HUNGARIAN **DEMOGRAPHIC** RESEARCH INSTITUTE

WORKING PAPERS

ON POPULATION, FAMILY AND WELFARE

№ 38

GROWING UP IN HUNGARY
COHORT '18 HUNGARIAN BIRTH COHORT STUDY
TECHNICAL REPORT 2. PRENATAL WAVE

Edited by
Laura Szabó

demografia.hu

WORKING PAPERS

ON POPULATION, FAMILY AND WELFARE

№ 38

GROWING UP IN HUNGARY
COHORT '18 HUNGARIAN BIRTH COHORT STUDY
TECHNICAL REPORT 2.
PRENATAL WAVE

Edited by
Laura Szabó

Hungarian Demographic Research Institute
2021

ISSN 1588-3396
ISBN 978-963-959-760-0

Director: Zsolt Spéder
Series editor: Róbert Iván Gál

Suggested citation:
Laura Szabó, Julianna Boros, Ildikó Fábián, Gabriella Gresits, Anna Hortobágyi, Balázs Kapitány, Réka Kis, Krisztina Kopcsó, Fruzsina Leitheiser, Adél Rohr, Zsolt Spéder and Zsuzsanna Veroszta (2021):
Growing up in Hungary - Cohort '18 Hungarian Birth Cohort Study. Technical report 2. Prenatal wave.
Working Papers on Population, Family and Welfare, No. 38
Hungarian Demographic Research Institute, Budapest.

Address: Hungarian Demographic Research Institute
Budapest, Buday László utca 1–3. 1024 Hungary
e-mail: speder@demografia.hu

HUNGARIAN **DEMOGRAPHIC**
RESEARCH INSTITUTE

The research is financed within the framework of the EFOP 1.9.4. – VEKOP-16 invitation (renewing methodology and informatics in the social sector) issued by the Hungarian Ministry of Human Capacities.

ABSTRACT

This volume presents the results of the Cohort '18 Hungarian Birth Cohort Study as the third output in the research series. The first contained the theoretical, methodological and organizational tasks preceding the prenatal data collection wave. The second volume presented the theoretical background and the conceptualization of the Cohort '18 Hungarian Birth Cohort Study. This third volume describes the methodology of the prenatal data collection wave and the technical background to the surveying and data processing. We present the sampling procedure, sample coverage and reliability of the raw data. This is followed by a summary and description of the documents and questionnaires used in the fieldwork, as well the quality control procedures. We also present the data recording, editing and cleaning process, and review the content of the different databases. Finally, we summarize the most important statistics of the fieldwork and the metadata of the survey and conclude with reviewing the research ethics guidelines.

Keywords: Cohort '18, Hungarian Birth Cohort, longitudinal, methodology, sampling, weighting, data quality, technical report

AUTHORS

Laura Szabó
Julianna Boros
Ildikó Fábián
Gabriella Gresits
Anna Hortobágyi,
Balázs Kapitány
Réka Kis
Krisztina Kopcsó
Fruzsina Leitheiser
Adél Rohr
Zsolt Spéder
Zsuzsanna Veroszta

ACKNOWLEDGEMENT

We are grateful to Lukács Heynes, the English translator of our original Hungarian report and to Clive Liddiard, the English language proof-reader.

DOI: 10.21543/WP.2021.38

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The Cohort '18 Hungarian Birth Cohort Study is a longitudinal research project initiated by the Hungarian Demographic Research Institute. It aims to examine the growth of nearly 9,000 children born in 2018–2019, from foetal age to 3 years of age, through several data collection waves.[1]

Considered at a European level, the Cohort '18 Study is also a major research initiative in terms of scale, methodology and wide applicability. The Health and Demographic Study on Pregnant Women and Infants, launched in 1979, is an antecedent study in Hungary; this was followed by the National Longitudinal Child Growth Study (Joubert and Gárdos, 1991) and the Turning Points of Life Course study.[2] The first of these also provided a precedent for the main features of current research regarding pregnancy-based data collection. In addition to the Hungarian traditions, our study also utilized the lessons learned and the results of foreign longitudinal birth cohort studies (for a comprehensive description of these, see Blaskó, 2009 and Veroszta, 2018a; 2019).

This volume presents the results of the Cohort '18 Hungarian Birth Cohort Study as the third output in the research series. The first contained the theoretical, methodological and organizational tasks preceding the prenatal data collection wave (Veroszta, 2018a). We also reviewed the professional background and organizational work on which the study concept as a whole – and our future data collections – are based.

The second volume presented the theoretical background and the conceptualization of the Cohort '18 Hungarian Birth Cohort Study (Veroszta, 2019). Considering the interdisciplinary approach of our research, in that volume we presented the three main research areas of our longitudinal study (demographic, health-psychological and socio-economic), based on comprehensive theoretical frameworks such as the ecological model of child development, life-course research or inheritance (and regeneration) of social inequality between generations.

In this third volume, we present the methodology of the first wave (i.e. the prenatal data collection wave) and the technical background to the surveying and data processing. After a brief description of the Cohort '18 Study, we present the sampling procedure, sample coverage and reliability of the raw data. This is followed by a summary and description of the documents and questionnaires used in the fieldwork. In a separate section, we present the tools and procedures used throughout the research to ensure the quality and control of the data collected and the techniques we used to prevent respondent drop-out. Then we describe the data recording, editing and cleaning process, and review the content of the different databases created in each phase of the prenatal wave, their editing and merging. In Section 6, we summarize the most important statistics of the fieldwork and the metadata of the survey. Finally, we conclude our study by reviewing the research ethics guidelines and by presenting the legal framework and data protection guarantees we used.

1.1. METHODOLOGICAL FRAMEWORK OF THE STUDY

The collection of data examining the growth of children born in 2018–2019 began at foetal age, with an inquiry involving pregnant women in January 2018. The initial sample of nearly 9,000 children is to be repeatedly interviewed when the cohort babies are aged 1 year, 18 months and 3 years. The aim of the Hungarian Demographic Research

1 The study is funded through the EFOP 1.9.4. - VEKOP-16 EMMI (Ministry of Human Resources) call (Renewal of Methodological and Information Systems in the Social Sector).

2 The Hungarian Generation and Gender Survey – Turning Points of Life Course 2001–2016, HDRI, https://www.demografia.hu/en/ggp

Institute of the Hungarian Central Statistical Office (HDRI HCSO) is to follow and monitor the lives of the children until they are grown up, thus providing a comprehensive picture of the growth of children living in Hungary.

The most important advantage of longitudinal data collection (compared to cross-sectional or repeated cross-sectional surveys) is that we can measure changes at the individual level: what factors have influenced the development of the foetus, and later the behaviour, performance and well-being of the young child. Since the research already started at foetal age, we can take into account not only those factors that influence a child's development after birth, but also factors that occurred and influenced the foetal life (through the experiences of the pregnant woman and later, when she is already a mother). To achieve our goal, we first collected data using a nationally representative sample. Our initially planned sample size included pregnant women whose due date was between 1 April 2018 and 30 April 2019 and who were 28–31 weeks pregnant at the time of the first inquiry, which was performed by health visitors (within the selected health visitor districts). As the Hungarian network of health visitors is unique – in that it covers the whole country and comes into contact with almost 97% of pregnant women – we relied on this network during the first two waves of our data collection. That is a worthwhile strategy, because both during pregnancy and when the child is of toddler age, the women (mothers) are in frequent contact with the health visitors. At later stages of the data collection, when the child is 18 months or 3 years old, the data collection is no longer carried out by health visitors, but by social and market research professionals.

The preparation of the study began back in 2017, and the first data collection started on 1 January 2018 and ended in early April 2019. The data collection involving 6-month-olds started in July 2018, and with 18-month-olds in August 2019. In this volume, we present the technical report of the first wave of data collection – i.e. the prenatal data collection wave.

1.2. THEORETICAL BACKGROUND OF THE COHORT '18 STUDY

The Hungarian Birth Cohort Study takes an interdisciplinary approach: it aims to answer research questions related to psychology (developmental psychology), health sciences, sociology, demography and economics. It is therefore necessary for the research programme to draw on all of the above disciplinary traditions, as this allows us to answer research questions related to child development within different disciplines. Each of the disciplinary ideas and theories (and the research questions arising from them) is discussed in detail in the volume presenting the research concept (Veroszta, 2019), and so here we only provide a brief summary of them.

In the conceptualization of the Hungarian Birth Cohort Study, special emphasis is placed on the life-course perspective, the system of social inequalities, and social policy interventions.

The perspective of life-course studies is particularly useful for interpreting a child's development and family circumstances, as it provides a suitable and beneficial framework for interpreting child development in dynamic environments, in the midst of changing family relationships (Elder et al., 2003; Settersten, 2003; Amato, 2000; Mayer, 2004; McLeod and Almazan, 2003). From this perspective, the family circumstances of child development are embodied in the linked parental life courses ('linked lives'; see Greenfield and Marks, 2006). As the mother and father live together continuously with the child, all their lives are interconnected across several aspects – family life, parents' partnership status, parents' labour market careers, etc. -; these interconnections create the framework (or living space) for childbearing and childrearing.

At the same time, we hypothesize that parental influences, parental resources and parenting styles change over time and may vary at each developmental stage, allowing us to capture the extent to which parents play a role in the performance of a child and young adult (Amato, 2000). The focus on the intergenerational perspective also allows for a new kind of inquiry into the fundamental issues of demography and sociology, the interpretation of adulthood, family formation (Barber et al., 2002; Buchmann and Kriesi, 2011) and the classical issues concerning social mobility studies (Bukodi et al., 2015).

The decisive role of parental influences in a child's growth and development is beyond question; but it is undeniable that the child's genetic codes and temperament are also given characteristics; furthermore, that the child's development takes place within a society with a specific social and institutional structure. Our research programme, while not explicitly promoting medical approaches, provides an opportunity to analyse health and health-related topics, as these questions are relevant to both psychological and demographic approaches (Mills and Tropf, 2015).

Although the Cohort '18 Study does not explicitly seek to describe the national social and institutional structures, it inevitably takes them into account. By social structure, we mean both parental resources (income situation, economic capital, cultural capital, social capital) and the child's (later) entry into and participation in the institutional system, and his/her success within these institutions. Thus, our research question also focuses on the extent to which the welfare system in Hungary is able to mitigate inequalities caused by differences in parental resources.

Nor is the upbringing of children independent of currently prevailing cultural patterns – such as the dynamics, values and ideas of predominant parenting and family relationships (Hagestad and Neugarten, 1985; Buchmann, 1989). Even if only to a limited extent, cohort studies accept the hypothesis that parents' ideals and views about family, childbearing and parenting can be translated into parental practices that greatly influence a child's development. Furthermore, we have good reason to assume that parents' ideals and views, cultural resources and level of demand create a continuously present cultural climate that can influence children's progress and performance within the school system. This can also later be traced in the transition from school to labour market and in the family-formation practices of young adults (Barber et al., 2002).

We compiled our research questions taking all these theoretical considerations into account. They centre around issues related to demographic research, health and development, as well as social background.

Demographic research questions examine the conditions of childbearing and childrearing within key demographic groups, as well as the relationships between childbearing and partnership stability, and the factors that predict, motivate or hinder the fulfilment of future childbearing (to produce siblings). The topic of health and development raises issues such as the impact of expectant mothers' health status and health behaviours on childbirth/health/ foetal and child development; factors influencing preterm birth, low birth weight and reproductive health; and the psychological well-being of the mother and the subjective characteristics of her relationship. And it involves analysis of parental behaviour and its implications for the child's development and health. The third research unit of the study explores the impact of socio-economic background on child development. This unit examines issues such as inequalities deriving from birth; the success of life-course planning and realization; the employment of women with small children; and access to (and use of) the healthcare system and social benefits.

2. SAMPLE, COVERAGE AND WEIGHTING

The aim of this section is to briefly present the reference population of the prenatal wave of the Cohort '18 Hungarian Birth Cohort Study, the practical implementation of the sampling, the theory and practice of the weighting. The text is a logical continuation of the study published in the volume that presented the preparatory phase of the study (Kapitány, 2018), in which the reference population of the study and the sampling plan were presented. Those issues that were discussed in detail in the previous volume and that will help in understanding this text are briefly described here as well.

2.1. REFERENCE POPULATION AND SAMPLING

2.1.1. Population and coverage

In the sampling plan, the population of the study was based on the 7-month-old foetuses registered in the Hungarian prenatal care system, where the due date of delivery was between 1 April 2018 and 30 April 2019. Due to the very high coverage of the Hungarian prenatal care system and the relatively low rate of late foetal mortality, this sample population covers the group of children born in Hungary during this period quite well.

The planned timeframe for interviewing the expectant mothers was during weeks 28–31 of pregnancy, although the health visitor could deviate from this in certain cases (see below). However, the biggest coverage challenge in terms of fieldwork organization was that, although some women appeared in the health visitor system and agreed to participate in the research, for one reason or another the interview did not take place prior to delivery. This could be due to, for example, preterm birth, prenatal hospitalization, travel/stay abroad or even illness on the part of the health visitor. As this is a significant subset, in these cases we provided an opportunity for such foetuses (and their mothers) who otherwise belonged to the sample to join the research at a later date, by getting the mothers to fill in a supplementary (proxy) questionnaire at the time of the survey when the baby was 6 months old. In all, 394 proxy questionnaires were completed during the data collection phase at 6 months, covering 399 children. The questions referred back to the seventh month of pregnancy: 383 mothers were able to answer the pregnancy-related questions retrospectively; however, 11 respondents could not answer the pregnancy-related questions, as they were not the biological mothers of the children, but rather their foster (or adoptive) parents. Respondents involved at the 6-month phase became full members of the study from that point onwards in subsequent surveys. However, as they did not participate in the prenatal data collection, their responses were not included in this current report.

A problem of coverage also arose in connection with respondents (16 in total) who could not complete the survey in Hungarian, due to language barriers (and for whom an interpreter was not available). These women completed a significantly shorter, foreign-language (English, German, Chinese or Vietnamese) self-administered questionnaire, the data of which will be analysed later (i.e. not in this report).

