

SOEP Survey Papers

Series C – Data Documentations

SOEP – The German Socio-Economic Panel study at DIW Berlin

2015

Flowcharts for the Integrated Individual-Biography Questionnaire of the IAB-SOEP Migration Sample 2013

Klaudia Erhardt

Running since 1984, the German Socio-Economic Panel study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing.

The SOEP Survey Papers is comprised of the following series:

Series A – Survey Instruments (Erhebungsinstrumente)

Series B – Survey Reports (Methodenberichte)

Series C – Data Documentation (Datendokumentationen)

Series D – Variable Descriptions and Coding

Series E – SOEPmonitors

Series F – SOEP Newsletters

Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at
<http://www.diw.de/soepsurveypapers>

Editors:

Dr. Jan Goebel, DIW Berlin

Prof. Dr. Martin Kroh, DIW Berlin and Humboldt Universität Berlin

Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin

Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Klaudia Erhardt. 2015. Flowcharts for the Integrated Individual-Biography Questionnaire of the IAB-SOEP Migration Sample 2013. SOEP Survey Papers 261: Series C. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

Contact: DIW Berlin
SOEP
Mohrenstr. 58
10117 Berlin

Email: soeppapers@diw.de

Klaudia Erhardt

**Flowcharts for the Integrated Individual-Biography
Questionnaire of the IAB-SOEP Migration Sample 2013**

Berlin 2015

Klaudia Erhardt, SOEP, DIW
kerhardt@diw.de

last updated: 2015-01-22

Contents:

Introduction	3
Flowcharts	6
Citizenship and origin	6
Migration biography: Coming to Germany	8
Migration biography: Stays abroad	9
Cultural ties, language and social community	10
Childhood and parental home	12
Education	15
Vocational training or higher education in other countries than DE	17
Employment	20
Partnership	25
Income	26
Satisfaction, health, and sickness / Attitudes and religion	28
Appendix A: Annotations	31
Appendix B: Filter Definitions	32

Introduction

Filters in questionnaires lead to “gaps” in the data, because subgroups of respondents skip certain questions. One main reason for inserting filters into a questionnaire is that it makes no sense to ask some questions to some respondents. For example, it would be inappropriate to ask a respondent who reported being unemployed what her last month's salary was.

Beyond this, a second factor leads to increasing numbers of filters in questionnaires: the objective of keeping the burden of a participation in the survey as low as possible for the respondents. To this end, surveys avoid as far as possible asking questions where the information can be obtained by other means. For instance, in the IAB-SOEP migration sample, respondents are asked some questions relating to their mother and father only if they do not live in the same household. Otherwise—as the SOEP aims at interviewing every adult household member—the information is obtained by asking the parents themselves.

When using CAPI and online questionnaires, filters can be implemented easily while programming the user interface. As these techniques have spread widely, the filters in questionnaires have become increasingly sophisticated and complex.

For data analysis, it is crucial to keep track of the filter processes that took place during data collection. To know what part of a population the conclusions of the analysis apply to, it has to be unambiguously clear what the reference subgroup is for the distribution of the variables in question. Also, if questions had been skipped with subgroups of the respondents because the information was obtainable from other sources, the relevant information is spread among several variables and has to be collected before analyzing the data. This step also requires awareness of the underlying filters.

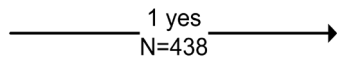
As filters grow more and more sophisticated, it becomes increasingly difficult to keep track of them, especially because the skip instructions in the documented questionnaires often do not explain the “does not apply” cases of a variable. To help determine which respondents have been asked a question and which ones have not, this survey paper presents comprehensive flowcharts for the integrated individual-biography questionnaire of wave 2013 of the IAB-SOEP Migration Sample.

How to read the flowcharts

The flowcharts are more or less self-explanatory. However, the meaning of some design elements should be noted:

- Questions are displayed within rectangular boxes framed by a continuous line.

- Shapes framed by a dashed line indicate questions or modules outside the module displayed in the current chart.
- Questions allowing for multiple responses are represented by light blue rectangular shapes. Questions allowing for multiple responses change the totals from N of respondents to N of answers.
- Flows between questions are represented by an arrow that points towards the next question:



- In a few cases, for lack of space, it was not possible to display each flow separately, some had to be depicted as flowing into an other flow. These “confluences” are represented as an arrow with a round tip and a pointed tail:



- References to annotations are written in red.
- Complex filters are explained in the appendix. In the flowcharts, they are referred to with short labels for the different groups of respondents created by the respective filters.

In general, the difference between the total number of cases (respondents) and the number of cases that enter a question amounts to the number of “does not apply” cases of the variables resulting from that question. (An exception to this rule is described in annotation 3, see Appendix A). Thus the flowcharts provide a means of following each respondent group through the questionnaire and to see precisely which groups make up the “does not apply” cases.

Abbreviations used in the flowcharts

biDE born in Germany

boDE born outside of Germany

DE Germany

FRG Federal Republic of Germany (former “West Germany”)

GDR German Democratic Republic (former “East Germany”)

hh household

n.a. no answer

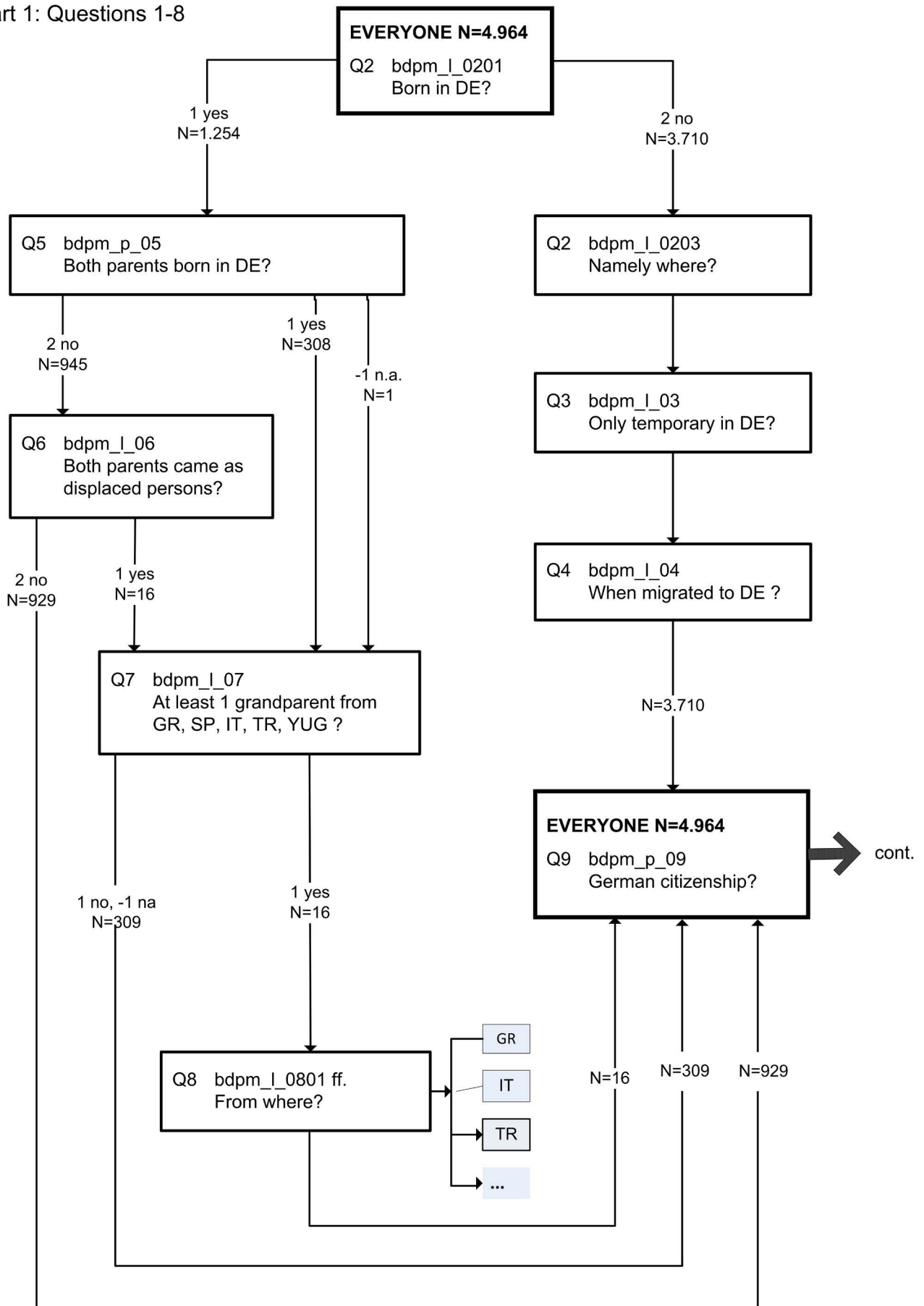
Related documents

- The original questionnaires are reproduced as SOEP Survey Paper No. 218 (see SOEP Survey Paper No. 219 for an English translation).
- Useful for understanding the variables of the migration biography¹ are the two synopses, pp. 15 and 16 in: Klaudia Erhardt. 2014. How to Generate Spell Data from Data in “Wide” Format (SOEP Survey Paper No. 228).

¹ In wide format the migration biography is part of the dataset bdp_mig. The dataset migspell contains the migration biography in spell format. Both datasets are part of the IAB-SOEP Migration Sample 2013

