants wanted to state experienced discrimination, but accidently specified ‘no’ in dis0100 or overlooked the item. For this recoding, the burden of discrimination needs to be specified greater than 1 because another kind of filtering inconsistency (see 4.) should not be included due to probable misunderstandings of several discrimination items.

2) The participant answered the question for experienced discrimination with ‘yes’ but did not fill in any of the following items on factor and burden.

   Syntax: If dis0100 was answered with ‘yes’ but no factor or burden was specified, the newly coded item dis0799_rec was set to 1 indicating that the person prefers not to answer specific questions about his or her discrimination experiences or skipped the questions by accident.

3) The participant answered the question for experienced discrimination with ‘yes’ and specified the perceived burden without choosing the factor (by checking the box).

   Syntax: If dis0100 was answered with ‘yes’ and the participant specified the burden of discrimination for at least one factor as greater than 1, the new variables (dis0200_rec – dis0700_rec) for the corresponding factors (dis0200 – dis0700) are set to 1, as we assumed the affected participants simply overlooked or forgot to specify the factor itself in addition to rating the perceived burden.

4) The participant answered the question on experienced discrimination with either ‘yes’ or ‘no’, rated the perceived burden for at least one factor as not stressful at all, but did not specify the corresponding factor of discrimination.

   Syntax: In the recoded version the variables dis0201_rec – dis0701_rec are set to a missing value (-95) if the participant specified the corresponding burden as not stressful at all (1) and did not specify the according factor of discrimination. We assumed the participants did not forget to state the factor by accident, but presumably, wanted to emphasize that they are not stressed because of any kind of discrimination by stating a burden of 1 (‘not at all’) for the factor. This assumption especially applies to those cases where the participants did not specify any factor of discrimination but rated the burden for all factors as ‘not stressful at all’. The recoding to a missing value also provides correct information if participants specified some factors of discrimination and the corresponding burden as greater than ‘not stressful at all’, but rated all of the remaining factors as ‘not stressful at all’ to underline they did not experience discrimination based on these factors.

Nevertheless, by recoding the discrimination variables not all mistakes can be removed 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 really wanted to express with their pattern of answers.