2.1.2. The sampling procedure

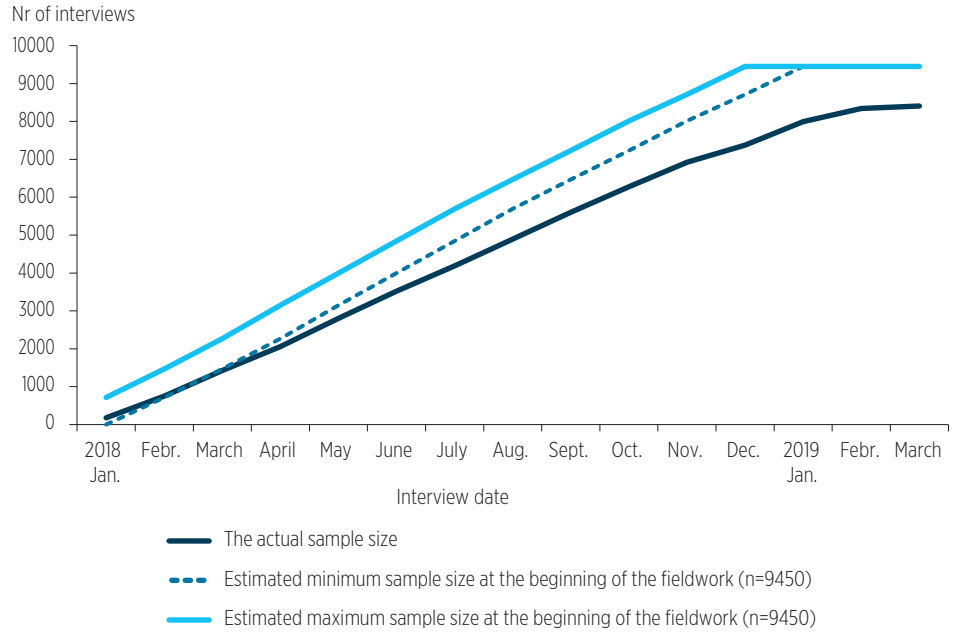
We applied a complex multi-stage sampling procedure, where the primary sampling unit was the local health visitor district. The sample selection was based on the expected number of births in the district, the geographical location and social status of the given district, and the estimated willingness to respond. For organizational and financial reasons, the districts selected were geographically concentrated (Kapitány, 2018). In

those cases where the health visitors refused to participate in the study (61 out of 628), we tried to replace the districts with health visitor districts of similar social characteristics. In some cases, we could not find a suitable replacement; and in one case, all the selected health visitor districts from a particular Budapest administrative area refused at the last minute to participate in the survey. Thus, the study fieldwork began in January 2018 in 608 health visitor districts. Within the selected districts, all women with an expected due date between 1 April 2018 and 30 April 2019 were invited to participate in the survey.

2.1.3. The initial sample and its adjustment during the fieldwork

When designing the study, our original goal was to have a sample size of foetuses (together with those who joined later through the proxy questionnaire at the 6-month stage) of 10,000. This sample design was based on live birth register data from previous years. It became apparent around the time the study was launched that, due to a 3.5% drop in birth rates between 2016 and 2018, we had to count on a lower number of respondents (about 350 fewer). As all the sampled health visitor districts from one of the capital city's areas refused to participate in the survey shortly before the start of the fieldwork, and since they could not be replaced at such short notice, this resulted in a further fall in the number of respondents (of 200). Given all these circumstances, it was already apparent at the start of the fieldwork that rather than the planned 10,000, only 9,450 foetuses could be included in the study.

However, the response rates measured in the first months of the fieldwork projected an even smaller number of cases for the prenatal wave. As the health visitors had a longer time span for the data collection, it was only possible to figure out a 'from-to' interval for the (estimated) number of interviews expected in a given timeframe. By March, we could already discern a clear lag; and by the first few months of the fieldwork (May 2018) there was a 20–25% shortfall, compared to the target sample of 10,000 (*Figure 1*).



Source: Cohort '18 Study prenatal wave database (children), N=8409.

Figure 1
The actual and the estimated fieldwork progress

On the one hand, we saw an average lag of 10% compared to the plan, due to a lower-than-expected willingness to respond. On the other hand, this lag proved more severe in some geographical regions: in Budapest, while willingness to respond generally met our expectations, the number of respondents was lower than expected due to the afore-mentioned refusal by health visitors to participate. In some parts of the Budapest metropolitan area (to the east of the capital), however, willingness to respond was much lower than we expected. We can assume that overload of the health visitor system lay behind this (the number of families with small children in these areas has soared due to the suburbanization process, so that the health visitor network could not keep pace). In addition, performance in some rural parts of Hungary fell short of expectations, typically due to some local, non-generalizable reason (e.g. organizational difficulties).

Because we had register data for new-borns from previous years, we could also monitor the development of the sample from a socio-demographic perspective. The results of the first months showed a worrying discrepancy: the rate of pregnant women with low levels of education (i.e. who had completed at most eight years of primary school) was significantly lower in our sample than in the reference population (15% vs. 20%). Other minor deviations (e.g. the proportion of economically inactive pregnant women was slightly lower than expected, while the proportion of respondents over the age of 35 was higher) are presumably also due to this educational bias (i.e. the small share of the poorly educated). According to feedback from the health visitors, willingness to respond was not necessarily lower within this group; rather, the difficulties associated with the fieldwork were more significant.

To mitigate the biases that occurred, the following changes were made from June 2018 onwards. (1) The data collection period was extended by a month (from 12 to 13 months). Thus, health visitors also now included in the study those women whose due date fell in April 2019. With this fieldwork procedure, the population of the study also changed slightly: foetuses due to be born between 1 April 2018 and 30 April 2019 were included in the sample. (2) Additional health visitor districts were selected in the sample, so that work could proceed using the initially planned 628 districts: in the administrative districts of the Pest part of the capital city and in the unexpectedly poorly performing parts of the Budapest metropolitan area.[3] When selecting the health visitor districts, we made sure that their social composition corresponded to that of the original sample districts, and that the geographical distance was kept to a minimum. (3) We significantly increased the fee for interviewing women who had completed at most eight years of primary school.

After sample adjustment, the proportion of expectant mothers with the lowest educational attainment increased from 15.2% to 16.9% (although that is still lower than their ratio in the population). Despite the sample correction carried out in the meantime and the downwardly modified targets (9,650 foetuses to be included in the study), the targeted sample size for the research could not be met, although the overall shortfall was not significant.

2.1.4. Estimating willingness to respond

When planning the sampling, willingness to respond was estimated at between 62% and 80%, depending on the characteristics of the health visitor district. We assumed that 14,265 women would be invited to respond, and that ultimately 10,000 successful interviews would be conducted, with a mean response rate of 70.1%.

At the time of writing, the actual response rate is not yet known: it will only be calculated on the basis of data from the 2019 yearly health visitor registers. Once these

3 A total of 20: 13 of them in Budapest, 4 in the Gödöllő subregion, 2 in the Monor subregion, 1 in the Gyál subregion.

have been handed over and edited, we can report the actual number of pregnant women (and those expecting twins) who were cared for in any given health visitor district during the sampling period. This will be comparable to the total number of successful interviews, pregnancy-proxy interviews and foreign-language interviews. However, even when calculating these, there will be additional difficulties in managing those areas where there was a change of health visitor during fieldwork and where the new health visitor did not agree to participate in the survey. In any case, according to preliminary estimates based on health visitor reports submitted to us, the actual response rate may be around 61–63% (approximately 14,000–14,500 visits, of which 8,813 foetuses became involved in the study, thus being the 'total involved population').

2.2. WEIGHTING

2.2.1. Purpose of weighting

The fundamental purpose of weighting cross-sectional samples is to eliminate minor biases in the sample (due to sampling errors, differences in the response rate or other non-sampling errors), by matching the survey data to the reference population data as far as possible (Solon, et al. 2015; Rudas, 2006). In practice, this means that answers from respondents with a lower representation in the sample, compared to the population as a whole, are given greater weight, so that they better represent the actual target population. In this way, we modify the raw data, so that certain distributions in the weighted sample are the same as the distributions of the target population, the characteristics of which are known from external sources (register or census data). Another important function of weighting a cross-sectional sample is to counterbalance the effect of sampling design (so-called design weights); this was not necessary here, since the sample was not geographically or socially distorted.[4]

Several methods are accepted for the practical implementation of weighting, for example classical matrix weighting (Szelényi, 2003; Gabler, 1994) or complex calibration (Horváth and Mihályffy, 2008). In the case of the prenatal data collection wave of the Cohort '18 Study, a simple, conventional (classical) matrix weighting was applied.

In the case of weighting the current wave of the cohort study, perhaps the most important question is where the population data that can serve as a reference for weighting the database come from. In this case, it seems logical that register data reflecting all births should serve as the reference data. Many such registers are available in Hungary (address register, social security register, etc.) with different data content, availability and accuracy. After reviewing the advantages and disadvantages of each register database, the Vital Statistics Register (VSR) of the HCSO was chosen for our reference data.[5] The data content of this register is quite abundant, containing – among other information – data on all women who give birth in Hungary (e.g. their age, the number of children they have, official marital status, actual and official address, educational attainment, economic activity) and their children at birth (height, weight, gestational week, type of pregnancy, Apgar score, etc.).

2.2.2. The limitations of the HCSO vital statistics register

The HCSO Vital Statistics Register (VSR) does not precisely match the sampling population. The differences are as follows.

4 More specifically, what bias there is in terms of the sample (slightly higher selection rate within large cities) is not intended to overrepresent certain groups, but rather to compensate in advance for the lower expected response rates.

5 https://www.ksh.hu/population-and-vital-events

There is a time lag of a few months between the two populations: the survey population of our study contains women due to give birth between 1 April 2018 and 30 April 2019, and the VSR contains data on children born between 1 January and 31 December 2018 (as well as on their mothers).[6] However, this lag does not imply a substantial difference in terms of distributions, and thus it does not affect the accuracy of the weighting.

The target population of our survey is pregnant women (in the seventh month of pregnancy); but late foetal deaths (including stillbirths) mean that there are 35–40 cases where birth does not occur and so these women and their foetuses are not included in the VSR. Fortunately, the extent of this distortion is so marginal that we can disregard it altogether.[7] The VSR is based on area, and thus it also contains those women who give birth in Hungary, but do not reside in the country (e.g. Romanian citizens who give birth in Nyíregyháza Hospital, premature childbirth while on holiday in Hungary, etc.), as well as live births that cannot be linked to a particular population settlement. The former group accounts for a few hundred (619 in 2018) and the latter for a few dozen (18 in 2018) births per year. They are excluded from our survey, as this group does not reside in Hungary, and therefore does not fall within the purview of the Hungarian health visitor system.

The Vital Statistics Register contains data *at birth*, whereas the Cohort '18 study is carried out *two months before delivery* to record the living conditions of the woman and her foetus. The age (at least in round years) of the mother at the time of childbirth can be estimated with some degree of accuracy from our study, but other things may change between interview and birth: the pregnant woman may have moved to another settlement, married or divorced, graduated, and the number of her living children might, in some cases, also have changed. With the exception of age correction, these possible changes are disregarded, and it is assumed that the changes did not affect (or only minimally) the weighted variables; thus there can only be a negligible bias stemming from them. The mother's age at the time of giving birth was estimated on the basis of the mother's month and year of birth, and gestational week.

2.2.3. Weighting variables and their application

Within the prenatal data collection wave of the Cohort '18 Study, three weighting variables were computed, based on the same principles and taking very similar values.

The so-called cross-sectional database contains data on 8,287 expectant mothers and their 8,409 foetuses. This database contains data provided by those expectant mothers whose completed questionnaire (in Hungarian) and completed consent form attached to the questionnaire (see Appendix 8.2.2) was received by the HCSO HDRI by 20 June 2019. There are two types of database: in the first, the unit of analysis is the mother (or more precisely, the pregnant woman, N=8287); in the second, it is the child (or more precisely, the foetus, N=8409). The name of the former database is 'Prenatal wave database (mother)'; it contains data on 8,287 women. The name of the latter is 'Prenatal wave database (child)'; it contains data on 8,409 foetuses/children.

The reason for this distinction is the presence of twins (or more precisely, twin foetuses), because in these cases a pregnant woman can expect two or three children. It is always the research question that determines which population is more suitable for an analysis (e.g. if the research question refers to the proportion of pregnant women who smoke, it is the former; if the question concerns the proportion of children at risk of passive smoking at foetal age, then the latter is used). However, the low rate of twin births means that, in practice, there is very little difference between the two weighted databases.

6 In principle, we could have 'waited' with the weighting for the 2019 data, but those data would only have been included in the database in July 2020, and until then we could not have worked with the data.

7 There were 415 later-stage foetal deaths in Hungary in 2017.

The total number of participants in the survey – the so-called 'Total sample participating in Cohort '18 Study' – is somewhat larger than the above-mentioned datasets; that database also contains responses to foreign-language questionnaires and to the so-called proxy questionnaires, which enrolled expectant mothers (and their children) in the research at the 6-month stage. As the 6-month (proxy) data collection took place up until autumn 2019, the weighting of the entire database will only take place once all the databases are merged. This will be the complete database, and there will not be a 'mother' version. From a longitudinal point of view, the basic units of analysis cannot be pregnant women (mothers), but only their children, since it is the children, and not the mothers, who will be followed throughout the study: it is a longitudinal examination of them that will take place.

A separate weighting variable is computed for each of the three databases, albeit with a very similar value. The main characteristics of the three weights for the prenatal data collection wave is shown in *Table 1*.

The reason for the similarity of the weights (in addition to the large overlap) is that all three databases are weighted using the same reference population: the 2018 HCSO Vital Statistics Register, excluding those mothers who give birth without a valid Hungarian address (N=89150). The differences between the sub-populations are small, and thus they are also expected to be minimal between the three weighted samples. The weights were always made for the total samples of women and children: it was not necessary to create separate weights for the self-administered questionnaire database or for the prenatal care booklet dataset, as the rate of missing data was fairly low (see also Section 5).

Table 1
Main characteristics of weights computed for the datasets of prenatal wave

	Prenatal wave database (child)	Prenatal wave database (mother)	Total sample participating in the cohort study
Variable name	s_magzat_1	s_anya_1	s_magzat_v
No. of cases	8,409	8,287	8,813
Unit of analysis	foetus	pregnant women	foetus
Main objective	suited to general cross-sectional analysis		weight to be used to compute longitudinal weights

2.2.4. Merging the cells of variables used for weighting

The following dimensions were considered for weighting: mother's age, marital status, number of children living, educational attainment and place of residence. Mothers were divided into six age groups, based on their age at the time of giving birth: –19; 20–24; 25–29; 30–34; 35–39; 40+.