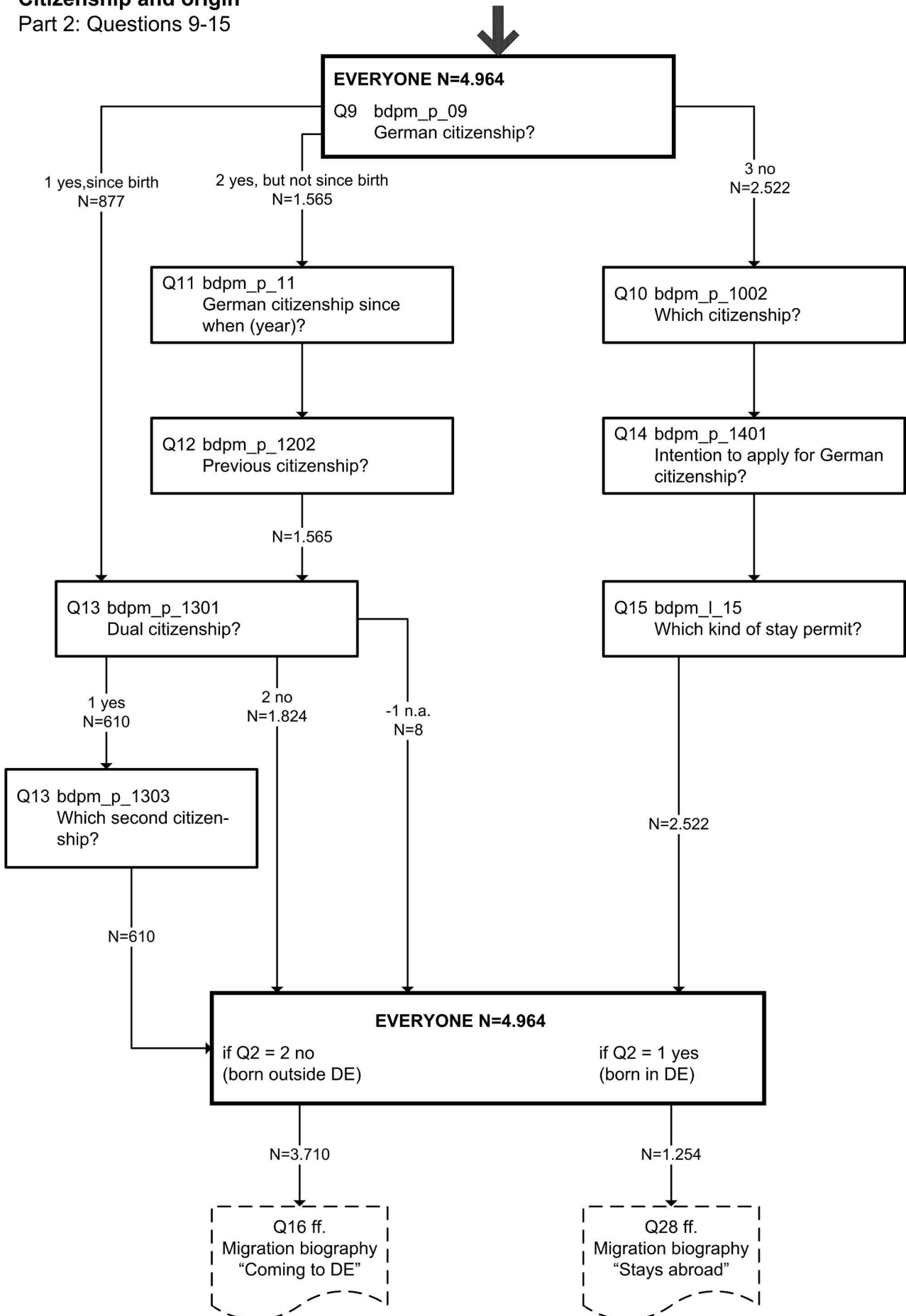
Citizenship and origin

Part 1: Questions 1-8



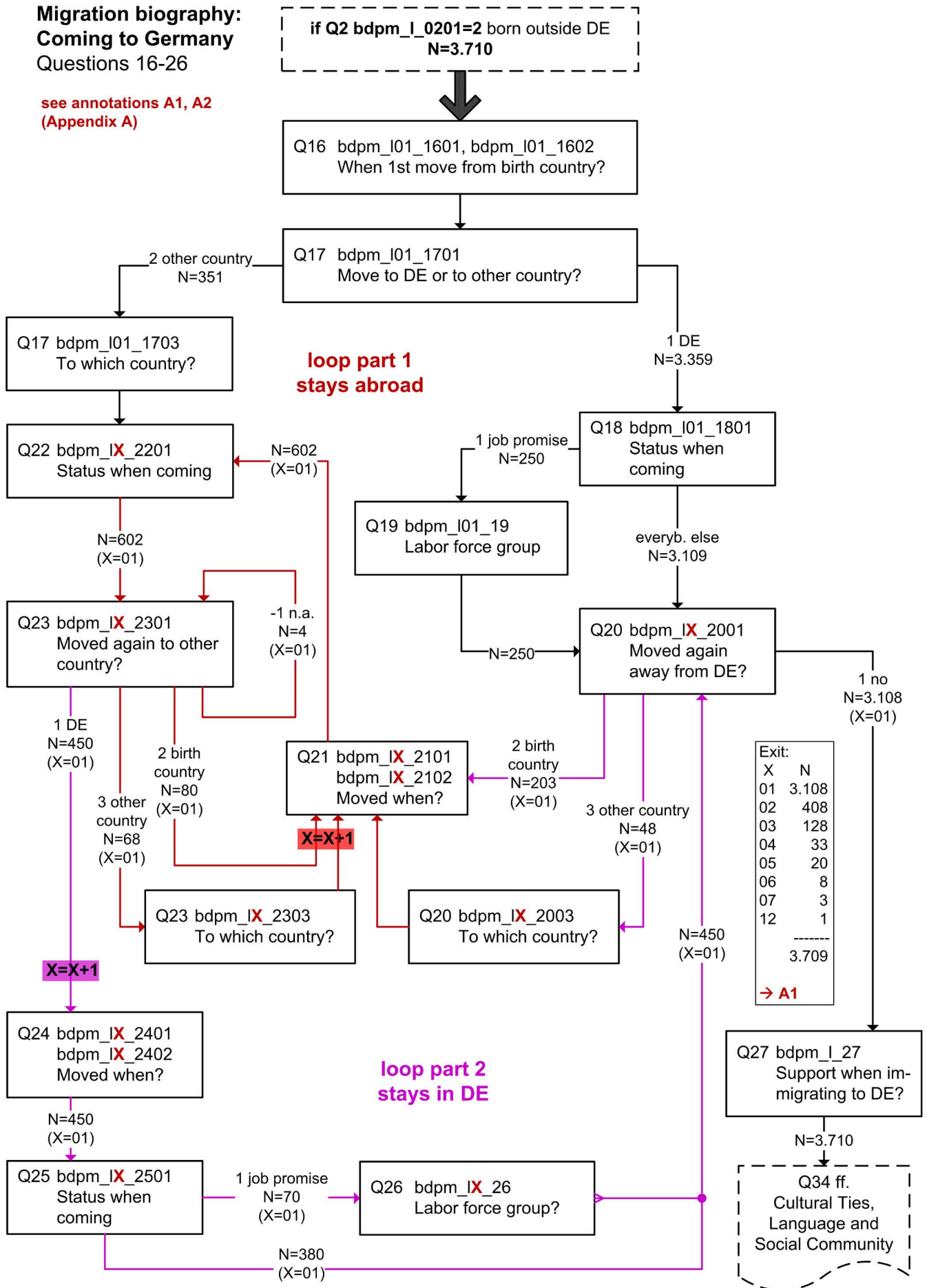
Citizenship and origin

Part 2: Questions 9-15



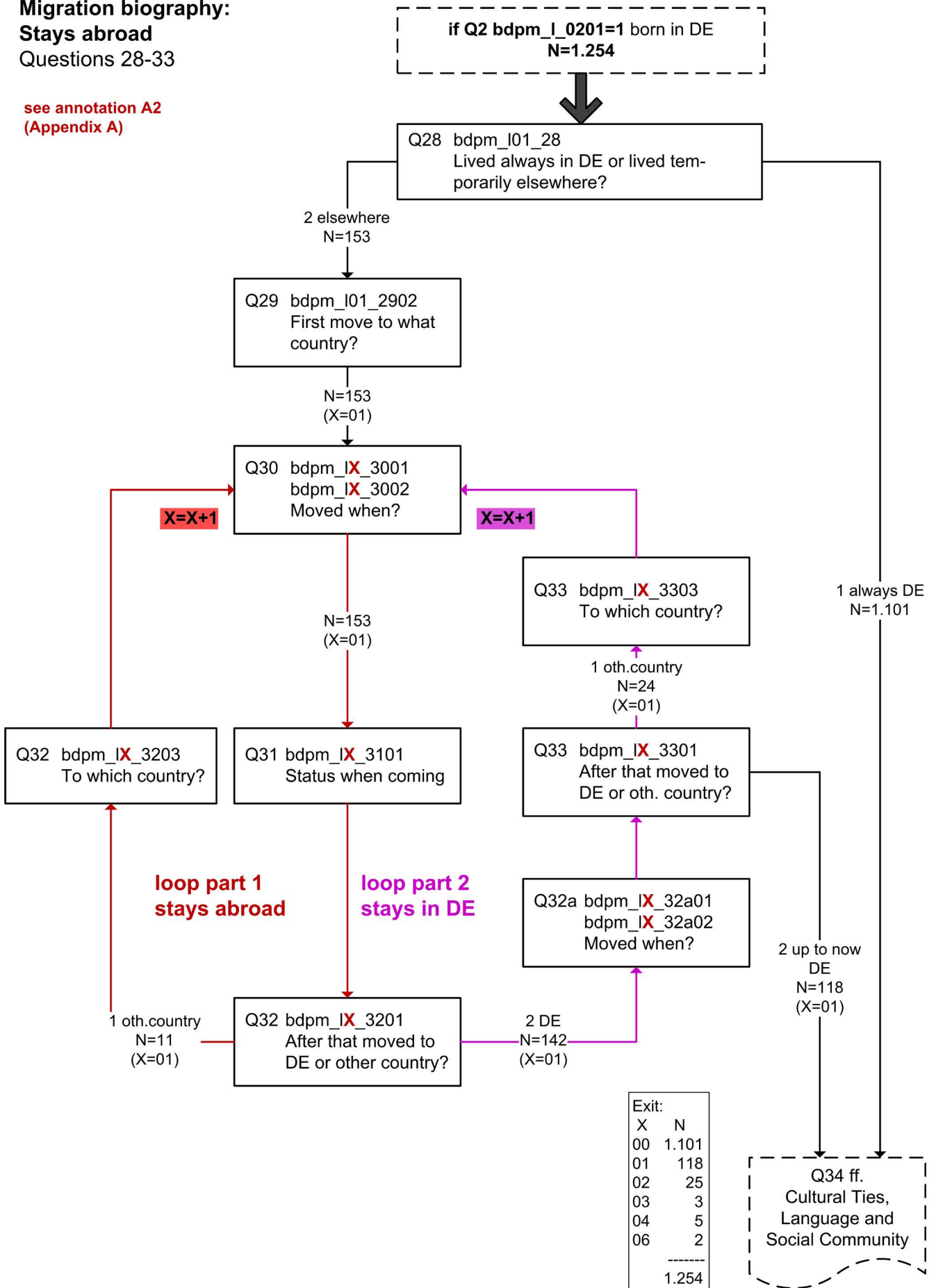
**Migration biography:
Coming to Germany**
Questions 16-26

see annotations A1, A2
(Appendix A)

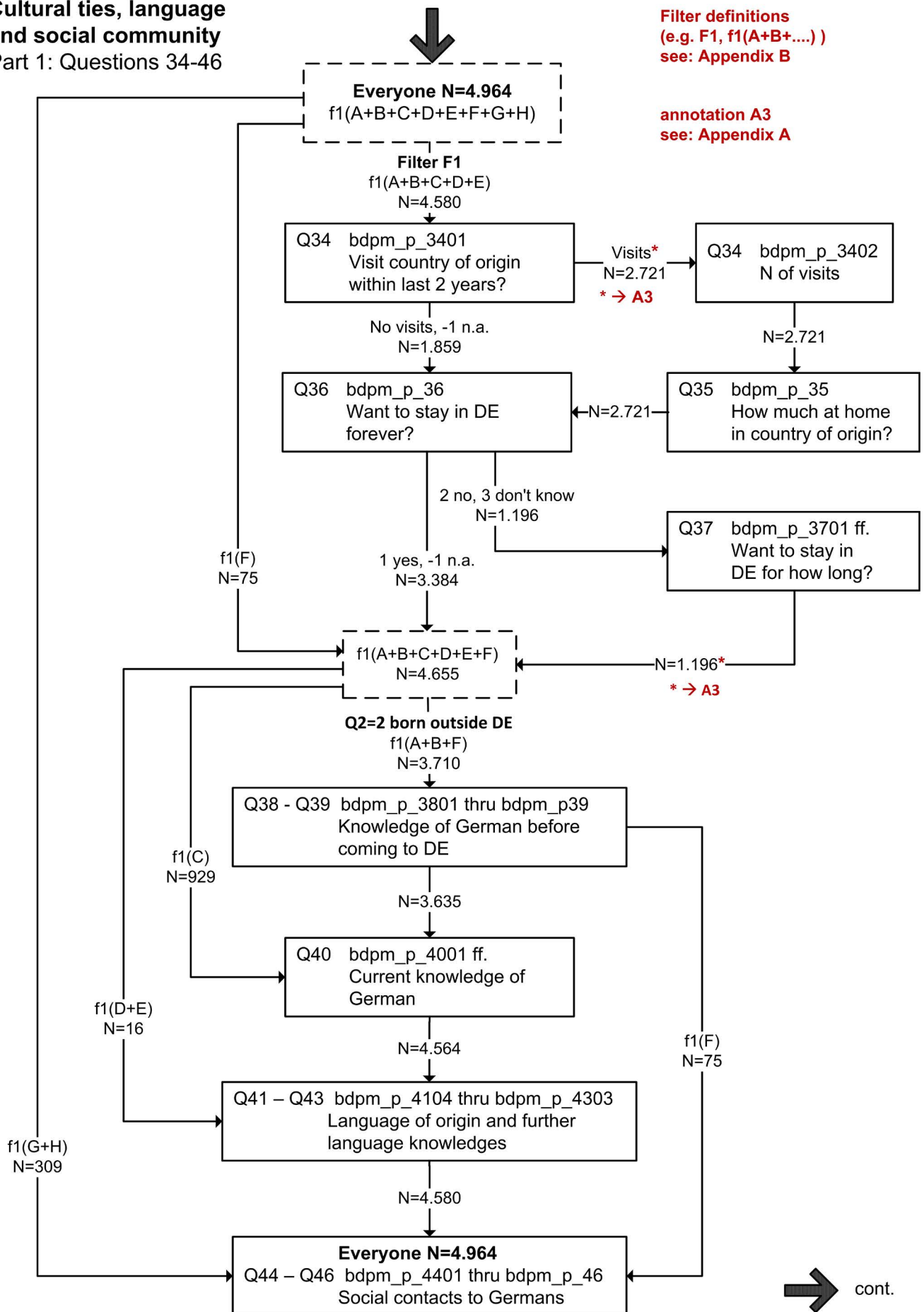


**Migration biography:
Stays abroad**
Questions 28-33

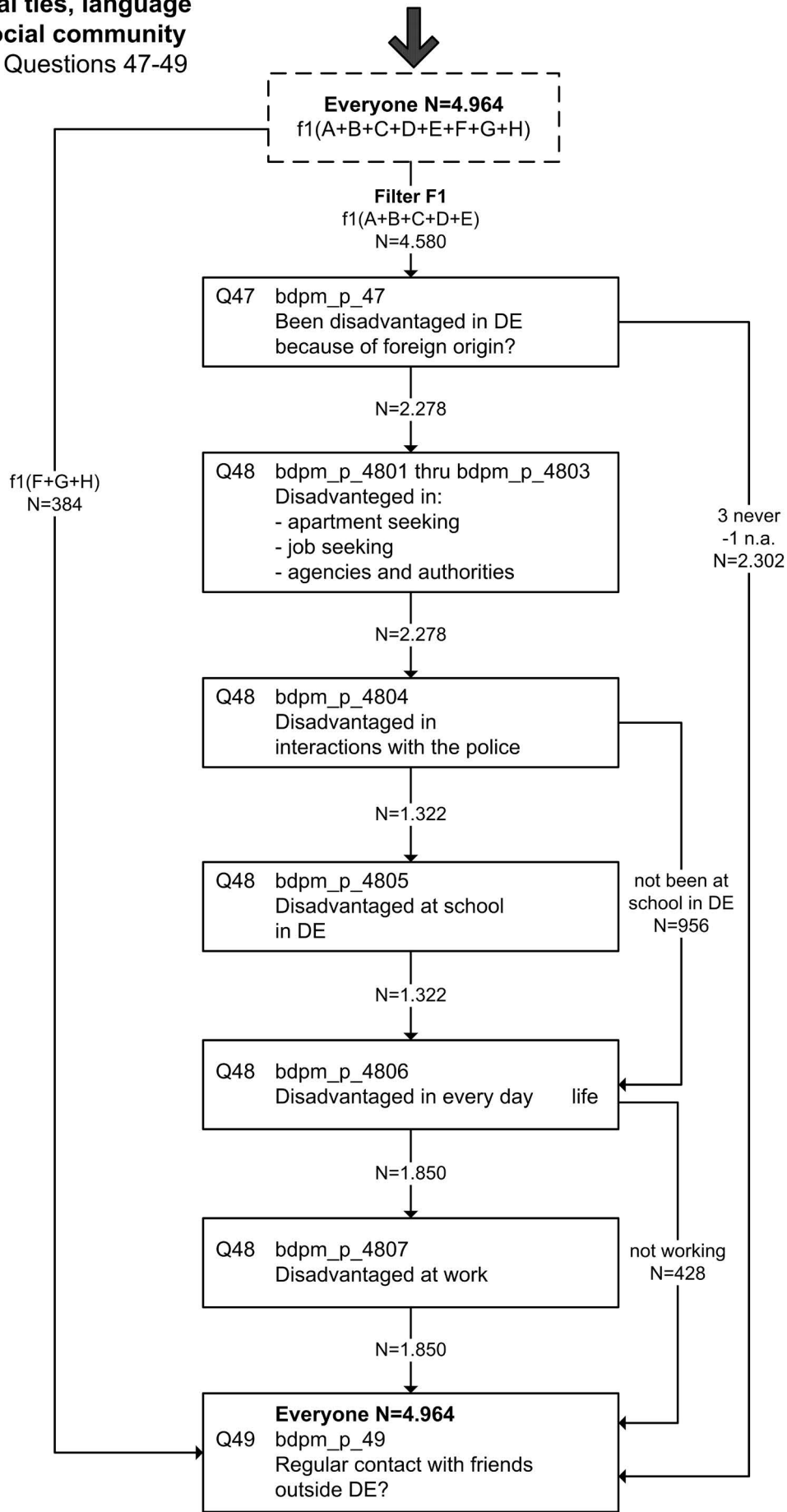
see annotation A2
(Appendix A)



Cultural ties, language and social community
Part 1: Questions 34-46



Cultural ties, language and social community
Part 2: Questions 47-49



➔ cont.

Childhood and parental home

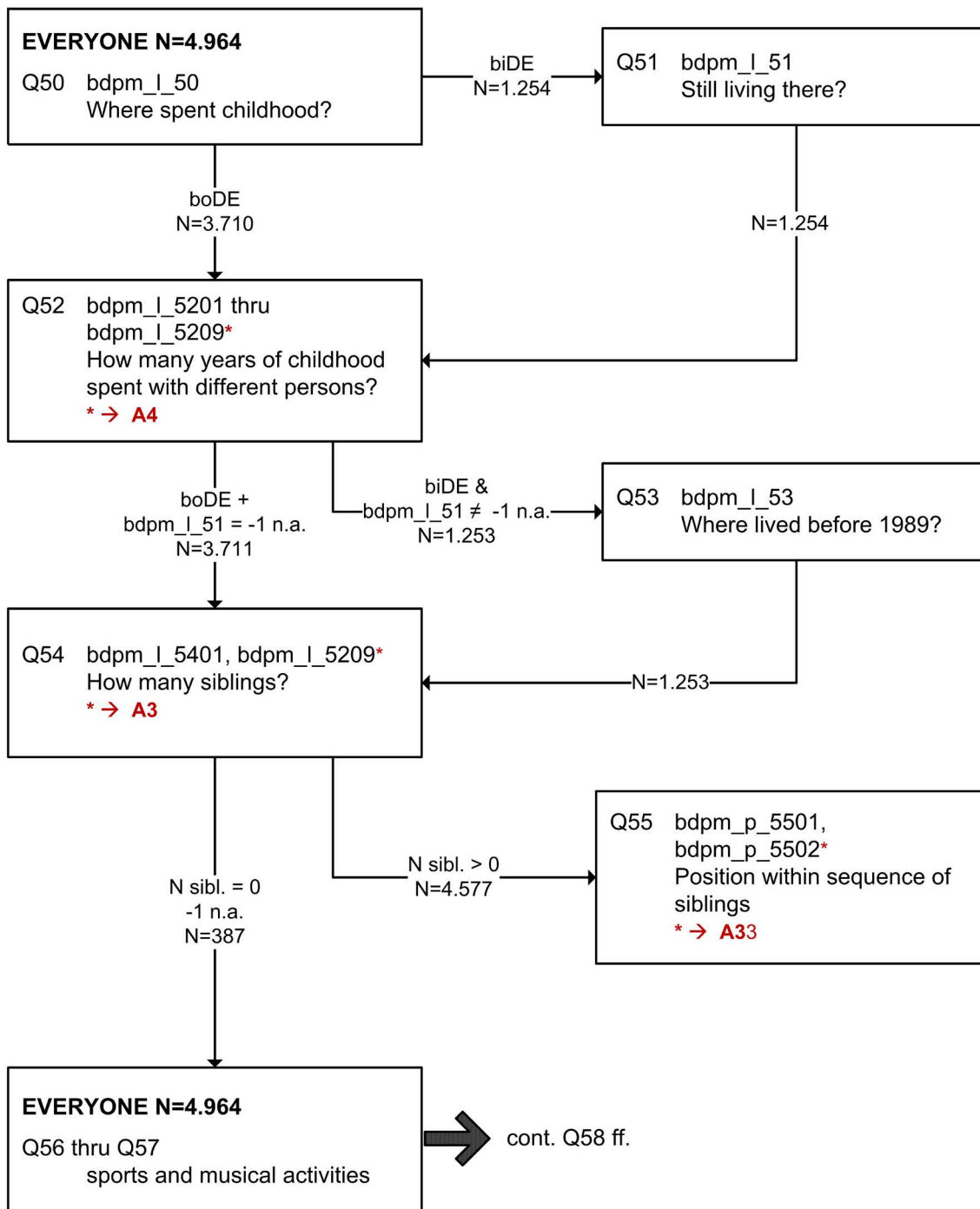
Part 1: Questions 50-57

annotations A3 and A4
see: Appendix A

Filters (→ Q2):