In terms of official marital status, we distinguished between married and unmarried women, the latter group also including single women, divorced women and widows. At the same time, in accordance with the procedure of the HCSO, any expectant mother living in a same-sex registered partnership was also classified as married.

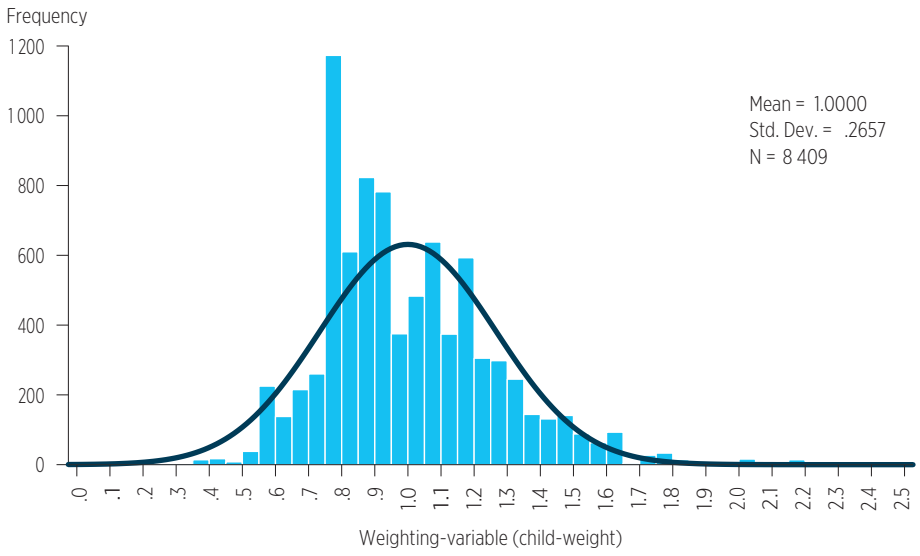
Regarding the number of children that the women already had, we classified the mothers-to-be into three groups: those expecting their first child, those expecting their second, and those expecting their third or subsequent child. In the case of twin pregnancies/births, we took into account the children born earlier as well: a pregnant woman who was expecting twins and who already had one child was categorized as 'expecting a second child'.

A detailed (10 categories) educational attainment scale was used in the questionnaire, in accordance with HCSO standards, but for the sake of analysis we grouped responses into four categories: 'at most 8 primary school years completed' (no years completed; years 1–3; years 4–5; years 6–7; year 8); vocational qualification (post-1960 vocational certificate; post-1975 vocational school diploma); high-school leaving certificate (high-school baccalaureate, qualification); university degree (college degree; university degree). This was the only variable used for weighting that had cases missing in the VSR (about 3%) – typically women living in the Central Hungarian region. Here, we proceeded as follows for imputation of the missing cases: within the given category (e.g. a woman aged 25–29, married, with one child, from Csongrád County (NUTS2) we distributed the missing data proportionally between the individual educational attainment categories.

As a result of the chosen sampling method, the sample is not representative of settlements or NUTS2 counties, but average social status (and the proportion of the Roma population) was taken into account when selecting the sample. For weighting, we created three territorial strata of approximately the same size in terms of birth rates: the Central Hungarian region (Budapest and Pest County); the 'more developed' provincial counties; and the 'less developed' provincial counties. The development classification was based on per capita GDP of the NUTS2 county (the latest available final data at the time of weighting was from 2016), and the 18 'provincial' counties were divided into two groups (each containing nine counties) according to this classification. The more developed NUTS2 counties were Bács-Kiskun, Csongrád, Fejér, Győr-Moson-Sopron, Komárom-Esztergom, Tolna, Vas, Veszprém and Zala. Meanwhile the less developed NUTS2 counties were Baranya, Békés, Borsod-Abaúj-Zemplén, Hajdu-Bihar, Heves, Jász-Nagykun-Szolnok, Nógrád, Somogy and Szabolcs-Szatmár-Bereg.

Based on the data presented above, in theory all births and all pregnant women could be classified into 432 (6 x 2 x 3 x 4 x 3) weighting cells. For example, a weighting cell could contain those women aged 30–34, married, who were having their first child, had a high-school leaving certificate, and who were living and giving birth in the Central Hungarian region. As the factors considered for weighting are interrelated, the number of cases in certain cells is very low (for example, in the Central Hungarian region there were very few married women under the age of 20, regardless of educational attainment, who gave birth to a third or subsequent child in 2019). In addition, it is typically not possible to obtain a degree in higher education under the age of 20, and thus some weighting cells had no one assigned to them. From the point of view of weighting procedure, that is not a problem, as the weighting variable takes a value of 0 for both the population and the sample. However, as we also collect some very rare events during the survey, that can result in extreme weights, even simply because of the necessary rounding of fractions. For example, a rare weighting cell might 'in theory' contain 0.14 sample members; but in reality, even the most accurate sample can naturally have only 0 or 1 such sample member. Therefore, the 113 cells from cross-tabulation where the number of the actual population was above 0 but below 50 (this means a sample size of approximately five people) were merged with a similar cell, based on the 'adjacent' age group, to reach the pooled number of 50 people in the population. By implication, the suppression of these 113 weighting cells results in the aggregation of so many age groups that we should talk about the role of relative age groups in the weighting procedure, rather than of absolute age groups. For some of the rarer demographic combinations (e.g. unmarried, university-educated women with large families), it could happen that two of the six age groups remained after suppression (however, typically 4–5 age groups remained). In the case of the weighting cells formed in this way, the actual number of cases was adjusted to the expected number of cases. The frequency distribution of the computed weighting

variables is shown in *Figure 2*.[8] The lowest value of the variable was 0.35 and the highest was 2.7; however, 90% of the values ranged from 0.63 to 1.48.



Source: Cohort '18 Study prenatal wave database (children), N=8409.

Figure 2
The histogram of weighting variable (child)

2.3. RELIABILITY OF RAW DATA

In the following, we present the unweighted and weighted sample distributions of given variables, compared to the population values. The results show that in terms of territorial strata, the proportion of respondents in the unweighted sample of the Central Hungarian region and the 'more developed' NUTS2 counties is somewhat higher than in the 'less developed' counties. This bias clearly corresponds to the bias in the responses by educational attainment: those with at most eight years of primary school, despite sample correction during the fieldwork, are significantly less represented in the sample than their share in the population would imply (16.3% vs. 20.1%; *Table 2*). In contrast, the surplus of respondents with a university degree is significant (40.8% vs. 34.6%). However, the data observed among the groups with secondary education suggest that the distortion is non-linear: there is a slight 'surplus' among those with a vocational qualification and a 'deficit' among those with a high-school leaving certificate. The combined proportion of the two groups with lower educational attainment is already significantly more accurate (31.6% vs. 29.5%); thus, in general, we cannot speak of a significant underrepresentation of the low-educated groups in our sample. After weighting, the distributions by educational attainment and territorial stratum already align precisely with the reference population data.

In terms of parity (number of children) and official marital status, the unweighted sample distribution already approximates to the population ratios rather well. In terms of age group, the fit is also good: in the case of the youngest (under 20 years of age) we can observe a difference of up to 1% between the population and the sample, with

8 In this section, we report the values of the prenatal wave database (child). Deviations from the values of the maternal database weight and the entire involved population weight are so minimal that they are not reported separately.

teenage pregnant women slightly underrepresented in the sample. As a result of the weighting, and due to the aggregation of the age groups, a complete sample fit cannot be achieved here, but the weighted sample approximates to the population very closely.

Overall, we can say that the prenatal sample is quite accurate in terms of the variables examined: the most serious distortions can be observed in the case of educational attainment, but these could be adjusted by weighting, given the sufficiently large sample size of the cohort study.

Table 2
Distribution of population and sample data, by components of weighting variable

	Population		Sample			
			Unweighted		Weighted	
	N	%	N	%	N	%
Territorial stratum of actual residence						
Central Hungarian region	27,335	30.7	2,691	32.0	2,578	30.7
'More developed' NUTS2 counties	27,404	30.7	2,658	31.6	2,585	30.7
'Less developed' NUTS2 counties	34,411	38.6	3,060	36.4	3,246	38.6
Parity: The Cohort '18 baby will be the ...						
First child of the mother	41,482	46.5	4,044	48.1	3,913	46.5
Second child of the mother	29,342	32.9	2,754	32.8	2,768	32.9
Third or subsequent child of the mother	18,326	20.6	1,611	19.2	1,729	20.6
Mother's marital status						
Married	49,466	55.5	4,638	55.2	4,665	55.5
Unmarried	39,684	44.5	3,771	44.8	3,744	44.5
Mother's educational attainment						
8 years primary school	17,902	20.1	1,372	16.3	1,688	20.1
Vocational qualification	10,277	11.5	1,107	13.2	969	11.5
High-school certificate	30,081	33.7	2,498	29.7	2,836	33.7
University degree	30,890	34.6	3,429	40.8	2,914	34.6
No response			3	0.0	3	0.0
Mother's age group at childbirth						
–19	5,269	5.9	383	4.6	468	5.6
20–24	12,764	14.3	1,139	13.5	1,218	14.5
25–29	23,093	25.9	2,241	26.7	2,193	26.1
30–34	26,982	30.3	2,615	31.1	2,550	30.3
35–39	16,265	18.2	1,567	18.6	1,545	18.4
40+	4,777	5.4	464	5.5	435	5.2
Total	89,150	100	8,409	100	8,409	100

Source: Cohort '18 Study prenatal wave database (children), N=8409.

3. CONDUCT OF FIELDWORK: MATERIALS, BRIEFINGS AND PROCEDURES

3.1. OVERVIEW OF THE DEVELOPMENT WORK

The Cohort '18 survey team used both qualitative and quantitative methods, as well as experts' opinions, in the developmental phase of the research. Before the project was launched in spring 2017, HDRI invited five social, economic, statistics and health researchers to present some *background studies and analyses* that could help the initiation and development of the survey. When the first version of the prenatal wave questionnaire was compiled by the HDRI survey team, we sent it out for review to 13 researchers whose fields of interest were covered by the topics of the questionnaire. The transparency of the research methodology and instrumentation was ensured through the organization of public events, where we gave the *audience* the opportunity to comment and ask questions. Moreover, we invited researchers and social and health policy experts to form three *advisory boards* to monitor and support the study: the Advisory Board of Experts on Social Policy, the Advisory Board of Scientific Experts and the Ethics Committee (Veroszta, 2018b). In the qualitative phase of the developmental work, four *focus-group sessions* were held among health visitors and pregnant women in the spring of 2017. During these group conversations, health visitors and pregnant women alike explored their opinions and attitudes regarding the fieldwork procedures and data collection: whether the location and the timeframe presented was appropriate for such a long-term study; what the topics were that mothers and mothers-to-be considered important and were happy to respond to; what were considered sensitive questions; how accurate, comprehensible, informative and welcoming the information booklets, the informed consent form and the data protection and data management information were (Gresits, 2018).

The prenatal data collection survey was preceded by a *pilot study.* Its aim was to test the main questionnaire and the self-administered questionnaire, in addition to the various survey materials (letter of invitation, information booklet, health visitor's questionnaire): comprehensibility and clarity of the question wording; the appropriate order of sensitive questions within the questionnaire; whether the main sections of the questionnaire and the groups of questions were in the right place in the questionnaire; which questions were difficult to answer and took longer; and which questions respondents were unable or unwilling to answer. The role of the health visitors at this stage was to carry out the pilot testing, gather information and give their opinions. Their opinions and evaluations helped the HDRI survey team write a grounded 'Interviewer Handbook (Briefing)' on the instructions related to the research (Szabó, 2018).

The *adaptation of psychological scales* to the Hungarian context of the self-administered questionnaires was also included at the preparatory stage. As most of the tests had a Hungarian translation and available relevant psychometric data, they could be used either in full or in part (drawing on previous studies) with unchanged wording, as part of the Cohort '18 Study. However, no sufficiently short measuring instrument adapted to Hungarian was available to HDRI to assess pregnancy-specific anxiety and relationship interaction. Thus, as part of an online survey study, we created a Hungarian adaptation and also analysed the reliability and validity (Kopcsó, 2018) of the Pregnancy Related Thoughts questionnaire (Rini et al., 1999) and the Gilford–Bengtson Scale (Gilford and Bengtson, 1979).

3.2. THE ORGANIZATION OF THE FIELDWORK

In the first, prenatal, data collection wave of the Cohort '18 Study, the fieldwork was carried out by the network of health visitors. This cooperation included, on the one hand, the coordination of the health visitors and, on the other, frequent contact with the leader of the sampled health visitor districts. All these tasks required continuous and attentive contact between the staff of HDRI and the health visitors.

3.2.1. Contacting the lead health visitors

Within the health visitor system, each health visitor district, including those in the study sample, belongs to a larger district, the work of which is supervised by a lead health visitor. Therefore, during the organizational phase and later, during the fieldwork, our staff had to work closely with the lead health visitor to support and facilitate the flow of information and to carry out work organization tasks, mainly with regard to logistics. The cooperation was based on mailing lists, by means of which we provided regular information and updates on the progress of the data collection. The flow of information on personnel management also took place on an ad hoc basis, via email and telephone, as changes and replacements within the pool of health visitors were handled by the lead health visitors.

3.2.2. The logistics of fieldwork: local logistical collection points

In order to ensure the conditions for data collection (i.e. the delivery, distribution and subsequent collection of the interviewer packs put together for each health visitor), it was necessary to establish local logistics points within the health visitor system. The research materials that the health visitor worked with during the fieldwork were collected at these designated points, together with the completed self-administered questionnaires and the informed consent forms. The logistics points were mostly the office premises of the lead health visitor; but, especially in the case of dispersed health visitor districts, there were also logistics units at the individual district level. Before data collection started on 1 January 2018, we sent a package of survey materials to all logistics points, sufficient for at least half a year. From there, the health visitors could pick up the number of packs assigned to them by their lead health visitor. We continuously supplied additional packs during the one-year data collection period.