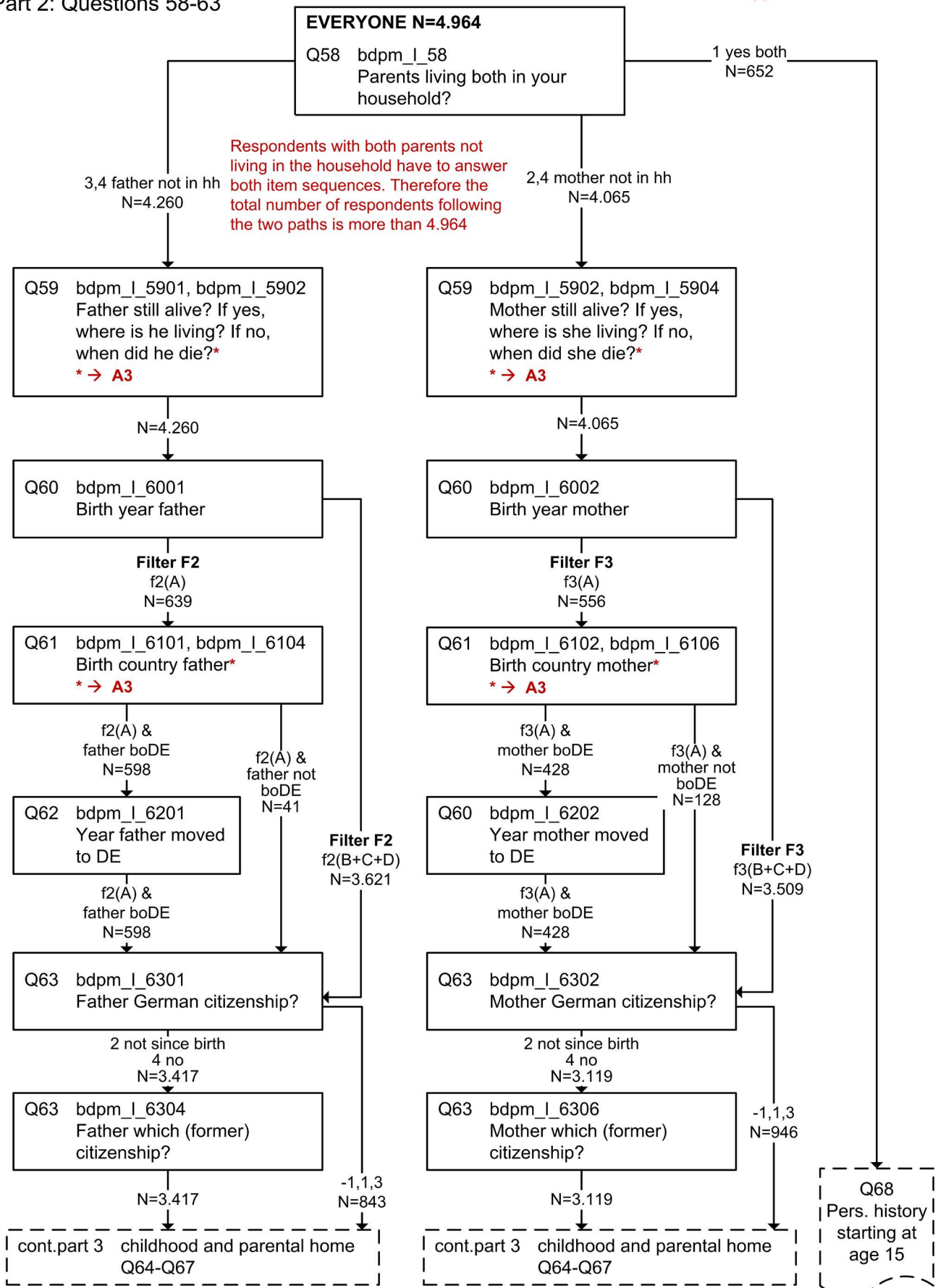
biDE born in Germany,
N = 1.254

boDE born outside of Germany,
N = 3.710

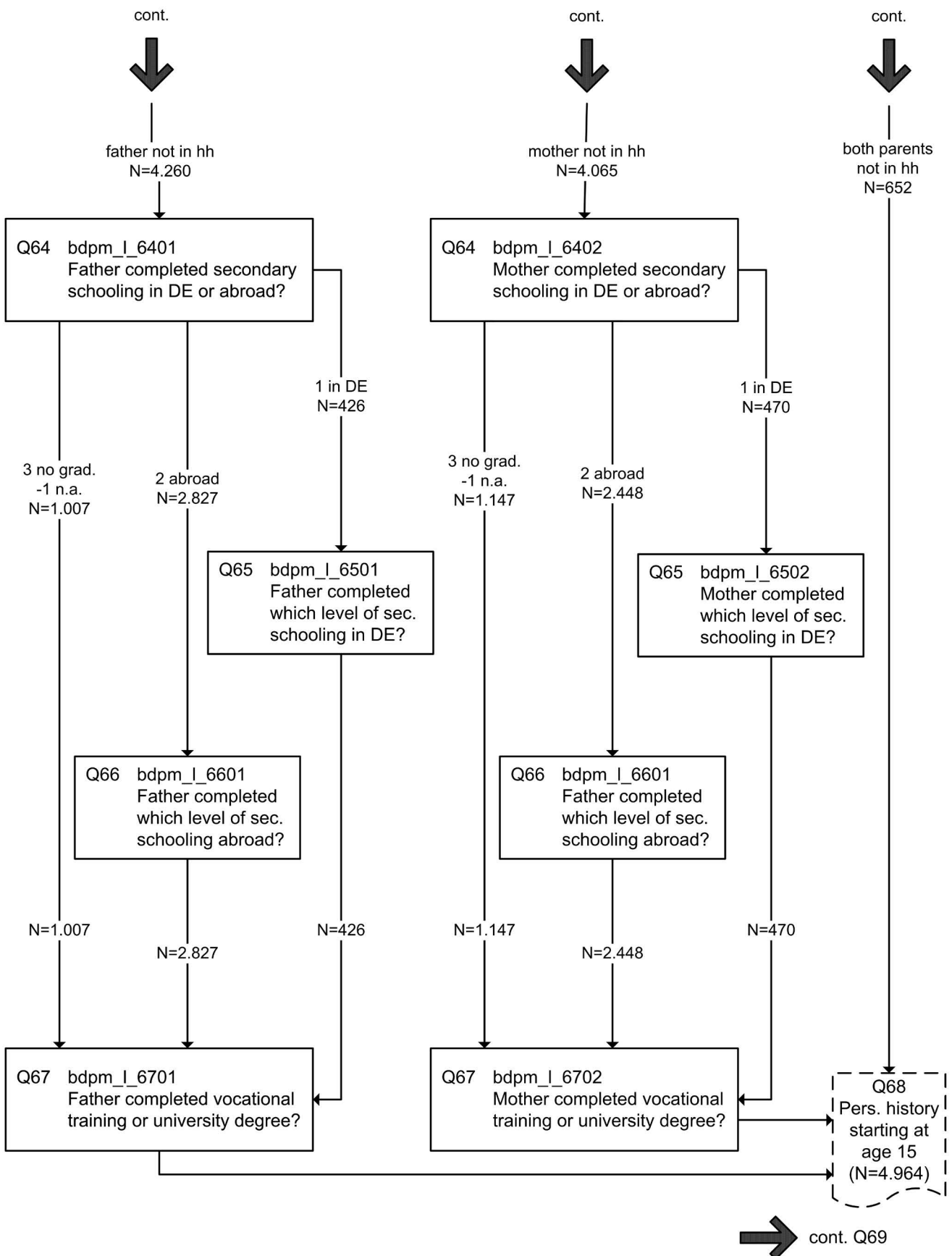


Childhood and parental home
Part 2: Questions 58-63

annotation A3
see: Appendix A



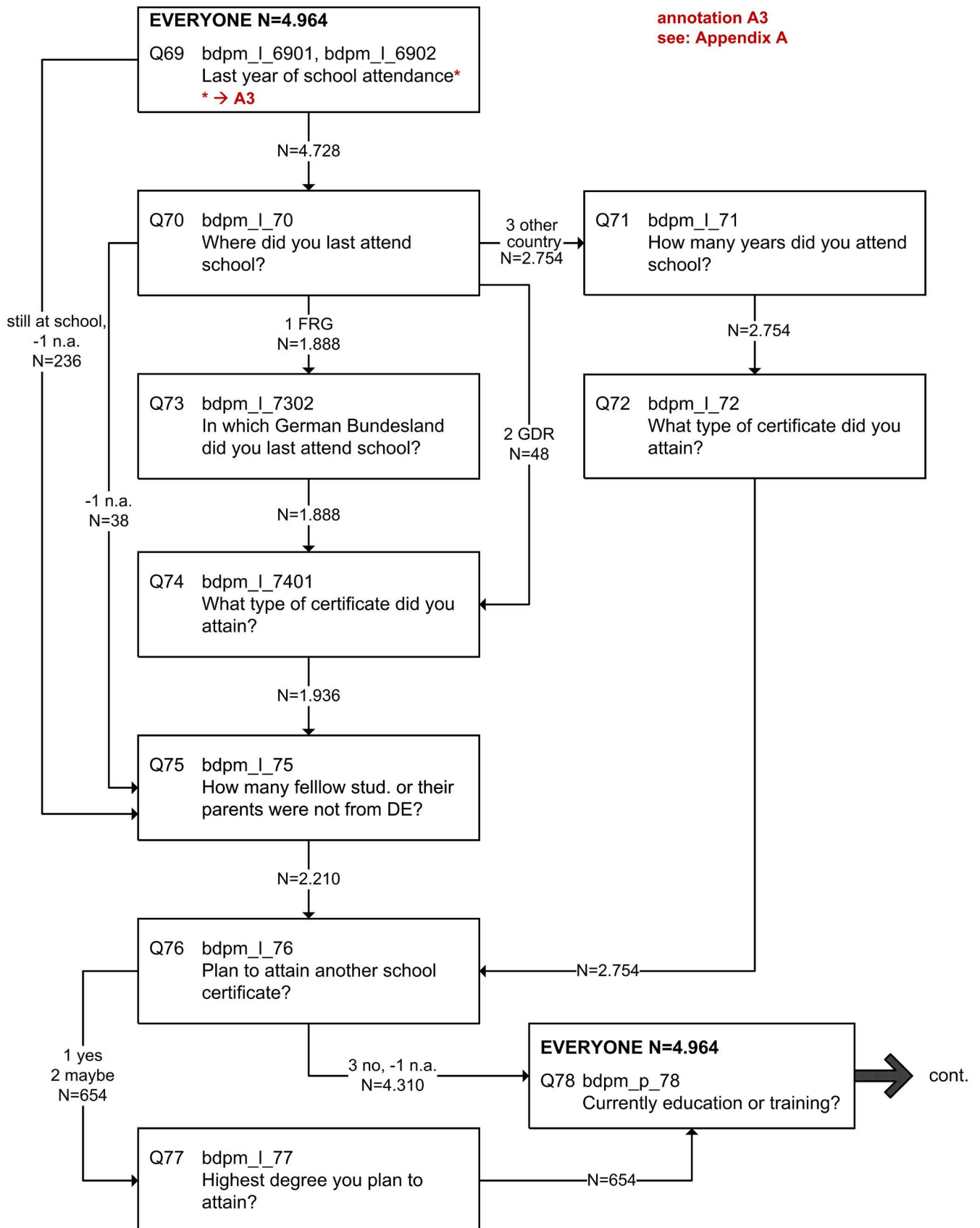
Childhood and parental home
Part 3: Questions 64-67

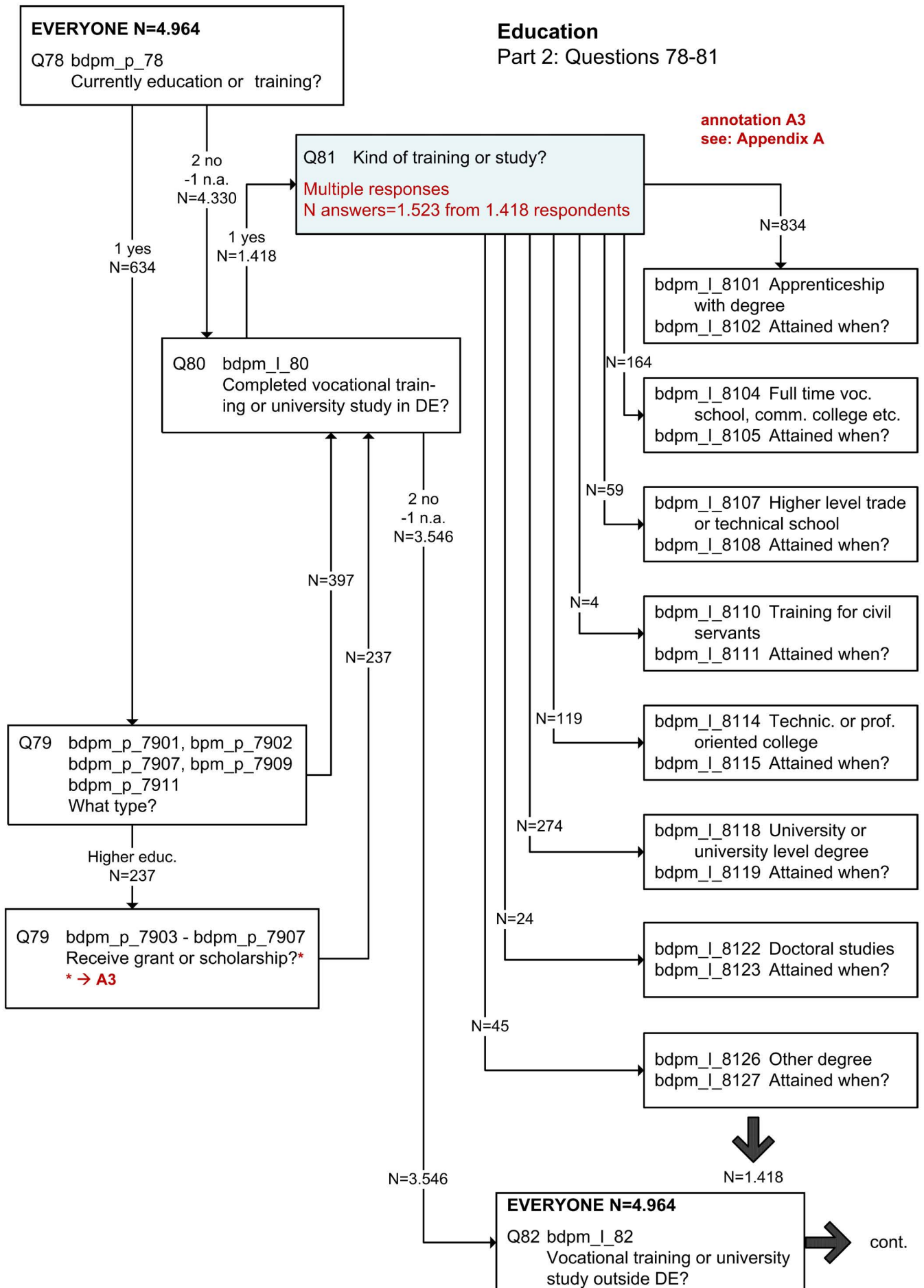


Education

Part 1: Questions 69-77

**annotation A3
see: Appendix A**





Vocational training or higher education in other countries than DE
Part 1: Questions 82-83

Annotation:

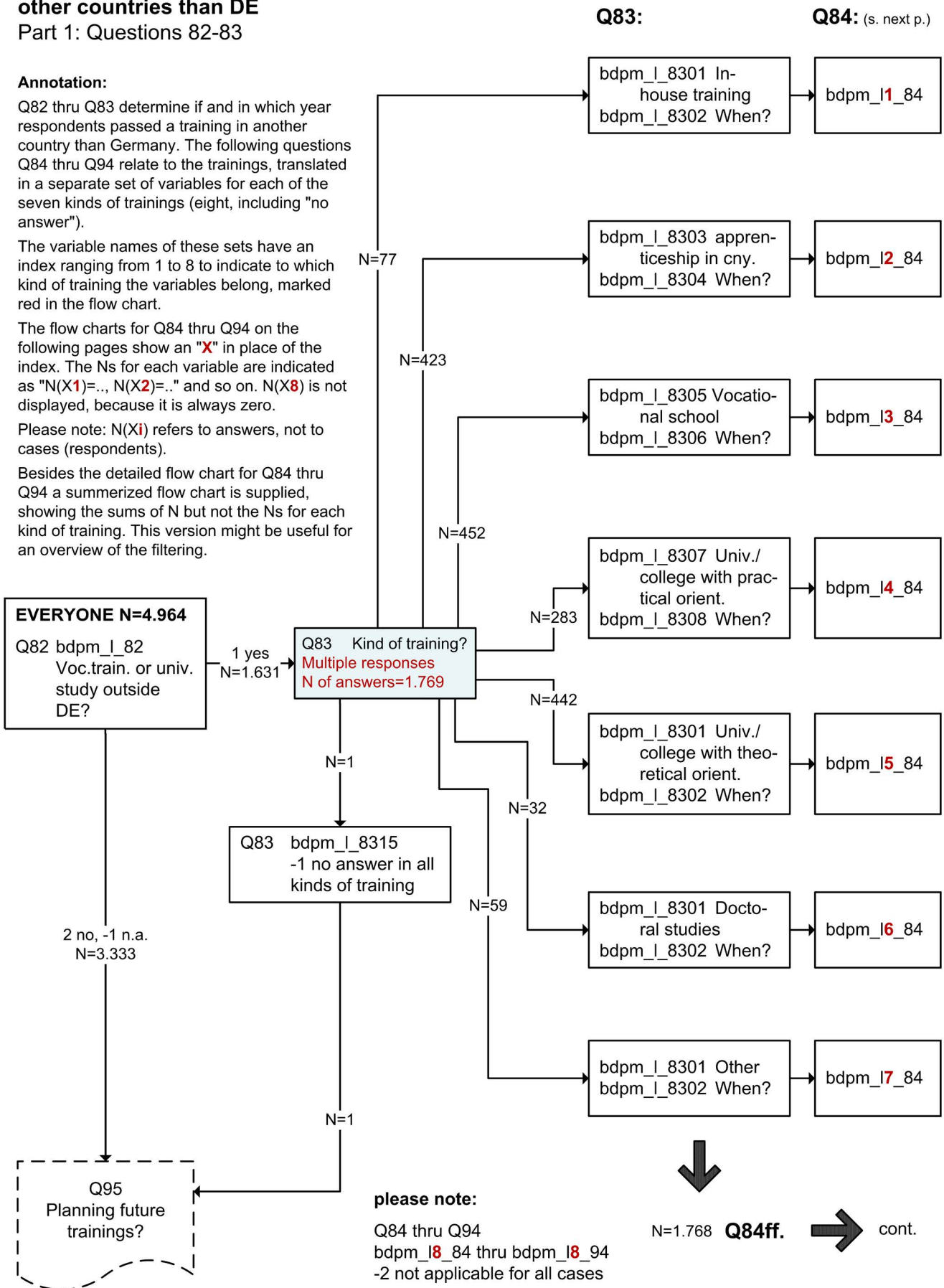
Q82 thru Q83 determine if and in which year respondents passed a training in another country than Germany. The following questions Q84 thru Q94 relate to the trainings, translated in a separate set of variables for each of the seven kinds of trainings (eight, including "no answer").

The variable names of these sets have an index ranging from 1 to 8 to indicate to which kind of training the variables belong, marked red in the flow chart.

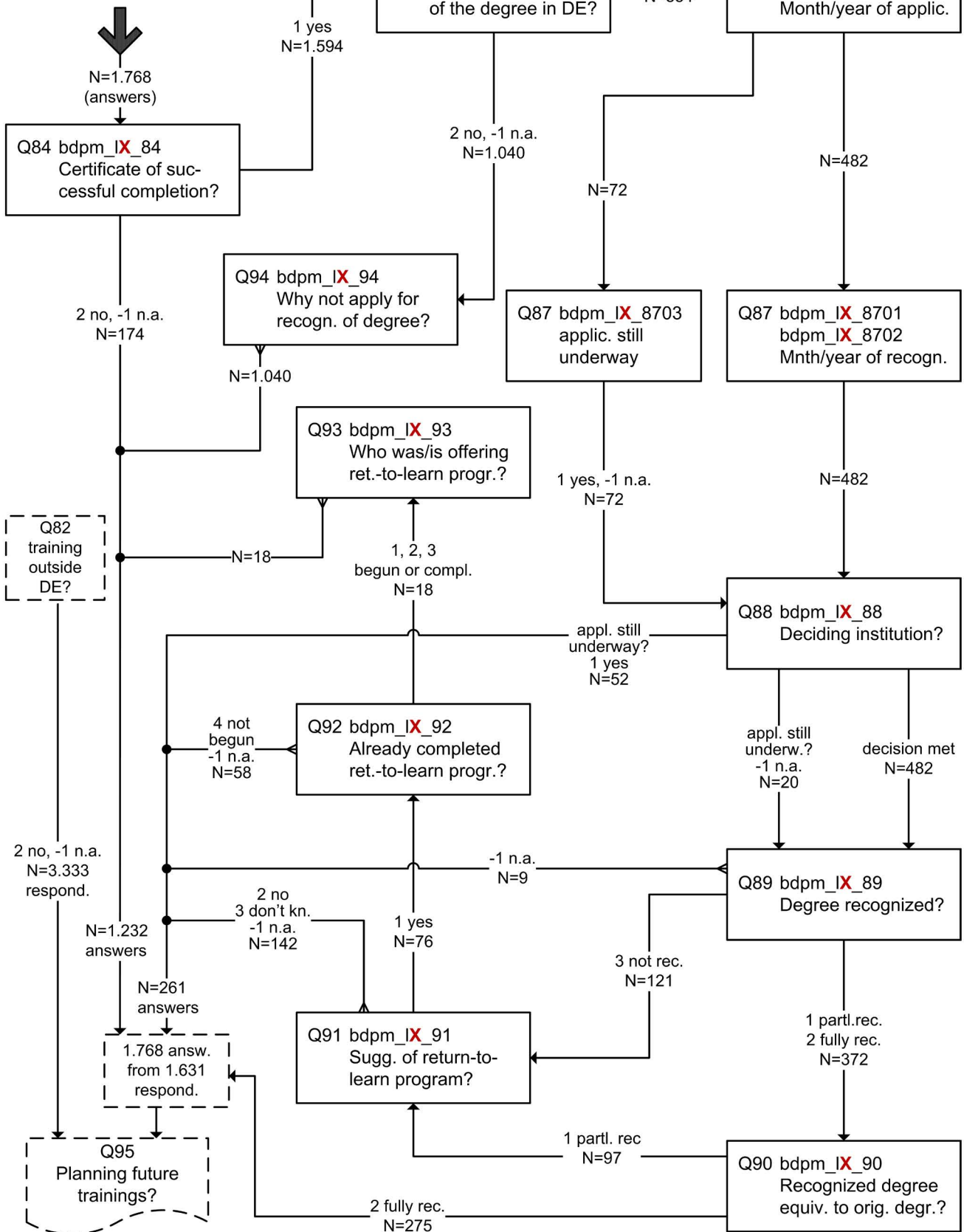
The flow charts for Q84 thru Q94 on the following pages show an "X" in place of the index. The Ns for each variable are indicated as "N(X1)=...", "N(X2)=..." and so on. N(X8) is not displayed, because it is always zero.

Please note: N(Xi) refers to answers, not to cases (respondents).

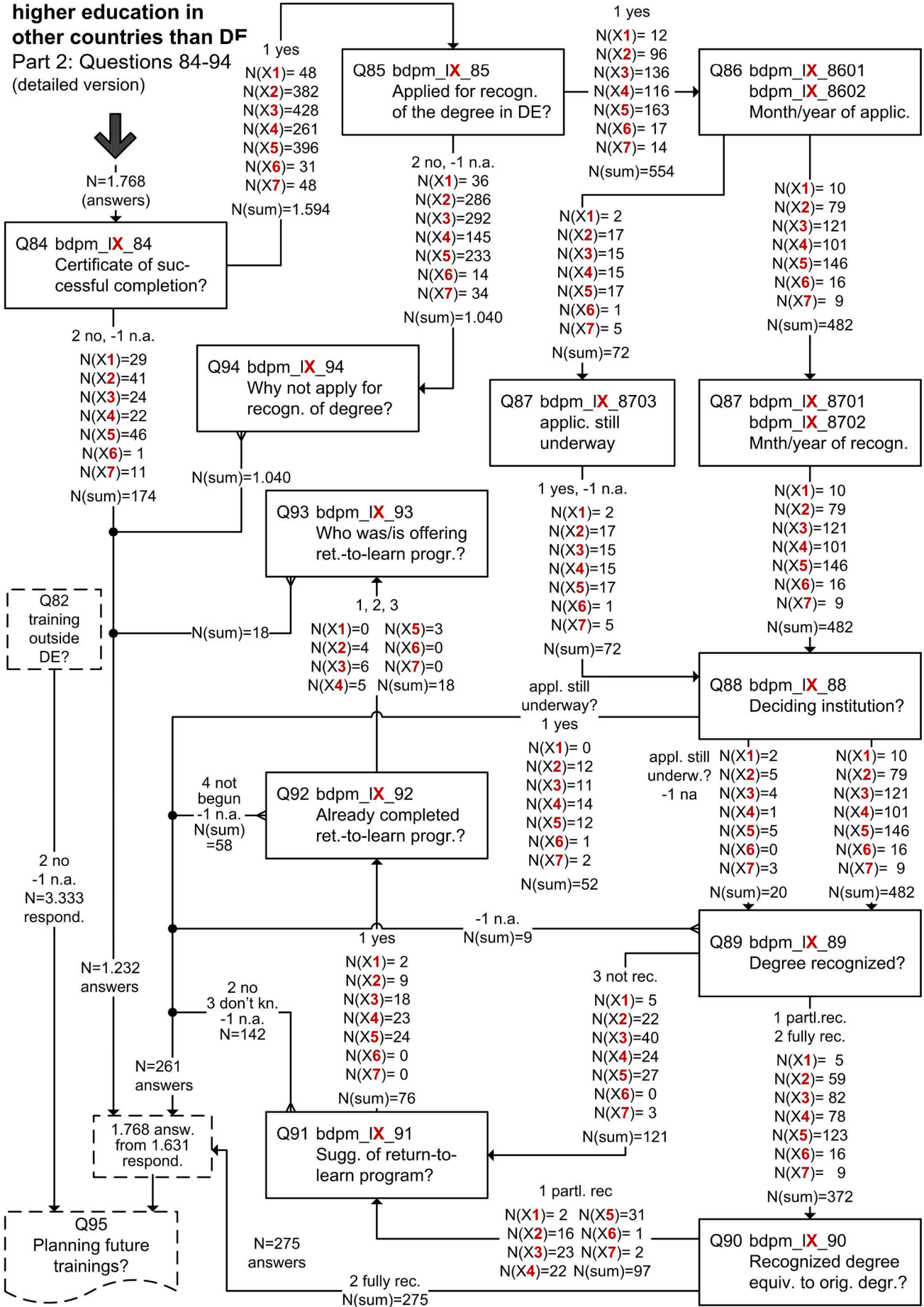
Besides the detailed flow chart for Q84 thru Q94 a summarized flow chart is supplied, showing the sums of N but not the Ns for each kind of training. This version might be useful for an overview of the filtering.



Vocational training or higher education in other countries than DE
Part 2: Questions 84-94
 (summarized version)



Vocational training or higher education in other countries than DF
 Part 2: Questions 84-94
 (detailed version)