3.2.3. Strengthening the commitment of health visitors

After sampling each health visitor district to be included in the survey, we contacted both the lead health visitor and the local health visitors by means of a letter of invitation, in which we not only informed them of the uniqueness of the research and the importance of the expected results, but also pointed out that data collection was impossible without their participation. In this way, we hoped to bolster their commitment to the research. We also addressed the health visitors through various forums. As a first step, we made a general presentation about the Cohort '18 survey at the annual National Lead Health Visitor Meeting in June 2017. After this plenary meeting, the research staff of HDRI held small group discussions with the heads of health visitor districts, at which there was discussion of the contact information and of the upcoming county-level health visitor meetings. However, the backbone of the fieldwork progress consisted of regular contacts with health visitors throughout the period of data collection. As a general channel for this, we sent out a health visitor

newsletter containing all information related to data collection, as well as information on deadlines and payments. In addition, as part of a continuous helpdesk service, we contacted the health visitors via email and telephone, in order to resolve any issues or problems related to data collection (for example, notifications related to relocation, clarification of problems with data entry).

Both the lead health visitors and the health visitors received financial compensation for their data collecting work, and there was an uplift to the fee if the pregnant woman had a twin pregnancy or (later) if she had no more than eight years of primary education. Payments were scheduled by HDRI on a quarterly basis, based on the number of questionnaires and informed consent forms received by a predetermined quarterly date.

At Christmas, we sent cards to all health visitors involved in the study. We also kept the health visitors informed about the progress of the research and the first preliminary results of the study, via our website and Facebook page.

3.3. ALLOCATION AND MANAGEMENT OF IDENTIFIERS

The unique identification of the cases included in the prenatal study was provided by a six-digit number called a 'token'. Tokens are centrally generated at HDRI and remain unchanged throughout the research.

The first two digits of a token identified the districts of a given lead health visitor. The third and fourth digits identified the health visitor (i.e. the interviewer who carried out the prenatal wave data collection). The last two digits identified the questionnaires of the health visitors. This six-digit number thus uniquely identified the unit of analysis (i.e. a particular pregnant woman in the prenatal wave). The token of a single foetus is the same as the token/identifier of the mother. In the case of twin pregnancies, however, the first child's ID matches the mother's token/identifier, while the twin sibling(s) receives a different six-digit ID that cannot be used as a token in other cases.

Tokens and child identifiers in later study phases form the basic variables of the longitudinal study, making identification and data linking possible in future waves.

3.4. RESEARCH MATERIALS

The following provides a description of the research materials provided to the health visitors to enable them to carry out the survey. First, the paper-based informed consent form was completed by both the health visitor and the survey participant. It was possible to fill in the main questionnaire and the prenatal care booklet either using the internet (CAPI) or using a paper-based questionnaire (PAPI), while the self-administered questionnaire could only be completed on paper. Finally, the health visitor's questionnaire could be filled in via the internet.

Table 3
Cohort '18 Study data collection types and query methods during the prenatal wave

Type of data collection	Informed consent form	Main questionnaire	Self-administered questionnaire	Prenatal care booklet	Health visitor's questionnaire
Method of query	PAPI	CAPI or PAPI	PAPI	CAPI	CAPI

3.4.1. The informed consent form

The informed consent form is a basic document and a precondition for the participation of pregnant women in the survey. An interview with an expectant woman could only begin once the respondent had signed the consent form, which the health visitor filled in with the woman's personal details. After filling in the address record form and the informed consent form, the health visitor separated these two documents and sent the informed consent form in a sealed envelope to the designated logistics point; the address record form was kept in the health visitor's folder until collection was organized by HDRI.

For the informed consent form, the respondent was required to provide her name (birth name and mother's name), place and time of birth, and actual and official addresses. In addition, she could optionally provide her phone number and email address, as well as her social security number, if she consented to data-linkage based on this information.

3.4.2. Personal data protection and data management protocol

If possible, health visitors informed all pregnant women from the sampled health visitor districts about the survey before the 28th week of their pregnancy, and invited them to participate. This was done at a scheduled pregnancy care appointment, during a face-to-face meeting between the health visitor and the client. It was a precondition for starting data collection that the woman should read (or otherwise get to know) the Personal Data Protection and Data Management Protocol document and sign the informed consent form. The protocol document provided details about the consent form clause by clause. The expectant mother received and could study both documents well before the interview; if she wished, she could also take the documents away with her to study prior to the interview.

3.4.3. The main questionnaire

The topics and the questions of the main questionnaire were built up stage by stage, taking into account the opinions of the expert reviewer-researchers and the conclusions of the pilot survey. The main questionnaire covered such topics as opinions about family, parenting and paternal roles; fertility history and childbearing plans; the history of the pregnant couple's relationship; household, housing, social and labour market situation; health status and social relationships (Appendix 8.1.1).

The main questionnaire was also sent to the health visitors in paper-based format; however, the interviewers had the opportunity to record the respondent's answers immediately, using an online interface created specifically for this purpose. The questionnaires completed on paper subsequently had to be entered into the web-based database as well. Storage of the completed paper-based questionnaires was also the responsibility of health visitors, until HDRI collected them during each collection wave.

3.4.4. Show cards

HDRI made the health visitors aware that, in order to ensure the uniformity of the survey, it was important that the questions in the questionnaire should be asked by everyone in the same order and using the same wording, and that the answers should be recorded uniformly in accordance with the given criteria. This was also the thinking behind the creation of 'show cards'. It was compulsory to use these cards, which had to be given to a respondent during the interview in order to make longer, more complex questions

and answer categories easier to review and understand (e.g. sources of income, benefits, education, occupation classification). The answer categories on the show card were clearly linked to the questions in the questionnaire. Two packs of show cards were given to each health visitor, and one was handed to the expectant mother during the interview. The respondent gave the pack back to the health visitor at the end of the interview.

3.4.5. The self-administered questionnaire

After the respondents had answered the main questionnaire, the health visitors were to give them the self-administered questionnaire, the completion of which was anonymous and voluntary. As with the main questionnaire, the expectant mother's identification number had to be inserted at the beginning of the self-administered questionnaire. However, unlike the main questionnaire, the self-administered questionnaire could only be completed on paper; their answers were recorded not by the health visitors, but by our staff, once these questionnaires arrived at HDRI. In this questionnaire, which took about 20 minutes to complete and covered more sensitive topics, we asked questions about the subjective feelings and thoughts of the women.

If there was a problem in completing the questionnaire (e.g. due to difficulty reading or visual impairment), the health visitor assisted by reading out the questions and answer categories from her own copy, so that the expectant mother could mark her answer herself, if she wanted to, while it remained hidden from the health visitor. If the woman had issues with understanding the question, the health visitor could volunteer interpretative remarks in a way that avoided influencing the woman (see also 'Data checking and cleaning of main questionnaire' in Section 5.2.2). Immediately after completing the self-administered questionnaire, the respondent (or, if requested, the health visitor) placed the questionnaire in an envelope and sealed it. The sealed envelopes were delivered by the health visitors to the designated logistics points, from where they were collected.

3.4.6. The prenatal care booklet

Pregnancy care in Hungary is covered by a professional protocol detailed in decree 26/2014 on pregnancy care (IV. 8.) of the Ministry of Human Resources. Pregnancy care is defined there as a complex healthcare service, based on the cooperation of a health visitor, general practitioner, obstetrician/gynaecologist, midwife (if the pregnant woman wants one) and the expectant mother herself. In accordance with the protocol, the health visitor issues the pregnant women with a prenatal care booklet, which can be used to track both the course of the pregnancy and the condition of the mother and foetus. This booklet must include all events, appointments/examinations, and results that relate to the pregnancy and the expectant mother.

The last pages of the Cohort '18 Study's prenatal main questionnaire included a series of questions that referred to and were based on the information from the prenatal care booklet. So that the research could make use of the information from the prenatal care booklet, we asked respondents for their special agreement on the informed consent form. Thus, in order to reduce the burden on the respondents, interviewers did not need to ask that information. Having collected it from the prenatal care booklet, a health visitor could enter it either on a paper form or input it directly using the internet interface. HDRI suggested that, if possible, health visitors should enter the prenatal care booklet information while the mother-to-be was filling in the self-administered questionnaire. It was also important for the health visitor to remind the pregnant woman to bring her prenatal care booklet with her to the examination. It was not necessary to enter all the information recorded

in the prenatal care booklet: only information that the researchers considered to be most important in terms of the development of the foetus – such as data on previous pregnancies, births and miscarriages, data on blood group, oncological examination, data on developmental abnormalities in the family, hereditary diseases, drug allergies and a risk assessment.

3.4.7. The health visitor's questionnaire

At the end of the main questionnaire, the health visitor's questionnaire appeared on the web interface. This survey material included questions about the circumstances of each interview. Through closed-ended questions, health visitors had to indicate how the interview had been carried out; who had been present during the interview (other than the interviewer and interviewee); where the interview had taken place; which factors had made it difficult for the pregnant woman to answer the questions; whether she had needed help with the self-administered questionnaire; whether the interview had been interrupted; and in what week of the pregnancy the woman had started receiving pregnancy care.

If the health visitor input the main questionnaire immediately, using the web interface, the health visitor's questionnaire appeared right at the end, so that she could fill it in at once. However, if the health visitor used a paper-based questionnaire for the interview and only later input the data using the web interface, she had to answer the health visitor's questionnaire retrospectively, based on her recollections.

3.4.8. The address record form

As well as the burden on the participants in the research, another reason for dropping out of the study sample was because they had moved to another location. Therefore, we impressed upon the expectant mothers and health visitors that we need to visit the same families and children throughout the course of the survey. That is why we have to record the address of a respondent and why, if the mother moves, it is important for her to provide HDRI with her new address and contact information.[9] To this end, as part of the informed consent form, HDRI prepared an address recording form, on which the health visitor recorded the contact details of the expectant mother – e.g. address and telephone number and the dates of the interviews during the various waves of data collection (Appendix 8.2.2). If she moved, the new contact information was added to the address recording form, which was kept by the health visitor responsible for conducting the 6-month interview. By signing the statement at the bottom of the address recording form, the health visitor guaranteed the authenticity and confidentiality of the data gleaned from an interview. The declaration was intended to certify fulfilment of the tasks included in the health visitor's contract. Moreover, by signing the document, the health visitors also consented to the monitoring and quality control of their fieldwork procedures and interviews.

3.4.9. Additional materials to promote and support the study

In order to promote the study and to engage the participants, each health visitor received a glossy poster about the Cohort '18 Study to display in the office in a prominent position. In addition, we designed a colourful information booklet about various aspects of the study. In this short booklet, we tried to address the most important questions concerning the survey: Why was the study being carried out? Who could participate? How was the

9 We made a form available to the health visitors through a closed, password-protected interface on the study website, where they could mark the occurrence of the relocation of a participant between the prenatal and the 6-month data collection waves, allowing us to follow the movers and, if necessary, involve new health visitors.

research being done? Why was it good/useful to be involved? The information booklets also included the availability of the Cohort '18 Study website and Facebook page, as well as our email and mailing addresses (see also Section 4.3).

3.5. SPECIAL PROCEDURES DURING THE INTERVIEWS

3.5.1. Interviewing pregnant minors

In the case of a pregnant minor (under the age of 18 and unmarried), special consent was required from her legal guardian before she could participate. Naturally, the health visitors also provided the woman herself with information and asked for her explicit oral consent. The prenatal wave database included 149 unmarried women under the age of 18.

3.5.2. Foreign-language questionnaires

The study's sample included all women belonging to selected health visitor districts whose due date of childbirth was between 1 April 2018 and 30 April 2019, according to the data entered in the prenatal care booklet. By definition, those expectant mothers who did not understand or speak Hungarian were also considered to be part of the sample. During the developmental phase of the study, and before the prenatal interviews were carried out, some health visitors indicated that their districts contained non-Hungarian-speaking mothers-to-be, who would certainly not be able to fill in the questionnaire in Hungarian. Following this important observation, we asked all the health visitors participating in the research to tell us whether, based on their experiences so far, questionnaires in languages other than Hungarian would be needed in their district; and if so, in which languages the questionnaires should be available. Based on the feedback received, we decided to translate the questionnaires into four languages: English, German, Chinese and Vietnamese (there would have been some demand for other languages as well – Ukrainian, Russian and Arabic – but we decided not to translate the questionnaire into those additional languages, due to the significantly lower demand). Certainly, we could not expect the health visitors to speak the foreign languages and to ask the translated questionnaires themselves; and so, as a bridging solution, we prepared an abbreviated version of 48 questions, drawn from the Hungarian questionnaire, in the form of a self-administered questionnaire. Thus, the health visitor's job in this case was to hand over the chosen foreign-language questionnaire to the pregnant woman, who filled it in herself and returned it to the health visitor in a sealed envelope. The translation into all four languages was done by professional translators. Contrary to expectations, a very small number of foreign-language questionnaires were completed: only nine English, five Chinese and two Vietnamese questionnaires were received back. Due to the minimal demand, we did not prepare foreign-language questionnaires for subsequent stages of the study. This does not mean that the children of foreign mothers living in Hungary are completely excluded from the research: feedback received from the health visitors indicates that, in some cases, the women completed the Hungarian questionnaire with help (through the interpretation efforts of their Hungarian-speaking husbands or other companions).

3.5.3. Interviewing pregnant women with learning disabilities

Mothers with learning disabilities are a very special group in the Cohort '18 Study, and our aim was also to collect data on their life situation as well. If the disabled person had a guardian, a separate informed consent form had to be signed by the guardian

(as in the case of minors). If it was challenging for the woman to answer certain factual questions, the guardian or an accompanying relative was allowed to help them answer the question. On the other hand, all other questions referring to opinions, plans and feelings, which were therefore subjective in nature, had to be answered independently by the expectant mother herself, as best she understood the concepts. (We recorded three cases where a guardian signed the consent form, even though the woman was over the age of 18.)

3.5.4. Interviewing pregnant women who were planning to give up their child for adoption

Because the main focus of the study is on the child, we also asked health visitors to interview any expectant mother who was considering giving her child up for adoption (provided, of course, she agreed to participate in the research). We plan to continue to follow the child through open adoption or placement in an institution. (The data collection wave when the child was 6 months old revealed whether the infant had been put up for adoption, in 18 cases.)