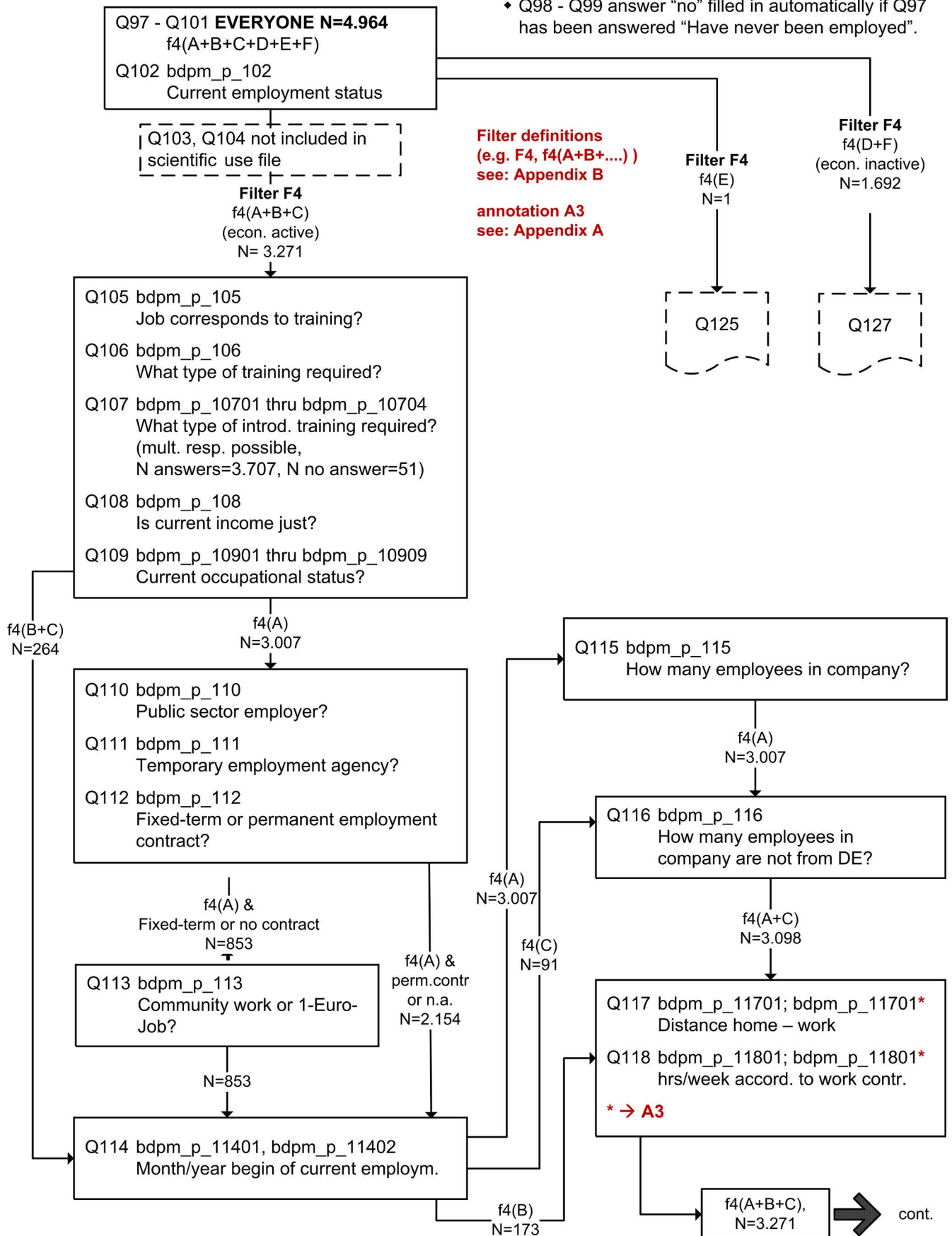


Employment

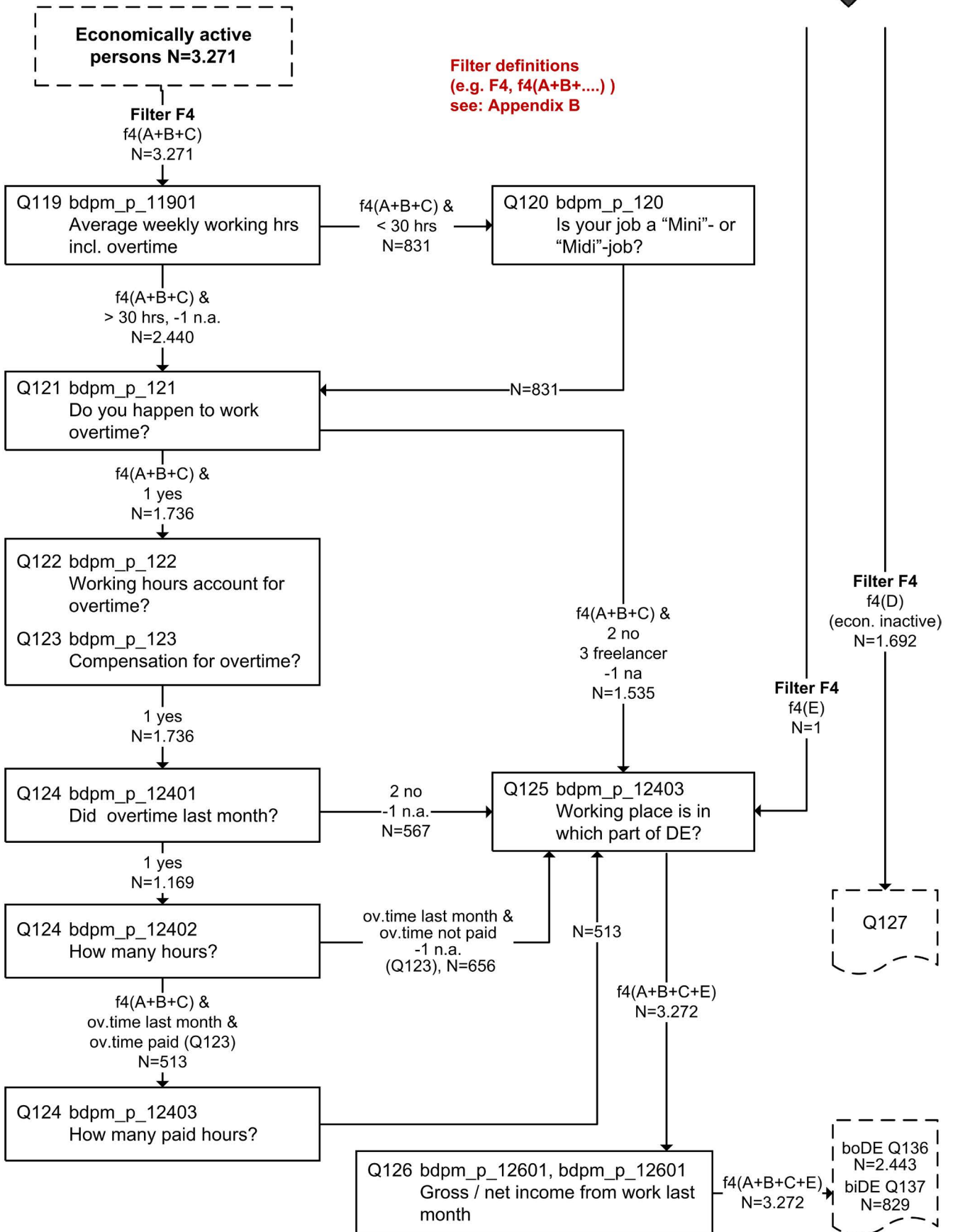
Part 1: Questions 97-118

No filters in Q97 - Q101 (bdpm_l_9701 - bdpm_p_101)

- ♦ bdpm_l_9701, bdpm_l_9702 → **A3**
- ♦ Q98 - Q99 answer "no" filled in automatically if Q97 has been answered "Have never been employed".