4. DATA QUALITY ASSURANCE AND CONTROL

4.1. SUPPORTING THE UNIFORM INQUIRY PROCESS: BRIEFING AND TRAINING OF INTERVIEWERS

One of the elements of ensuring data quality is that interviews and data entering should be complete and should always take place in the same way. Both the training and the briefing of health visitors served this purpose.

One of the most important aspects of our research was the training of the health visitors who conducted the fieldwork and were in direct contact with the study participants. During the training, all health visitors received the same information about the principles of the study, the content of the questionnaires and the methodology on which the survey was based. The health visitor training took place between 6 November and 1 December 2017, in 36 locations, with a maximum of 25 participants per briefing. During the training sessions, the health visitors were briefed on how to inform the expectant mothers about the study and how to ask for their participation. The success

Table 4
Chronological guidance on fieldwork process for health visitorse

Prenatal wave fieldwork period: 1 January 2018 – April 2019	
WHAT?	WHEN?
Participation in the health visitor training, signing of contract	Prior to the start of the fieldwork, during November and December 2017.
Receiving survey packs	Prior to the start of the fieldwork, before 1 January 2018. The survey packs were delivered to the logistics points in December 2017 and could be picked up from there.
Informing and inviting expectant mothers to participate	During the period before the 28th week of pregnancy, during pregnancy care. Handing over an information booklet and a letter of invitation.
Organizing the interview	During the 28th–31st weeks of pregnancy.
Issuing of identification numbers	Before starting the interview, to be recorded in the health visitor's folder and on the informed consent form.
Signing the informed consent form	Before the start of the interview, at the site of the interview.
Interview	In the 28th–31st weeks of pregnancy, at a prearranged time and location.
Filling in the data of the prenatal care booklet	Immediately after asking the main questionnaire, during completion of the self-administered questionnaire by the participant.
Completion of self-administered questionnaire	Immediately after the main questionnaire, while the data from the prenatal care booklet were being entered.
Health visitor folder management: fieldwork administration	After a successful interview, recorded on a spreadsheet of the health visitor folder. In case of unsuccessful recruitment, following a failed request or a refusal, recorded on a spreadsheet in the health visitor folder.
Entering the questionnaires on web interface	After the interview (as soon as possible or at the same time).
Sending of the research materials to HDRI	Envelopes containing the informed consent forms and self-administered questionnaires were regularly collected by HDRI from the logistics points: it depended on the location of the given logistics point, but typically occurred every 2 months. It was therefore necessary to send them to the collection points as soon as possible after the interview. The completed paper-based questionnaires were delivered simultaneously from the logistics points at the end of the study wave.
Monitoring and conducting the 6-month study wave	The second wave of interviews – when the children were 6 months old – began in early June 2018 in the case of premature births, but typically took place from autumn 2018. Monitoring this was also the responsibility of the health visitors.

Source: own design.

of the recruitment was also supported by the communication materials and channels presented in Section 4.3. We gave precise guidance on the process by which health visitors would select the expectant mothers for the sample. We guided them step by step in the order of their tasks (*Table 4*). The health visitors' feedback confirmed that these training sessions were extremely helpful (see also Boros, 2018).

Support for health visitors and the interviews also came from the so-called 'Interviewer Handbook (Briefing)' and the Frequently Asked Questions (FAQ) section on the study website (health visitor interface) prepared by the research team. Everything that was presented at the training sessions was also summarized in the handbook that each health visitor received. This also served as a detailed methodological guide on how to input the paper-based interviews onto the web interface prepared for the purpose (see Rohr, 2018 and Section 5.2.1). The FAQ section on the study website provided quick assistance on key issues and prepared health visitors for situations where a pregnant woman might ask them why they were being asked to join, how long they would be involved for, what would happen to their data, etc.

4.2. MONITORING AND CHECKING OF DATA COLLECTION

One key point in ensuring data quality is that interviews have to be complete, and a detailed overview was required of those expectant mothers who failed to participate in the survey. The research material packs were delivered, distributed and collected through local logistics points within the health visitor system. It was the responsibility of the lead health visitors to monitor and control the processes from these logistics points. They monitored the stocks of the various elements in the survey material packs and kept HDRI informed. The sealed envelopes of the self-administered questionnaires and informed consent forms had to be delivered to these logistics points by the health visitors as soon as possible after the interview. HDRI collected these survey documents from the points by courier every three months. The two documents had to be placed in separate, sealed envelopes, and the envelopes had to be stored in a lockable, secure place, and care had to be taken to ensure that the data did not fall into unauthorized hands. As the survey materials arrived at HDRI, they were checked by the responsible researcher from the survey staff at HDRI, primarily to make sure the questionnaire had been correctly completed and was accompanied by a signed informed consent form. If any deficiencies were found, the health visitor responsible was consulted.

The progress of the prenatal data collection wave was monitored at several stages. The monitoring was based on a so-called 'successful interviews count' database. We recorded the estimated number of births and calculated the expected number of successful interviews for each health visitor district selected in the sample (used as a point of reference). The actual number of interviews was compared to the expected number at the district level. We performed continuous monitoring between May and August 2018. In health visitor districts where fewer interviews were conducted than expected, we collected further information and asked the lead health visitor for a likely explanation for the low number of interviews. The lead health visitors were informed of the results of the monitoring in email circulars, and they sent us back their answers on discrepancies.

Tracking the number of interviews and failed recruitments was the responsibility of the health visitors. For this, the survey staff prepared a so-called *health visitor folder*, which could be regarded as a paper-based registration interface. Throughout the first two waves of the survey (the prenatal and the 6-month data collection), the health visitors collected respondents' contact data and provided a summary of successful and unsuccessful recruitments. Successfully recruited respondents were those who agreed

to participate and who completed the questionnaires (either the prenatal or later, the 6-month questionnaire). Based on the information entered (date of prenatal wave interview, due date of delivery, actual date of delivery; and later, expected and actual date of the 6-month interview), the health visitor was able to track how many successful interviews she had completed and when the 6-month interviews were due to take place. The spreadsheet with recruitment information was also used to track unsuccessful requests. Due to the voluntary nature of the study, it was expected that some pregnant women would not wish to participate in the research, even once they were in possession of the facts. For them, the study was over at this point; but in order to make the results of the study reliable, it was important for HDRI to see who had refused to participate. Therefore, we also asked for some non-personally identifiable data on these pregnant women to be recorded: how many children they had; what age groups they belonged to; what their marital status was; and what their highest level of education was.

4.3. COMMUNICATION OF THE COHORT '18 STUDY'S PRENATAL WAVE

The communication strategy related to the study had several target groups at the time of the prenatal wave. Communications thus served various purposes and were pursued through several channels. The goals included sparking the interest of *respondents*, increasing and maintaining their willingness to participate, and creating a sense of identity related to the study. These goals were served by the information booklet prepared for respondents, information available online, and the visibility of the Cohort '18 Study at various events. In addition, we were in constant contact with the *health visitors* who recruited and interviewed the women during the first two data collection waves. The primary platforms for informing the *professional community* were the study website and the Facebook page, as well as publications related to the study.

4.3.1. Communication towards respondents

At the beginning of the study, as part of the survey material pack, we provided the respondents with a selection of information material, via the health visitors. As well as inviting expectant mothers to participate, the *letter of invitation* also offered them fundamental information about the study. An *information booklet* provided key information on the study in detail and in a comprehensible manner. A *poster* promoting the study and smaller leaflets were placed by health visitors in prominent places in the offices of the health visitor districts involved in the research. The *Personal Data Protection and Data Management Information* booklet explained to the participants in detail the purpose of the study, what legal guarantees we offer for the personal data provided by them, and how we would guarantee their anonymity during the data processing.

Before data collection started, the website *www.kohorsz18.hu/en/* was launched in Hungarian and in English; it lists separate menu items for respondents to consult, and contains a detailed presentation of the study, frequently asked questions and all the ways in which respondents can contact the research staff directly. There was also an opportunity to subscribe to a newsletter via the website (which remains live): we send a quarterly report approx every three months to anybody interested in the progress of the study. (Professional news related to the study is also published in the HDRI newsletters.) By January 2020, menu items presenting the research in general had received more than 3,000 'hits'.

The Facebook page of the Cohort '18 Study was launched on 2 February 2018 (in Hungarian), and has been posting twice a week ever since; in January 2020 it had 619

followers. Its primary goal is to be an additional channel through which we can reach out to our respondents and thus increase their willingness to respond; but it is also suitable for regularly informing professionals and families raising young children, and maintaining their interest. Above all, the Facebook page provides news and fresh content about the progress and results of the study. In addition, we present, for example, interesting results and insights from international cohort studies; these aim to give our followers an idea of the types of benefits that social science research can bring, and thus to make them more committed to responding. We sometimes also share other research findings relevant to pregnancy and early childhood, as well as the content of popular scientific and family-related pages that may be of interest to families with young children. These posts reach an average of 200–300 people.

Another means of informing respondents and reaching a wider audience is to have the news and results concerning the study published in the *media*. Finally – and although this may only have succeeded in reaching a smaller proportion of respondents – the research was showcased at various *events*. Since these events took place while prenatal data collection was still going on, our goal was partly to allow respondents (and potential respondents) to meet the researchers in person, and partly to allow families with small children to find out about the study and its results, in case they were interested.

4.3.2. Contact and communication with health visitors

Prior to the start of the Cohort '18 Study, all those health visitors involved in the research took part in small-group training, to prepare them for data collection (see Boros, 2018 and Section 4.1 of this volume). During an audit following the first few months of fieldwork during the prenatal data collection wave, it became clear that the response rate among certain groups and in certain territorial areas was falling short of our expectations. As the prenatal data collection period spanned a year, and since we were able to pinpoint a number of factors that, in principle, affected data collection in various geographically dispersed health visitor districts in different ways, we were able to develop a new support briefing based on insights from the current study wave. Some practical considerations related to the 6-month wave were also discussed in this new briefing. With this material, we revisited some of the health visitor districts, with special emphasis and priority given to those in areas of the country where we had encountered a problem with the fieldwork. Health visitors received new, up-to-date, important information about the research through email newsletters as well. In addition, we maintained a closed Facebook group for those health visitors carrying out the data collection, where they could also ask questions directly of the research team and the organizers of the study.

In addition, health visitors could gain further information via the *Health Visitor Journal*, by means of which not only those who were participating in the data collection, but the whole network of health visitors could be informed of progress in the study – research that had been started and was being implemented with the help of their colleagues.

Finally, we created a closed health visitor interface on the Cohort '18 Study website, through which health visitors could access links to various questionnaires, their documents and the FAQ section specifically created for them.

5. DATA EDITING, CHECKING AND CLEANING

The data collection was followed by entering, checking and cleaning the data from the survey materials: questionnaires, address record forms, informed consent forms and fieldwork progress information. In all cases, we kept in mind that the databases created during the prenatal data collection wave should be user-friendly and should reflect the reliability and quality of the original, raw data. We had to solve the dilemma of what to do with data that appeared in multiple data sources and that were in apparent contradiction with each other. We took each problem individually and tried to resolve it. These procedures are described in the following section, after a brief presentation of the databases created during the fieldwork.

5.1. THE SURVEY DATABASES

Different databases were created for input of the different information gleaned from the survey materials used during the fieldwork.

Table 5
Databases created during the prenatal data collection wave

Name of the instrument	Database
Main questionnaire	Database on the main questionnaire entered via the web interface
Self-administered questionnaire	Database on the data recorded on paper-based questionnaire and then subsequently entered by HDRI
Prenatal care booklet	Database on the prenatal care booklets recorded via the web interface
Health visitor questionnaire	Data from the health visitor's questionnaire recorded via the web interface
Informed consent form	Database on data recorded on paper and then subsequently entered by HDRI
Address record form	Data recorded on paper and online; the database was updated continuously during the fieldwork by the health visitors whenever a respondent moved house

5.1.1. Database of the main questionnaire

The information from the main questionnaire was either entered by the health visitors online during the questioning, or was recorded on a paper-based questionnaire, from which the answers were subsequently entered via the web interface. This database included the unique identifiers of the health visitor district, the health visitor and the foetus(es), the week and type of pregnancy (single, double or triple foetuses), the date and time of the interview, as well as all the answers from the main questionnaire.

The two most important variables in the database are two identifier variables: the serial number and the token. In the web interface system, each individual case or interview is assigned a *serial number* that cannot ever be deleted or modified under any circumstances. The case-level cleaning procedure for the token or any other variable is done with the help of this serial number. All cleaning procedures are linked to this case-specific serial number. The so-called *token* identifier is the identifier of the respondent, or more precisely, of the foetuses. The web interface system only allowed each token to be used once, and so no duplicates could occur during data entry. However, in some cases, tokens that did not belong to women/foetuses could be recorded by mistake. The reasons and methods for token cleaning are detailed in the section on data editing. The main questionnaire database contains 561 variables in total. Most of the questions

were closed-ended, but 12 of them permitted respondents to give various possible answers. For some questions, we expected an answer in the form of a date (on the web interface, it was possible to select the appropriate date from a calendar), and there were several open-ended questions where the answer was a number (e.g. week of pregnancy, or the weight and height of the woman, or the number of children they already had, including those adopted). We expected a fully open-ended, textual answer for three questions in all: the occupation of the expectant mother, the occupation of her partner, and a description of the institution (if the woman was living in an institutional setting). The topics of the questionnaire included: the demographic characteristics of the expectant mother; her attitudes towards gender and family relations; female–male roles and paternal roles; her fertility and relationship history; household and housing data; the social and labour market situation of the pregnant woman and her partner; the health status of the woman; and her social network (Appendix 8.1.1).

5.1.2. Self-administered questionnaire database

Certain variables (such as those related to psychology) were measured using self-reporting paper-and-pencil tests. These were handed to the pregnant woman in the form of a booklet by the health visitor after the interview (Appendix 8.1.2). The self-administered questionnaire was sealed in an envelope immediately upon completion, and the health visitor delivered it to HDRI, where it was entered. The expected number of respondents to the self-administered questionnaire was 8,287, as the booklet was handed to all respondents in the prenatal wave survey; however, the actual number of complete or partial interviews (returns) was 8,191.