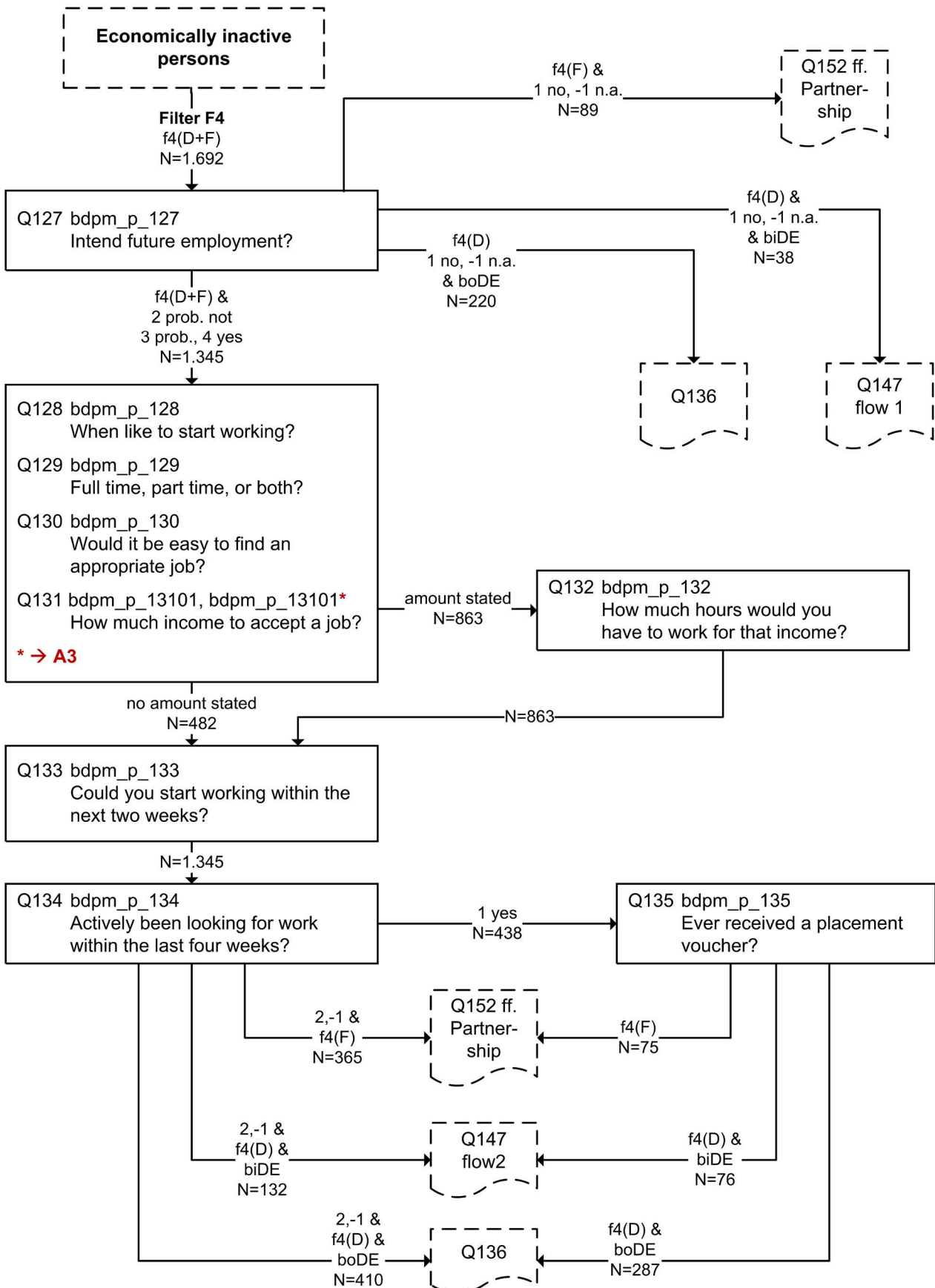
Employment
Part 2: Questions 119-126



Employment
Part 3: Questions 127-135

Filter definitions
(e.g. F4, f4(A+B+....))
see: Appendix B

annotation A3
see: Appendix A

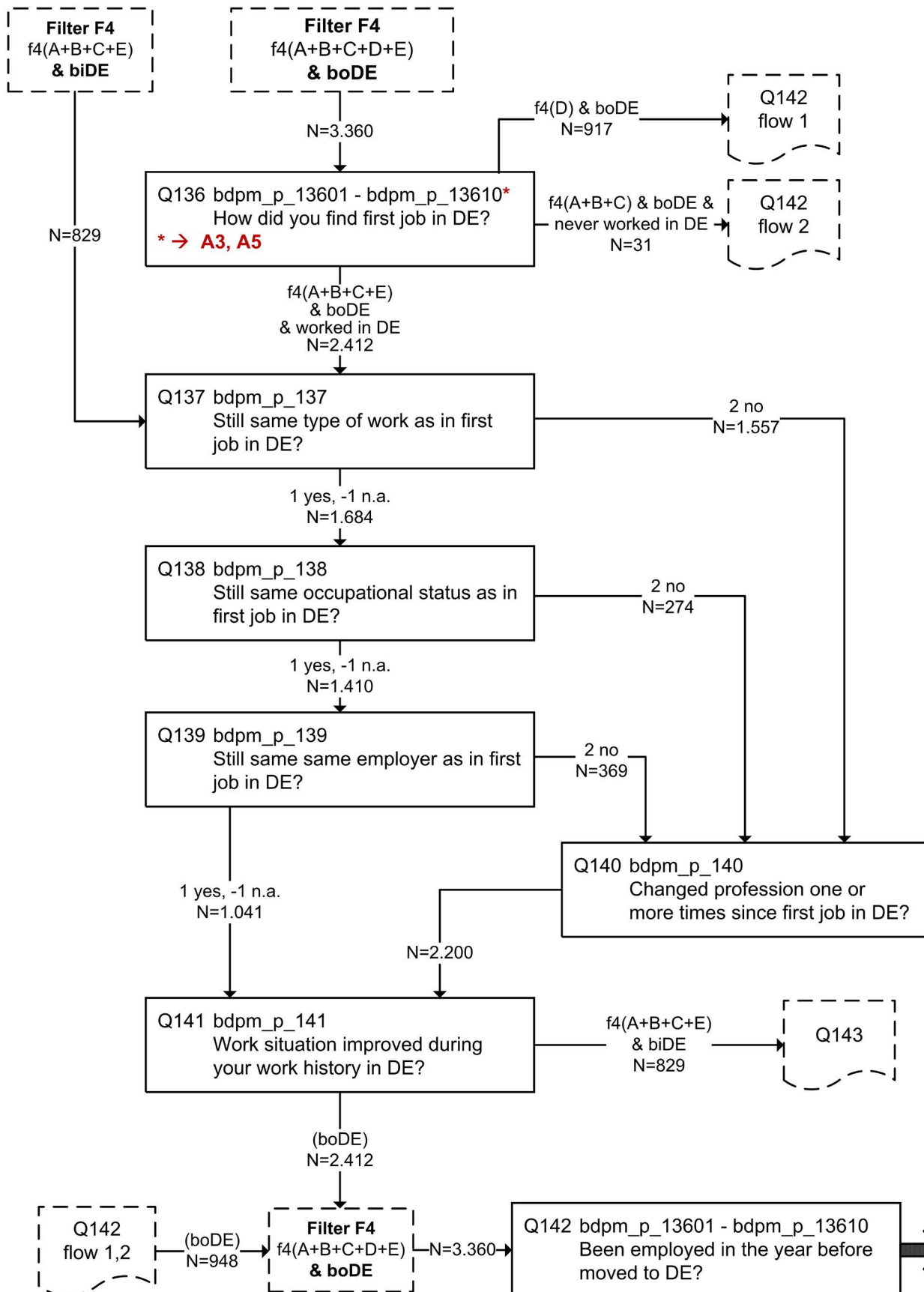


Employment

Part 4: Questions 136-142

Filter definitions
(e.g. F4, f4(A+B+....))
see: Appendix B

annotations A3, A5
see: Appendix A

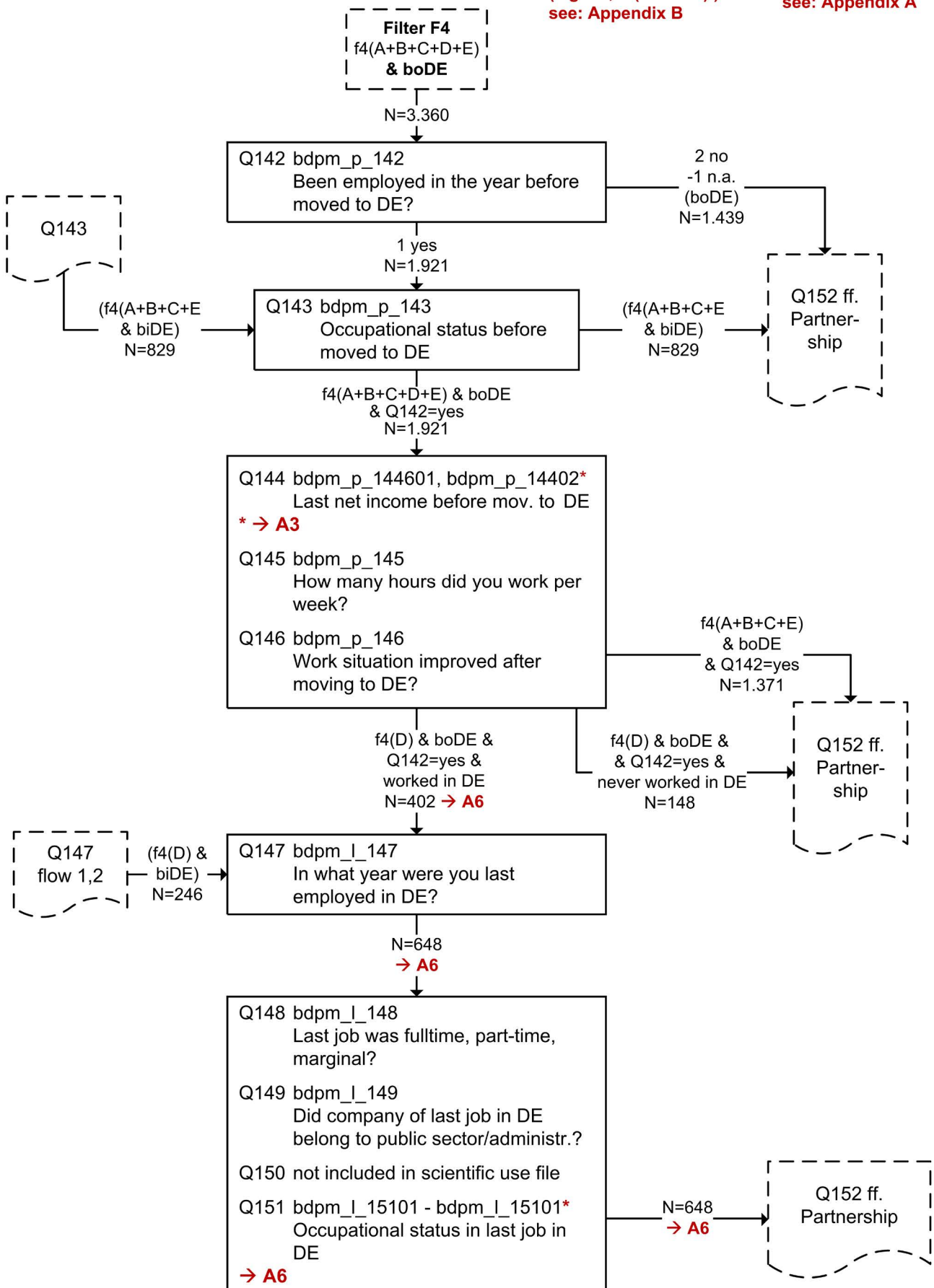


Employment

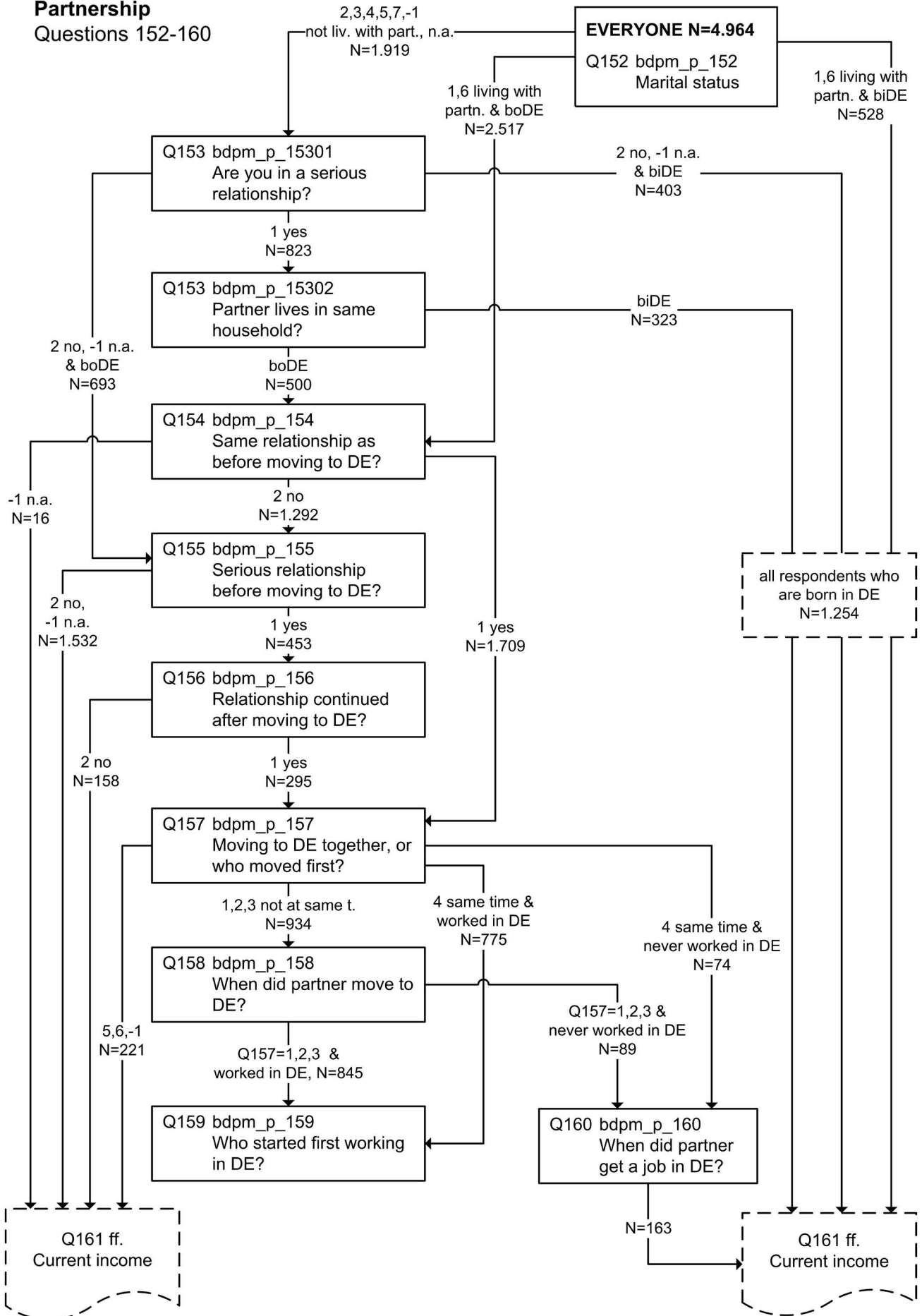
Part 5: Questions 142-151

Filter definitions
(e.g. F4, f4(A+B+....))
see: Appendix B

annotations A3, A6
see: Appendix A

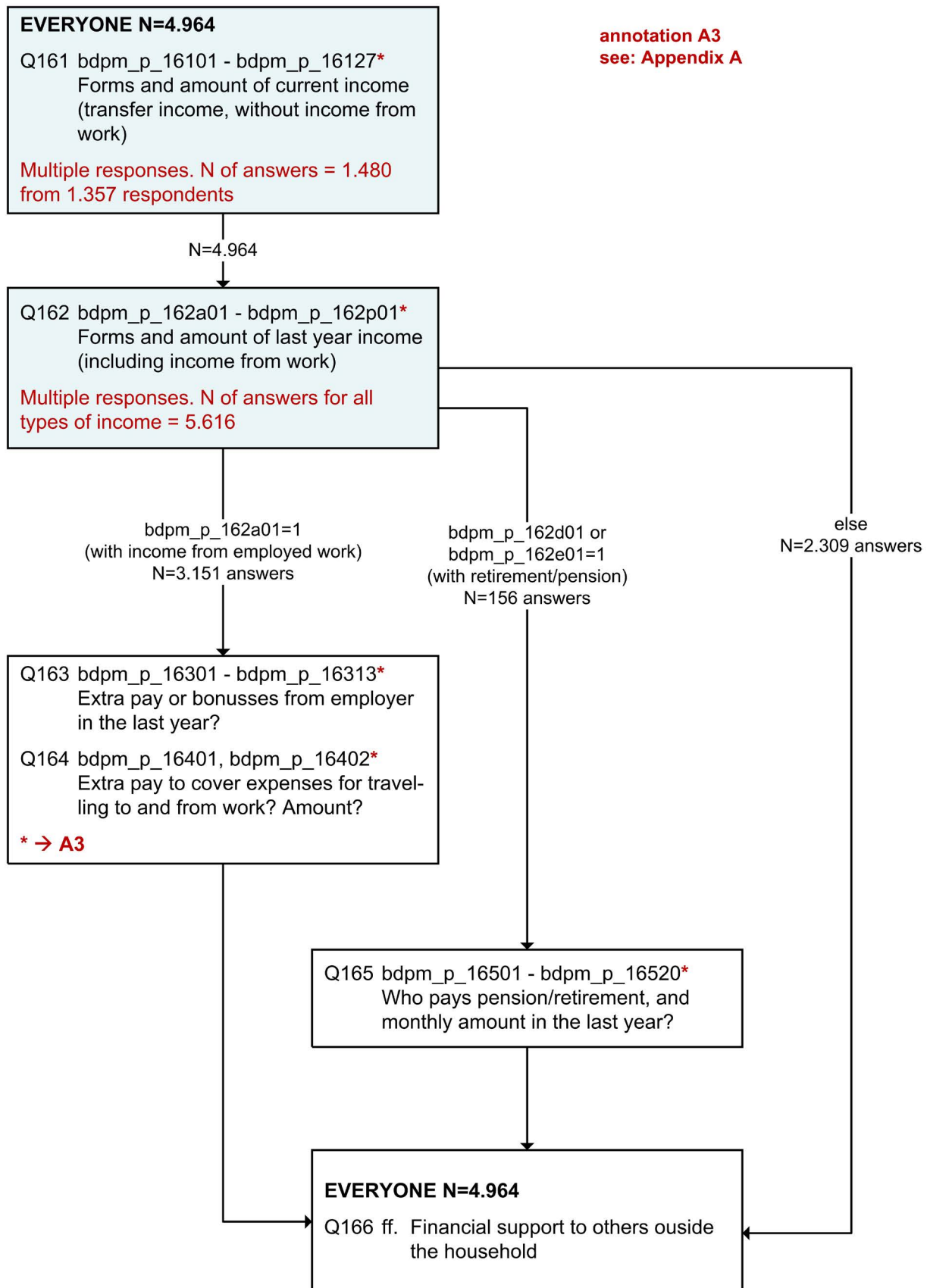


Partnership
Questions 152-160



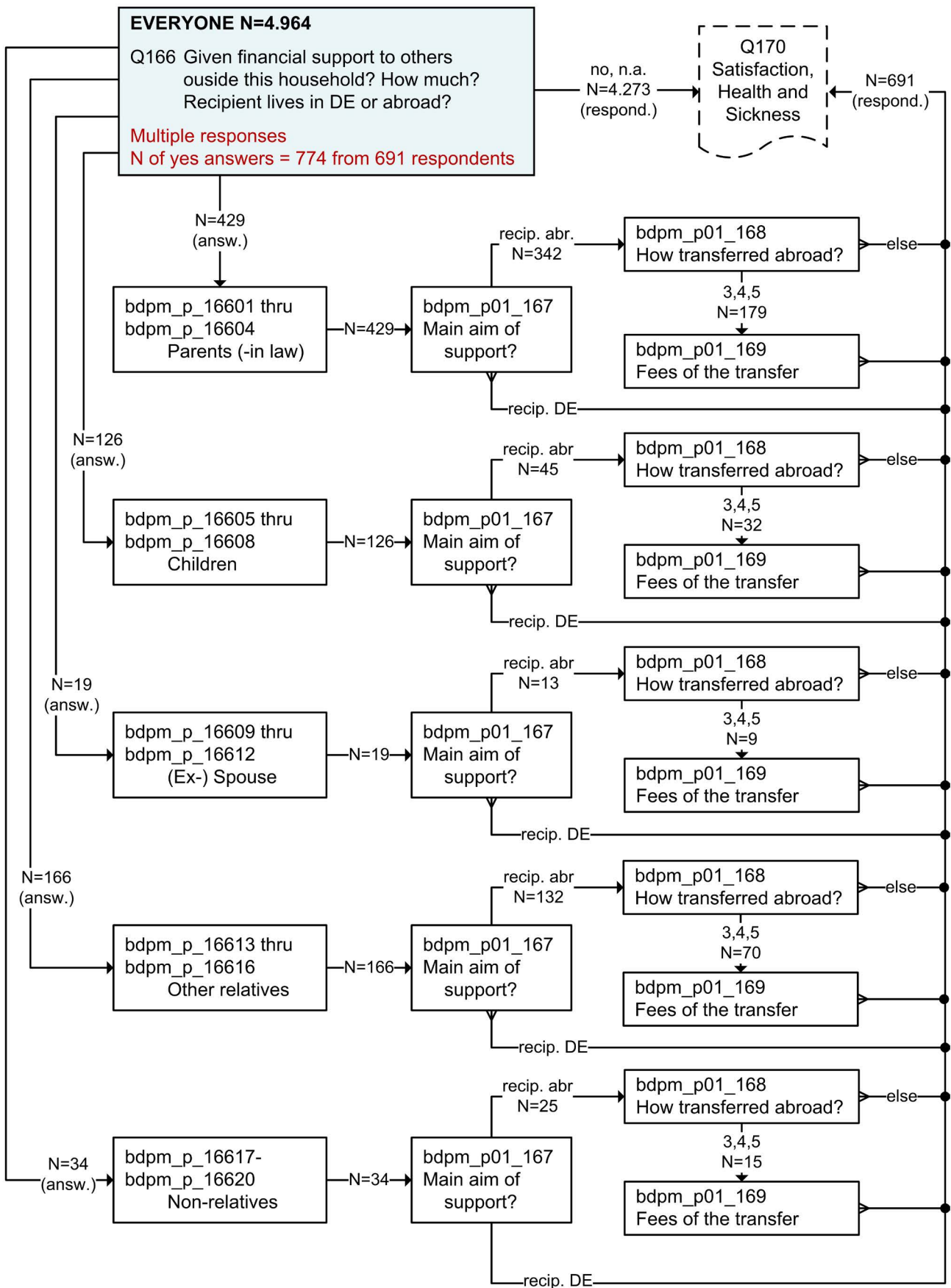
Income

Part 1: Questions 161-165



Income

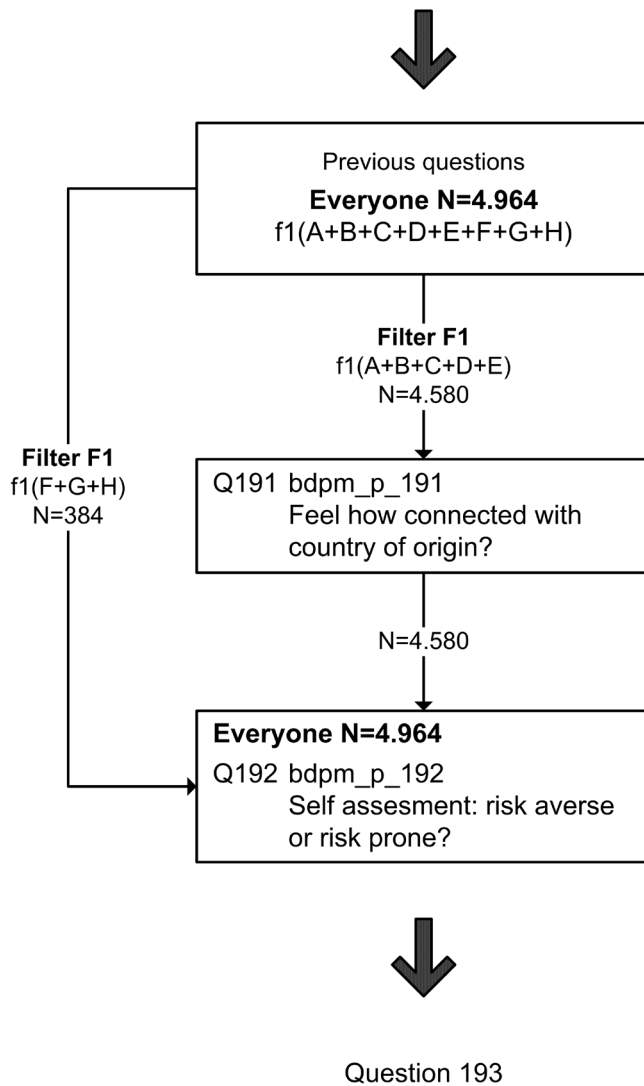
Part 2: Questions 166-169



Satisfaction, health and sickness, and Attitudes and religion

Questions 171-193

filtering is transparent and easy to understand, except for question 191



Appendix A: Annotations

A1 Annotation to the migration biography, section “Coming to Germany”

The variable names contain X as a two-digit marker for the respective loop transit number. With each transit over one of the two interlaced loops (one for stays in DE and one for stays in other countries) X is increased by 1. At any one time the “does not apply” group in the loop variables is composed of a) the 1.254 German-born respondents who skip the whole battery of questions, b) the respondents who already exited the loop, and c) the respondents who passed the loop transit in the other one of the interlaced loops.

The loop is exited once and for all when the answer to Q20 is: “no, have not moved any more, stayed in DE up to now.” The “exit” list in the right lower corner shows how many respondents exited with each loop transit. This ought to add up to 3.710 (all respondents born outside DE), but the sum is only 3.709. This is because the maximum of 15 loop transits provided were not enough for one respondent. In consequence, for this person, it cannot be established which moves happened after the first 15 moves and when the last move to DE took place.

A2 Annotation to the migration biography sections “Coming to Germany” and “Stays abroad”

Synopsis 1 and synopsis 2 of the documentation “How to Generate Spell Data...” (SOEP Survey Paper No. 228) further illustrate the structure of the migration biography data.

Especially for the section “Coming to Germany”, synopsis 2 help to clarify at what point in the loop the marker for the transit number changes. The marker is notated “X” in the variable names in the flowcharts. Please note that in section “Coming to Germany” of the migration biography, the marker is not the same for every variable in the same loop transit.

The structure of the section “Stays abroad” is much easier to understand. The marker in the variable names changes with each loop transit. All variables belonging to the same loop transit have the same marker.

A3 Annotation to value “-2 not applicable”

Sometimes the value “-2” stands not only for “-2 not applicable” but also for “the answer is to be found in the next variable(s)”. Often this is the case if a question asks for the occurrence of something and at the same time for the number or quantity of the

occurrences, while the answers are coded in two or more variables. Examples of this are questions 34 and 37. This is also the case with multiple-response questions.

Normally, the flowcharts explain the number of “-2” values of each variable as the sample size minus the sum of the “N”s of those groups that encounter a variable (i.e., that do not skip a question due to filtering). But if the value “-2” stands for something other than “-2 not applicable (due to filtering)”, this no longer holds. To eliminate the ambiguity of the value -2, one could recode the variables so as to assign semantically appropriate codes to “-2” values that do not stand for “not applicable”.

A4 Annotation to question 52 (bdpm_I_5201 through bdpm_I_5209)

Variables bdpm_I_5201 through bdpm_I_5208 show which person the respondent lived with during the first 15 years of his or her life. The values add up to 15 for each respondent except for respondents who refused to answer. The variables take on the value “-2” if the respondent spent the 15 years with other persons than the person that the specific variable represents.

If a respondent did not answer question 52 at all, variable bdpm_I_5209 takes on the value “-1” and otherwise “-2”.

A5 Annotation to question 136 (bdpm_p_13601 through bdpm_p_12610)