The metadata for the self-administered database was the token ID. The (mostly closed) questions in the self-administered questionnaire produced a total of 113 variables 3 variables. In some cases, we expected a numerical answer; and three questions were open-ended. We then computed 10 new variables (indices) from the questions of the self-administered questionnaire, and these also form part of the database (see also 'Data checking and cleaning of the self-administered questionnaire' in Section 5.2.2, and Appendix 8.3). The topics covered by the self-administered questionnaire were: childbirth and childbearing, emotional and mood state, lifestyle and living conditions, and partnership.

Each of the variables in the focus of the Cohort '18 Study – and particularly those covered by the field of psychology – was measured not by means of an individual question, but by using a battery of questions or statements, which were included in the self-administered questionnaire (Appendix 8.3). By using multi-item measurement tools, we could get an idea of abstract phenomena that could not be measured directly through simply one question. When selecting the scales used, in addition to their fit to our research questions, we took account of existing data on their validity and reliability, which indicated whether the tool was indeed suited to accurately assessing the phenomenon that was to be studied.

5.1.3. Database of prenatal care booklet information

After administering the main questionnaire, the health visitors then input selected information from the prenatal care booklet via the web interface developed for this purpose (Appendix 8.1.1). The information from the prenatal care booklet questionnaire yielded a total of 155 variables, and this information is available for a total of 8,269 women. From the questions in the booklet, we created 9 new variables (indices) that also form part of the database. These measure such things as the level of pregnancy risk, the causes of high-risk pregnancies, surgery and medical history, and the expectant mother's drug susceptibility. The data from the prenatal care booklet were: test results relevant to

childbirth; data on previous births and pregnancies; data on previous failed pregnancies; data from general practice examinations and medical history; determination of due date of delivery; and risk classification.

5.1.4. Health visitor's questionnaire database

The health visitor's questionnaire referred to the circumstances of the interview. It was entered immediately following the entry of data from the main questionnaire (i.e. the two types of data entry occurred at approximately the same time). The mostly closed-ended questions (plus one open-ended question) yielded 20 different variables. The questions referred to the method of interview; who had been present during the interview; where the interview had taken place; what had caused the respondent difficulty; whether the expectant mother had needed help in completing the self-administered questionnaire; whether the questioning had been interrupted; and in what week of pregnancy the woman had started with pregnancy care.

5.1.5. Informed consent form database

The informed consent form information was entered at HDRI: the name of the respondent (including her name at birth and her mother's name), place and time of birth, and actual and official address. In addition, the woman had had the opportunity to provide additional information in the shape of her telephone number and email address, as well as her social security number. These data are stored separately from the questionnaire databases, in a secure, password-protected file, so that neither a third party nor any unauthorized employees of HDRI can access them, in accordance with applicable data protection regulations. Of the data from the informed consent form, the date of birth of the mother was compared with the same data from the main pregnancy questionnaire for the purpose of verification (Appendix 8.2.2).

5.1.6. Address recording form database

The health visitors' folder also included a spreadsheet indicating whether the interview had been successful or not; background information on successful and failed interviews; the planned stages of the next data collection; and all moves undertaken by the women. If the pregnant woman moved away from her initial address, the health visitor could input the new address card data via the web interface, and thus the address record form database was continuously built up and maintained (Appendix 8.2.2).

5.2. DATA EDITING

5.2.1. Data entry

The interviews and data entry were both performed by the health visitors. During the initial briefing sessions, we provided the health visitors with comprehensive instructions. Throughout the period of fieldwork, we also provided training in data entry and in the easy-to-use data entry web interface. To this end, HDRI developed an IT system that could be easily adapted to the diverse and widely divergent IT skills and equipment of the health visitors taking part in our research. The development of this data entry interface was preceded by a needs and infrastructure asset survey, which revealed the computer expertise of the health visitors. The upshot of this was that, in addition to direct online data entry (CAPI), we had to make the system compatible with data entry from paper-

based questionnaires (PAPI) – without having access to health visitors' computers, i.e. without being able to install software and provide assistance.

Thus, in order to ensure continuous data entry, we developed an online questionnaire and also operated an online data entry interface. The Cohort '18 Study used the *online-kerdoiv.com* interface, developed by Hungarian programmers for this purpose, with the questionnaire software designed around the special needs of the Cohort '18 Study data collection. Thanks to the built-in system checks, the data entry operators (i.e. the health visitors) could immediately check for discrepancies within the data and correct any errors. A questionnaire was considered to be accepted by the system if it was error-free and complete. Since data entry occurred continuously – either in parallel with the data collection (CAPI) or shortly thereafter (PAPI) – both progress monitoring and data cleaning could take place in real time. As a result, the sample adjustment decisions (based on projections) could be based on statistics measured at the half-way point of the data collection period. In addition, this system provided us with an opportunity to produce interim interview databases that served as the basis for the publication of preliminary research findings.

5.2.2. Data checking and cleaning

Data checking of informed consent forms

Once they were collected by HDRI, the informed consent forms were sorted on the basis of their token identification number. They were stored in a lockable safe, to which only those HDRI research staff responsible for data entry and data cleaning had access. If certain personal information was missing from the informed consent form, or if the form was not clearly completed, we contacted the relevant health visitor for clarification.

Data checking and cleaning of main questionnaire

The online data entry process was designed in such a way that only one questionnaire could be entered and finalized with a token. However, the system allowed health visitors to input the data using tokens belonging to other health visitors. We provided this feature in order to ensure the continuous use of this unique identifier in the case of respondents who moved to another health visitor district during the fieldwork period. Another health visitor had to interview the woman if she moved to another health visitor district, but the same unchangeable identifier had to be used. As we mentioned in Section 3.3, we apply two identifiers in the Cohort '18 Study. The first is a serial number that uniquely identifies every case entered via the web interface; we do not use serial numbers to identify sample members or for data linkage, but only for data cleaning (as will be shown below). The second ID in the study is the more important one, as it serves to identify each member of the sample (i.e. the foetuses/children) and to link the data in the different databases of the study. However, this ID does double duty, serving as two different variables in the study databases: the '*token' variable* and the '*child ID' variable*. These two variables are identical only in the case of single pregnancy; in the case of a double or triple pregnancy, the foetuses have IDs that vary by one or two digits (i.e. in the case of twins, we have three identifier variables in the database, for example: token=101102; child id1=101102; child id2=101103). A typing error or an incorrectly filled-in form could mean that sometimes the pair of the *token* and the *child ID* variables differed, so that the child ID was entered as the token for another case. HDRI can identify, document and clean these types of errors on the basis of feedback from the responsible health visitors.

Checking and cleaning the Cohort '18 Study databases proved to be a complex process. The data entry program already had a built-in check during the recording process: certain values that fell outside the possible range of answers to questions were not allowed to be entered. The skip-the-question shortcuts were also built into the program: that is, if a question did not have to be asked of a given expectant mother, that question did not even appear on the screen. By and large, the questionnaire contained closed-ended questions, with predetermined answer categories where missteps or mis-recordings were not possible.

The data were checked and cleaned primarily at the individual record level: we enumerated and documented all the erroneous cases at the individual record level. We also checked the logical connections between the different variable-groups (i.e. we looked at the logically related answers given to different questions). To correct the data inconsistencies, we not only reviewed the logically related data within the main questionnaire, but in many cases we also included the information from the informed consent form, the prenatal care booklet and the 6-month questionnaire. However, as well as verifiable and correctable errors, there were also some responses that could not be cleaned, even after checking for any logical connections (such as variables related to household size or physical activity of the expectant mother). In those cases, we kept the original answers.

Data checking and cleaning of the self-administered questionnaire

The final self-administered database contained 8,191 recorded cases: this corresponds to the number of unique self-administered questionnaires that were returned to HDRI for data entry. The checking of the self-administered prenatal database (as with the main database) began with an overview of the distributions of the variables and the elimination of erroneous values. As the self-administered questionnaire was filled in by the participants on paper, with their answers subsequently entered by a person commissioned by HDRI, a significant part of the data cleaning involved errors in manual entering. Cleaning thus consisted of correcting or deleting irrelevant responses. Outlier values (e.g. in relation to income questions) were coded as missing cases, and several textual responses classified as 'other' were recoded into predefined categories (e.g. for maternity plans).

Data checking and cleaning of the prenatal care booklet

The administrative data on the complex care of health visitors, GPs and specialist doctors generated during maternity care in Hungary are collected in the prenatal care booklet issued by the health visitor at the beginning of the pregnancy care period. This booklet is given to the expectant mother, but it is filled in by the health visitor. The data content of the prenatal care booklet formed the basis on which individual-level administrative data linkage took place in the first wave of the Cohort '18 Study. Prior to the start of data collection, respondents agreed to this in a dedicated section of the informed consent form. Because the information from the prenatal care booklet is not recorded and integrated into a unified database by the Hungarian system, HDRI had to create such a database for the Cohort '18 Study sample. That was why we asked health visitors to enter some important information from the prenatal care booklet, with the consent of the mother-to-be. The focus was on incorporating data that we were unable to collect during the interview with the expectant mother (typically medical data). This information was inserted at the final stage of the prenatal interview – essentially without a break – in the last pages of the main questionnaire. The health visitors could enter the prenatal care booklet data via the web interface while the expectant mother was filling in the self-administered questionnaire.

As a first step, by checking and cleaning the token and child ID identifiers, we ensured the data linkage of the prenatal care booklet database with the other research databases. The prenatal care booklet database contains 8,269 cases that can be directly linked to the database of the main questionnaire (N=8287). Data cleaning was possible if the relevant data were included in both the main questionnaire and the prenatal care booklet (specifically inserted in both questionnaires partly for control purposes). In these cases, the discrepant data were cleaned in the prenatal care booklet database, based on the individual-level data from the main questionnaire.

5.2.3. Data linkage and merging different survey datasets

Merging different survey datasets of the Cohort '18 Study

Merging the different datasets of the Cohort '18 Study was possible only after the checking and cleaning of the six-digit token identification number. The data linkages had to be performed not only once the prenatal study wave had ended, but also during several phases of data collection, as the payments made to the health visitors were linked to the availability of the different research documents at HDRI and the appropriate linkage of these survey materials using the token. In the prenatal wave, we had a total of four survey databases that we needed to merge: the main questionnaire, the self-administered questionnaire, data taken from the prenatal care booklet, and the database containing the information on the informed consent form. The first three databases were merged, edited and cleaned only once the informed consent form had been collected by HDRI. The informed consent form and the main questionnaire both contained the pregnant woman's date of birth (year, month); thus, when merging these databases, we also used those data to check the linkage (see 'Data checking of informed consent forms' in Section 5.2.2). The self-administered questionnaire did not contain any variables that were included in the other questionnaires; thus, we were only able to merge the self-administered and the main questionnaire databases using the token identifier (see 'Data checking and cleaning of the self-administered questionnaire' in Section 5.2.2). The prenatal care booklet and the main pregnancy questionnaire both included previous births and existing children's years of birth; thus, we used the consistency of the data across these datasets to check the linkage of these databases (see 'Data checking and cleaning of the prenatal care booklet' in Section 5.2.2).

Territorial data linkage to the Cohort '18 Study

In the case of the prenatal data collection wave, the territorial variables are linked at the settlement and administrative district level; but for reasons of data protection, the public data files do not contain settlement codes or variables enabling the identification of respondents at the settlement level. *The linkage of settlement-level data* was done according to the HCSO's own settlement identification registration number[10] and/or the 2018 Administrative Settlement Register of Hungary.[11] By default, the variable on which the linkage was based was the actual place of residence of the mother/child (which was recorded in the informed consent form). In some special cases, when such a settlement did not exist or when the settlement could not clearly be determined (because it was abroad, missing data, etc.) from the informed

10 See the Decree No. 31/2011 (X. 24.) of the Minister of Public Administration and Justice on the territorial registration identification code system: https://www.ksh.hu/docs/osztalyozasok/teruleti_szamjel/31_2011_decree.pdf

11 Available (in Hungarian and English) at: http://www.ksh.hu/docs/hun/hnk/hnk_2018.pdf

consent form, linkage was based on the settlement given in the official address. The *linkage of administrative district-level data* was done by using 'administrative district codes'.[12] The classification according to the district code is based on the settlement identification number.

5.3. DATABASE STORAGE, ACCESSIBILITY

The children included in the sample are identified through a unique study identifier, which was assigned by the health visitor during the prenatal interview and which was added to the mother's informed consent form. The individual cases from different study databases are identified only by the six-digit *token* identification number. Therefore, personal data cannot be extracted from the study datasets in such a way that a respondent can be identified. Personal data collected during each wave of data collection and transferred from administrative data systems are stored in separate research databases.

Researchers carry out scientific research and statistical calculations based on the study database, and the results of their research are published. However, these scientific research and statistical activities are not aimed at using individually identifiable personal data – and still less at publishing any identifiable personal data in official publications. As we have emphasized several times, the informed consent forms are handled and archived separately from the study database. The personal data of the sample of the Cohort '18 Study can only be accessed by HDRI's researchers and survey staff responsible for data management or, in the case of a data-processing agreement, by designated staff of the data-processing company. In the latter case, access is subject to a confidentiality agreement having been signed. HDRI intends to provide free access to external scholars and researchers on request, after the publication of the Cohort '18 Study reports, and after the data-use protocol agreement is signed by the external users. Such access will not include access to personal data.

12 https://net.jogtar.hu/jogszabaly?docid=a1100031.kim – Section 2.9.1.

6. DATA QUALITY: PRESENTATION OF SOME BACKGROUND DATA FROM INTERVIEWS

6.1. THE SAMPLE OF THE COHORT '18 STUDY

The first step in the sampling procedure for the Cohort '18 Study was the selection of 628 health visitor districts as primary sampling units. Of these, 77 did not agree to participate, but they were replaced by 57 districts before the start of the fieldwork, and 20 districts were added later. For further details, see Kapitány (2018). The number of health visitors changed constantly during the fieldwork, as a consequence of retirement, childbearing or illness; also additional health visitors were involved in interviewing those mothers-to-be who moved to another health visitor district. As a result, a total of *578 health visitors* took part in the fieldwork for the prenatal wave.

In total, 8,717 women participated in the study, with 8,844 foetuses (*Table 6*).

Table 6
Response outcomes, Cohort '18 Study

	Expectant mothers	Foetuses
Total sample participating in the study	8,717	8,844
Number of expectant mothers contacted	8,319	8,441
Of these:		
Foreign-language pregnancy questionnaire	16	
Lack of informed consent form (retrospective refusal)	5	
Incomplete questionnaire	6	
Out of sample	4	
Death of a foetus or infant	1	
Number of expectant mothers who answered the questionnaire	8,287	8,409
Number of women with retrospective prenatal proxy interviews during the 6-month survey	383	

Source: Cohort '18 Study prenatal wave database.

The 8,717 women represent the total sample of the longitudinal Cohort '18 Study. The figure includes both those women who participated in the prenatal data collection wave, and also some who became involved later, during the 6-month survey.[13] A total of 8,319 pregnant women were contacted during the prenatal wave fieldwork, and 8,287 of them (with 8,409 foetuses) answered the prenatal wave questionnaire.

6.2. NUMBER OF INTERVIEWS PER HEALTH VISITOR

The rate of both achievement and progress varied from district to district, but the number of interviews that we expected also differed fundamentally in each health visitor district. Each health visitor conducted a different number of interviews: there were some who carried out just one interview (e.g. because of the pregnant mother's relocation and follow-up), while there were others who conducted 80 interviews during the fieldwork (e.g. because they stepped in for another health visitor). Typically, however, the health visitors conducted 6–12 interviews.

13 We can already see how many respondents fell out (possibly) permanently after the prenatal survey (165 people) and after the 6-month survey (22 people); also, those who did not answer the 6-month questionnaire, but plan to continue later (652 people); or who did not become sample members because of the violation of some criterion (29 persons). Based on our current data, 7,849 people are continuously present in the sample and were queried during the next waves.

6.3. AVERAGE RESPONSE RATES PER QUESTIONNAIRE (UNIT AND ITEM RESPONSE RATES)

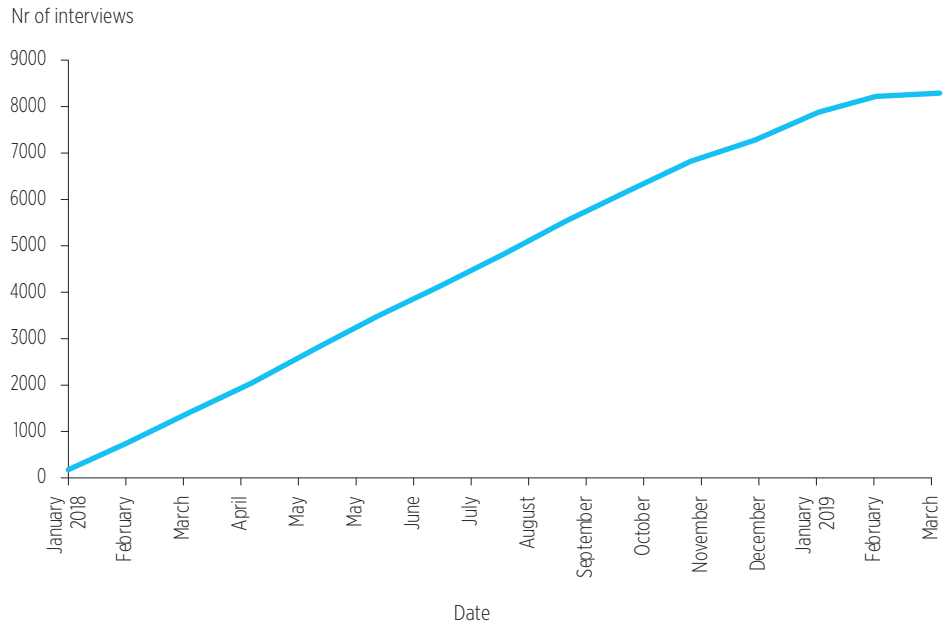
If all respondents had had to answer all the questions, a mother-to-be would have been asked a total of 561 questions: that is the total number of variables computed from the questions in the survey. However, as we know, not everyone had to answer every question: if someone was not affected by a particular situation, we did not ask unnecessary questions. The number of questions that required a response ranged from 224 to 385: that is, no pregnant woman needed to be asked all the questions. Thus, the number of 'don't know' and 'don't wish to answer' responses were always compared to the number of questions that had to be put to a given woman when calculating the *unit non-response rate* (i.e. 100% is based on the number of questions/variables that the woman had to answer). While the rate of 'don't knows' ranged from 0% to 17.6%, the rate of 'don't wish to answer' responses ranged from 0% to 37.2%, and the overall rate of 'don't know or don't wish to answer' ranged from 0% to 46.3%.

If, instead of looking at the case-by-case answers, we change our focus to calculate the rate of 'don't know' or 'don't wish to answer' categories at the level of the individual variables (*item non-response rate*), then we find that, for example, the women frequently could not say how much they weighed when they were born (17.2%) or did not want to say what they thought was the ideal age nowadays for a woman to give birth to her first child (15.5%).

The dataset of the self-administered questionnaire was missing entirely only in the case of 96 expectant mothers (1.2% of respondents). In the self-administered questionnaire, we did not use 'don't know' and 'don't wish to answer', but respondents occasionally left certain questions blank. Question skips were not used in this questionnaire, but there were questions that we did not expect everyone to answer – for example, questions about their relationship situation. Theoretically, the 'required response number' was set differently for single expectant mothers (required to answer 78 questions) and for those with a partner (96 questions). Accordingly, 63.7% of the women answered every self-administered question; 3.3% answered less than 80% of the questions; and 1.6% answered less than 50% (this figure includes the 1.2% who omitted to fill in any of the questionnaire).

6.4. FIELDWORK PROGRESS

The initial sample of the Cohort '18 Study consisted of children (or rather foetuses) whose expected date of birth was between 1 April 2018 and 30 April 2019. However, it was also a selection criterion that the expectant mother had to be in the 28th–31st week of her pregnancy. Health visitors had to carry out the interviews in that time window, so that the foetuses were at approximately the same stage of growth and development. If the health visitors thought that the pregnant woman was at risk or could not be interviewed in the 28th–31st week of pregnancy for some other reason, they could reschedule the interviews for earlier (or later). Consequently, the prenatal wave fieldwork started on 4 January 2018 and ended on 2 April 2019. The progress of interviews is shown in *Figure 3*. Babies born in January turned 6 months in July 2018; thus, from that point on, the health visitors had to schedule the 6-month interviews in parallel with the prenatal interviews. Also, in June 2018, we introduced supplementary payments for cases that were difficult to interview: if the woman had completed at most eight years of primary schooling.



Source: Cohort '18 Study, prenatal wave database (mother), N=8287.

Figure 3
Fieldwork progress 01/2018 and 04/2019

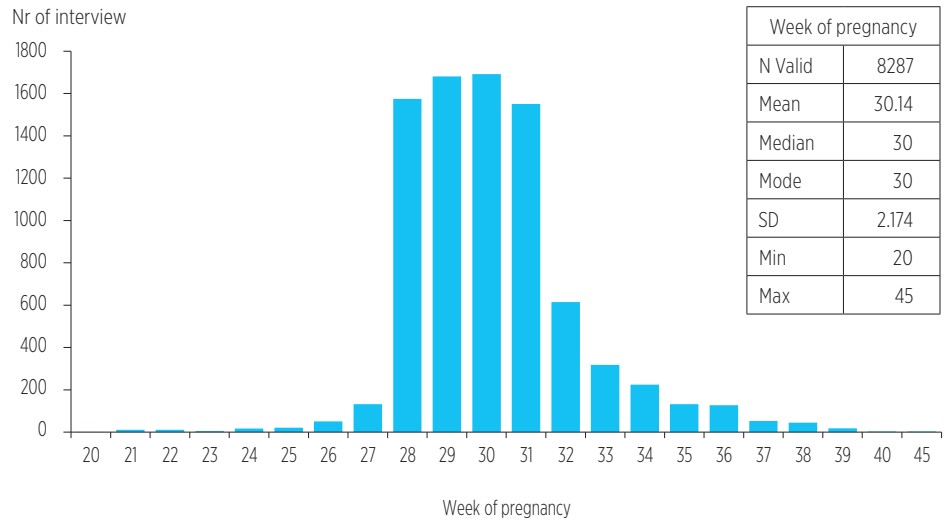
As already indicated, we asked the health visitors to interview expectant mothers in the 28th–31st week of their pregnancy, so that the foetuses were at approximately the same stage of growth and development. Therefore, right from the initial briefing, we drew the attention of the health visitors to the need to track the weeks of pregnancy of each expectant mother from the beginning of the pregnancy period, and to try to inform mothers about the study before their 28th week, to ask for their participation and to arrange a date for the interview.

However, it was not always possible to meet the required interview date: various factors could prevent the expectant mother or the health visitor from meeting for the interview during the four-week window designated for the fieldwork.[14] Despite the difficulties encountered, the majority of the surveys did take place within the indicated four weeks: 78% of interviews were successfully conducted during the 28th–31st week of pregnancy (*Figure 4*).

6.5. LENGTH OF INTERVIEW

The interviews based on the main questionnaire took an average of 52 minutes (SD=19 minutes). The modal value of the distribution was 45 minutes (i.e. for most expectant mothers it took 45 minutes to answer the main questionnaire) (*Figure 5*). In addition, there was the time taken to fill in the informed consent form, the self-administered questionnaire and the prenatal care booklet: that could take the total time to about one and a half hours.

14 In the open-ended question seeking comments at the end of the questionnaire, the health visitors conscientiously noted and indicated all those whose interviews could not be completed within the allotted four weeks. The HDRI research team has decided to keep these cases in the database, and to select the subsamples they need for the current health and psychological analyses later on, based on the occurring needs of the study.

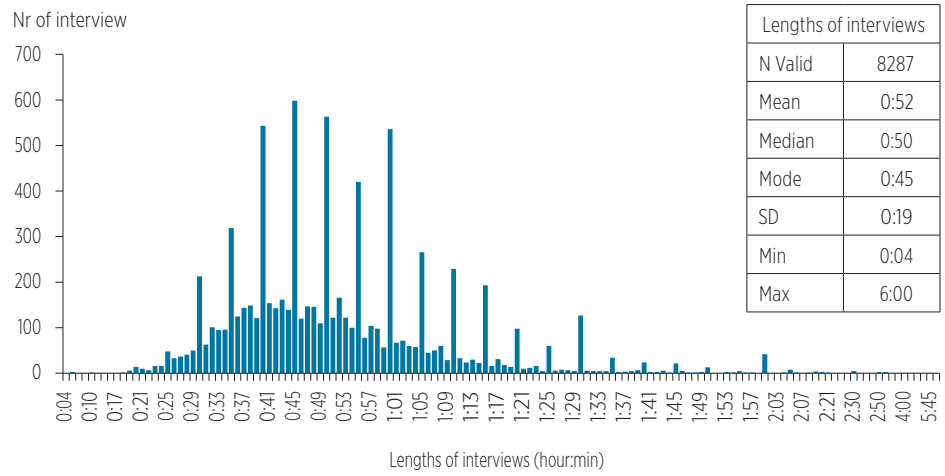


Week of pregnancy	
N Valid	8287
Mean	30.14
Median	30
Mode	30
SD	2.174
Min	20
Max	45

Source: Cohort '18 Study, prenatal wave database (mother), N=8287.

Figure 4
The number of interviews by week of pregnancy

In 292 cases (4%) the interview took less than 30 minutes, whereas in 50 cases (1%) it required more than 2 hours. The length of interview was less than 30 minutes (very short) among women who had completed at most eight years of primary schooling, who were under the age of 25, living in a cohabiting relationship, still working, and living in Jász-Nagykun-Szolnok or Szabolcs-Szatmár-Bereg county; and among women whose answers to the questionnaire were immediately entered online (and not first on paper), who answered the self-administered questionnaire with the help of the health visitor, completed the interview without interruption, were not hindered by any vision, hearing or other impairment and answered the questionnaire in the health visitor's office (Appendix 8.4, second column).



Lengths of interviews	
N Valid	8287
Mean	0:52
Median	0:50
Mode	0:45
SD	0:19
Min	0:04
Max	6:00

Source: Cohort '18 Study, prenatal wave database (mother), N=8287

Figure 5
Distribution of interviews by their length (hours:minutes)

Interviews tended to last more than 2 hours if: the woman was interviewed using a paper-based questionnaire; she could not complete the self-administered questionnaire on her own, but only with the help of a health visitor; the interview was interrupted; the woman was hampered by something in her responses (hearing, seeing, speaking, or some other difficulty); or if more than three children lived in the household (Appendix 8.4). *Figure 5* shows that for a relatively large number of interviews, the health visitors entered whole hours, half and quarter hours, thus probably not indicating the *exact* length in hours and minutes, but rather a rounded value.

6.6. DATA FROM THE HEALTH VISITOR'S QUESTIONNAIRE

Some information from the health visitor's questionnaire is summarized in the table in Appendix 8.4 (second column). This shows that more than half of the interviews (56%) were conducted in the office of the health visitor and 43% in the home of the expectant mother. Other locations included the woman's workplace, the central health visitor's office, a doctor's surgery and the mayor's office. Two thirds (67%) of the interviews were carried out using a paper-based questionnaire, while one third (33%) were entered directly by the health visitors during the interview, using the online interface. (Some 13% of the questionnaires administered in the woman's home were entered online during the interview, compared to 48% conducted in the health visitor's office.) The vast majority of interviews (82%) were conducted without interruption.

Both in the interviewer's handbook and during the briefing, we emphasized that an appointment should be made for the interviews, as it was necessary to create the right conditions for a smooth interview. Despite our request to the contrary, a quarter of the women (26%) were interviewed in the presence of someone else – most often a child under the age of 6 (16%) or a spouse/partner (8%). When asked whether anything had bothered the woman when answering, the most common thing mentioned by health visitors was that the interviewee had not understood (or had had difficulty in understanding) some of the questions (4%). For those who had language difficulties, in some cases their partner assisted by interpreting; but some had received help from their mother-in-law, a friend, a sibling or an interpreter.

The vast majority of the women (81%) completed the self-administered questionnaire on their own, as it contained sensitive questions. However, if the expectant mother requested it, the health visitor could help: 14% of the women asked for help in answering only a few questions, but 5% received assistance in filling in the entire self-administered questionnaire (Appendix 8.4).