The question allows for multiple answers.

Some of the respondents have the value “-2 does not apply” in all variables bdpm_p_13601 through bdpm_p_13610, although they satisfy the filter conditions for question 136. They have probably been asked question 136, but none of the categories applied. The instrument ought to have provided the category “other” to capture such cases and make them distinguishable from the respondents who really skipped that question.

A6 Annotation to question 147 through 151

The one additional case in the data compared to the flow that leads to question 147 (and following) is caused by a very unusual combination of values for one person, which was not captured by the filter conditions. This one respondent had been led to question 147 although the logic of the filter would have required him to skip the question. As the flowcharts have to show flows and filter conditions at the same time, it seemed preferable to specify the logical frequencies instead of the true frequencies for the remaining questions in the employment module.

Appendix B: Filter Definitions - Description of Complex Filters

Filter F1 in module “Cultural ties, language, and social community”

**F1 [Q2=2 and (Q9=2,3,KA or Q18=2 or Q25=2)] or [Q6=2 or Q7=1]
Included: f1(A+B+C+D+E), excluded: f1(F+G+H)**

- f1A Group A: not German-born & no German citizenship or German citizenship not since birth, respectively. (N=3.476)
- f1B Group B: not German-born & German citizenship since birth & Aussiedler (ethnic German from Eastern European Country) (N=159)
- f1C Group C: German-born & parents not German-born, but not World War II refugees (N=929)
- f1D Group D: German-born & parents German-born & grandparents from IT, SP, GR, former YUG, TR (N=14)
- f1E Group E: German-born & parents Aussiedler & grandparents from IT, SP, GR, former YUG, TR (N=2)
- f1F Group F: not German-born & German citizenship since birth & not Aussiedler (N=75)
- f1G Group G: German-born & parents German-born or no answer & grandparents not from IT, SP, GR, former YUG, TR or no answer (N=295)
- f1H Group H: German-born & parents World War II refugees & grandparents not from IT, SP, GR, former YUG, TR (N=14)

Filter F2, questions Q61 and Q62 in module “Childhood and parental home”, path “father not living in household”

F2 [Q58=3,4 and Q2=1 and Q6=2]

- f2A German-born & parents not World War II refugees & father not living in household (N=639)
- f2B Not German-born & father not living in household (N=3.307)
- f2C German-born & father not living in household & parents are World War II refugees (N=15)
- f2D German-born & father not living in household & (both parents born in DE or no answer if both parents are born in DE (N=299)
- f2E Father living in household (N=704)

Filter F3, questions Q61 and Q62 in module “Childhood and parental home”, path “mother not living in household”

F3 [Q58=2,4 and Q2=1 and Q6=2]

- f3A German-born & parents not World War II refugees & mother not living in household (N=556)
- f3B Not German-born & mother not living in household (N=3.196)
- f3C German-born & mother not living in household & parents are World War II refugees (N=16)
- f3D German-born & mother not living in household & (both parents born in DE or no answer if both parents are born in DE (N=297)
- f3E Mother living in household (N=899)

Filter F4 in module “Employment”

**F4 economically active / inactive (according to Q102, Q109, Q97).
Economically active: all kinds of employment (Q102=1,2,3,4,8). Economically inactive: Partial retirement with zero working hours, voluntary military service, volunteer service, not employed (Q102=5,6,7,9).**

- f4A Group A: Economically active persons (Q102=1,2,3,4,8), without freelancers, self-employed persons, and family members working for self-employed relative (according to Q109) (N=3.007).
- f4B Group B: Freelancers, self-employed persons without employees, family members working for self-employed relative (according to Q109) (N=173).
- f4C Group C: Freelancers, self-employed persons with employees (according to Q109) (N=91)
- f4D Group D: Economically inactive persons without persons in partial retirement, phase with zero working hours and without persons who have never been employed (Q102=6,7,9 & Q97 not 1) (N=1.163)
- f4E Group E: Persons in partial retirement, phase with zero working hours (Q102=5) (N=1)
- f4F Group F: Economically inactive persons who have never been employed (Q102=6,7,9 & Q97= 1) (N=529)

Please contact the author if you have questions and suggestions, or if you encounter any errors in the flowcharts.