7. RESEARCH ETHICS GUIDELINES

7.1. LEGAL FRAMEWORK CONCERNING THE STUDY

The legal framework for the Cohort '18 Study is set out in the following legislation:

- The Basic Law of Hungary
- Legislation 2016/679 of the European Parliament and of the EU Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing legislation no. 95/46/EC (hereafter GDPR)
- Act V of 2013 on the Civil Code (Civil Code)
- Act CXII of 2011 on the right to self-determination of information and freedom of information (Information Act)
- Act XLVII of 1997 on the handling and protection of health and related personal data (Act on Health Privacy)
- Act LXXVI of 1999 on Copyright Law (Copyright Law)
- Act I of 2012 on the Labour Code (Labour Code)
- Act XXXIII of 1992 on the Status of Civil Servants (Civil Servant Act)
- Act XXXIII of 1992 on the Status of Civil Servants and Government Decree 84/2011 (V.26.) on the implementation and supervision of budgetary bodies and other research and research supplementary institutes placing them under the authority of the Hungarian Academy of Sciences.

The GDPR guidelines, which provide the regulatory framework for research data management, set different – less stringent – conditions for personal data processing for scientific studies and statistical purposes. According to this, the processing of personal data for the purposes of scientific research should be interpreted broadly to include, inter alia, technological development and demonstration activities, basic research, applied research, and privately funded research. According to the GDPR, scientific research in the field of public health that is in the public interest should also be seen as such. Thus, the collection and processing of personal data for the purposes of statistical surveys or the calculation of statistical results is considered to be for statistical purposes. These statistical results can later be used for a variety of purposes, including scientific research. The personal data processing of the Cohort '18 Study is considered to be for scientific research and statistical purposes. The study applies the special rules of the GDPR regarding data processing for scientific research and statistical purposes, while regularly monitoring the changes in EU-level and national personal data-processing regulations, conditions and guarantees.

7.2. PROTECTION OF PERSONAL DATA

HDRI, as the data manager of the Cohort '18 Study, has taken a number of technical and organizational measures to ensure the appropriate level of security regarding the handling and storage of personal data; these are covered in detail by the study's Personal Data Protection and Data Management Protocol.[15] The main data protection elements of the study include pseudonymization (in our case, this was done using unique identification number assignment), the physical and IT security of data storage, and the ongoing review of these systems.

Due to the fact that the first and second waves of data collection were partly carried out using a paper-based questionnaire, we also took care in handling the completed

15 Hungarian Central Statistical Office Demographic Research Institute, Growing Up in Hungary - Cohort '18 Study, Personal Data Protection and Data Management Protocol, January 2018.

paper-based documents, when concluding these study phases. All paper-based research materials were collected and sorted. The statements of consent containing personal data and the address cards linked to the statements have been archived and are kept in a safe place by HDRI. Completed paper-based questionnaires (including the self-administered questionnaires) are stored for a time and then destroyed once the information they contain has been recorded and the verification phase has been completed.

7.3. PROVIDING DATA-PROCESSING INFORMATION AND COLLECTING CONSENT

The first, prenatal data collection wave of the Cohort '18 Study was also the period when the birth cohort of the study was built up for longitudinal follow-up. The expectant mothers, in the 28th–31st week of pregnancy, were contacted through the health visitor network. The personal data provided by the respondents during the collection of the main questionnaire and the self-administered questionnaire, as well as the personal data transferred by the health visitor from the prenatal care booklet, are handled by us, through the health visitors, who acted as data processors under individual contracts concluded with HDRI, the data manager. The legal basis for data processing is provided by the informed consent form signed by the respondents.

It was the responsibility of the health visitors to inform the expectant mothers they visited and to ask for their consent, which had to be forthcoming before data collection could start. The materials supporting this were the information booklet, the official letter of invitation and a Personal Data Protection and Data Management Protocol (Appendix 8.2.1), which aims to explain to respondents the legal framework ensuring their lawful participation in a comprehensible way, in particular citing data protection guarantees.

Following the provision of information, the consent form was signed (Appendix 8.2.2); this allows the personal data of the women to be used for research purposes and to be linked to other data sources throughout the survey. The informed consent form also included the address recording form, which was used to provide contact details. It will facilitate the continuity of interviews and allow the databases containing various individual responses to be merged, through the introduction of the six-digit unique identification number. For respondents who were not yet of legal age, an amended version of the informed consent form was prepared and was signed by their legal guardian.

As well as agreeing to participate in the study, by signing the informed consent form the respondent also consented to the transfer of personal data on herself and her child from the Birth Notification System operated by the National Centre for Public Health. The basis for the data linkage was provided by the respondent's social security identification number.

Without a valid, signed informed consent form, the data collection was considered invalid, and any cases entered on the system were not included in the study database.

7.4. DATA PROTECTION GUARANTEES

During the Cohort '18 Study, HDRI, as the data manager, always carries out its personal data-processing activities with the explicit consent of the data subject, in accordance with the provisions of Article 6 (1/A) of the GDPR and Article 9 (2/A) of the GDPR. Given that most of the personal data processed during the Cohort '18 Study relate to children aged 0–3, in their case the parent exercising parental guardianship is entitled to consent in their name, in accordance with the legal provisions in force. The respondent has the right to withdraw her consent at any time, by formal request.

The data manager shall ensure that the data subjects' statements of consent are kept for an adequate time, filed and recorded – whether electronically or on paper.

According to Chapter III of the GDPR, the data manager is required to ensure that the rights of respondents are upheld, through the following procedures:

- Providing transparent information
- Handling requests
- Right to access for data subjects
- Right to rectification
- Right of cancellation
- The right to restrict data processing
- The right of data subjects to data portability
- The right to object to data processing
- The right to appeal

7.5. CONDITIONS OF DATA TRANSFER

During the study, the use of external data processors is sometimes necessary for data collection, data entry or data processing. When this occurs, it is also necessary to ensure the legal guarantees concerning proper data management on the part of HDRI, which requires a number of technical and organizational measures to be in place to ensure the protection of the rights of respondents when handled by third parties.

In accordance with Article 28 of the GDPR, HDRI (the data manager) enters into a data-processing agreement with its data processors, which sets out the subject, duration, nature and purpose of data processing, the type of personal data, categories of data subjects and the obligations and rights of the data manager and the data processor. It further states that personal data are processed by the data processor only to the extent of the written instructions of the data manager. It also ensures that confidentiality is respected by the data processor and any authorized persons employed by the data processor. It specifies the security, technical and organizational measures to be applied by the processor. It sets out the procedure to be followed in the event of a so-called data protection breach and the necessary protocols of cooperation. It requires the deletion of personal data and copies upon completion of the service. And it obliges the processor to cooperate in the event of an audit or on-site inspection.

7.6. ETHICS COMMITTEE

In addition to the legal frameworks detailed above, since the start of the data collection we have been assisted in our work by a three-member Ethics Committee, which ensures that the handling of both personal and survey data within the Cohort '18 Study meets the ethical requirements of scientific research. The task of the Ethics Committee of the Cohort '18 Study is to provide ongoing background support during the research, by commenting on and validating methodological decisions and research ethics issues with regard to data use. On the initiative of HDRI, the committee has an ad hoc opinion-providing role at the planning phase of each stage of the study, and in ethical issues concerning the management and use of data. Resolutions, approvals and ethical permits may be requested from the committee in relation to certain ethical remarks and procedural research decisions directly related to the Cohort '18 Study. Requests for resolutions, approvals and ethical opinions may be initiated by external actors, such as stakeholders, employees, researchers, analysts and certain research bodies (e.g. the Scientific Advisory Board), during which HDRI acts as an intermediary.

8. APPENDIX

8.1. QUESTIONNAIRES OF THE PRENATAL WAVE

8.1.1. Main questionnaire

Web: https://demografia.hu/en/publicationsonline/index.php/workingpapers/article/view/974/747

8.1.2. Self-administered questionnaire

Web: https://demografia.hu/en/publicationsonline/index.php/workingpapers/article/view/974/748

8.2. RESEARCH DOCUMENTS

8.2.1. Personal data protection and data management information

of the Hungarian Central Statistical Office Demographic Research Institute, Growing Up in Hungary - Cohort '18 Study
Web: https://demografia.hu/en/publicationsonline/index.php/workingpapers/article/view/974/749

8.2.2. Informed consent

to participate in the research **Cohort '18 - Hungarian Birth Cohort Study** conducted by the Hungarian Demographic Research Institute
Web: https://demografia.hu/en/publicationsonline/index.php/workingpapers/article/view/974/750

8.2.3. Letter of invitation

Web: https://demografia.hu/en/publicationsonline/index.php/workingpapers/article/view/974/751

8.3. SELF-REPORTED SCALES

Web: https://demografia.hu/en/publicationsonline/index.php/workingpapers/article/view/974/752

8.4. DISTRIBUTION OF THE LENGTH OF INTERVIEWS BY SOME SELECTED BACKGROUND VARIABLES

Table 7

Distribution of the length of interviews by some selected background variables

	Distribution of the variable (col %)	Length of interview within a given category (row %; average, SD)				
		Less than 30 minutes	30–120 minutes	More than 2 hours	Average	SD
Total sample, N=8287	100.0%	3.5%	95.9%	0.6%	0:52	0:19
Type of query						
Paper-based	67.2%	2.2%	97.1%	0.7%	0:55	0:20
Online	32.8%	6.3%	93.4%	0.4%	0:47	0:16
Assistance in completing the self-administered questionnaire						
Completed individually	81.3%	3.8%	95.8%	0.5%	0:51	0:17
Asked for assistance regarding some questions	14.2%	1.4%	97.7%	0.9%	0:57	0:21
Health visitor asked, expectant filled in	1.0%	2.5%	94.9%	2.5%	1:08	0:29
Health visitor asked and filled in	3.5%	7.2%	90.4%	2.4%	0:59	0:30
Interruption of interview						
No, queried in one go	81.7%	4.0%	95.6%	0.5%	0:51	0:18
Yes, but only for a shorter period	18.1%	1.4%	97.5%	1.1%	0:57	0:20
Yes, for longer period (couple of hours, days)	0.2%	6.7%	66.7%	26.7%	1:44	1:44
Limitations of respondent						
Hindered (vision, hearing, other)	5.4%	1.6%	95.8%	2.6%	1:04	0:27
Was not hindered by anything	94.6%	3.6%	95.9%	0.5%	0:52	0:18
Location of the interview						
In the home of the expectant mother	43.1%	1.4%	98.0%	0.6%	0:55	0:19
In the office of the health visitor	56.2%	5.2%	94.2%	0.6%	0:50	0:19
Other location	0.7%	1.8%	96.4%	1.8%	0:51	0:16
Type of pregnancy						
One foetus	98.5%	3.6%	95.9%	0.6%	0:52	0:19
Twin pregnancy	1.4%	1.7%	96.7%	1.7%	0:56	0:18
Triplet (or more) pregnancy	0.0%		100.0%		1:40	

Note: Significantly higher values compared to the total sample with regards to the length of a given interview (0–29 minutes; 30–120 minutes; 120+ minutes) are highlighted with a grey background.
Source: Cohort '18 Study, prenatal database (mother), N=8287.

Table 7
Distribution of the length of interviews by some selected background variables (continued)

	Distribution of the variable (col %)	Length of interview within a given category (row %; average, SD)				
		Less than 30 minutes	30–120 minutes	More than 2 hours	Average	SD
Educational attainment of women						
Primary	29.6%	4.7%	94.5%	0.8%	0:54	0:20
Secondary	29.8%	2.8%	96.8%	0.5%	0:52	0:20
Tertiary	40.7%	3.2%	96.2%	0.6%	0:51	0:17
Age of women at the time of the interview						
13–25	23.0%	5.5%	93.8%	0.6%	0:53	0:19
26–34	53.9%	2.9%	96.3%	0.7%	0:52	0:19
35–49	23.1%	2.9%	96.8%	0.3%	0:53	0:17
Real relationship status						
Married	54.3%	3.0%	96.4%	0.6%	0:52	0:19
In a cohabiting relationship	41.8%	4.2%	95.3%	0.6%	0:53	0:18
Single or in visiting relationship	3.9%	4.6%	94.5%	0.9%	0:54	0:19
Labour market status of women						
Still working	20.8%	4.8%	94.4%	0.8%	0:50	0:19
Not working	79.2%	3.2%	96.2%	0.6%	0:53	0:19
Number of live births						
Childless (parity: 1)	48.1%	3.8%	95.6%	0.6%	0:51	0:18
Already existing child/children	51.9%	3.3%	96.1%	0.7%	0:53	0:19
Total number of children in the household						
0	49.3%	3.8%	95.6%	0.6%	0:52	0:18
1	33.1%	3.7%	95.7%	0.5%	0:52	0:19
2	12.7%	2.2%	97.4%	0.4%	0:53	0:16
3+	4.8%	3.0%	94.7%	2.3%	0:58	0:26
Total number of people in household						
1	0.9%	2.8%	97.2%		0:50	0:16
2	39.6%	3.6%	95.8%	0.6%	0:51	0:18
3	31.7%	3.5%	96.0%	0.5%	0:52	0:19
4+	27.9%	3.4%	95.8%	0.8%	0:54	0:20

Note: Significantly higher values compared to the total sample with regards to the length of a given interview (0–29 minutes; 30–120 minutes; 120+ minutes) are highlighted with a grey background.
Source: Cohort '18 Study, prenatal database (mother), N=8287.

REFERENCES

Amato, P.R. (2000). The consequences of divorce for adults and children. *Journal of Marriage and Family*, *62*(4), 1269–1287.

Barber, J. S., Axinn, W. G. and Thornton, A. (2002). The influence of attitudes on family formation processes. *Meaning and choice: Value orientations and life course decisions*, 37, 45–93.

Blaskó, Zs. (2009). Születési kohorsz-vizsgálatok itthon és külföldön–tapasztalatok és tanulságok [Birth cohort investigations at home and abroad – experiences and lessons learned]. *Esély*, *6*, 45–75.

Boros, J. (2018). Training sessions and their background materials. In: Zs. Veroszta (ed.), Technical report. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study. Prenatal research, preparational phase. Working Papers on Population, Family and Welfare, No. 30. Budapest: HDRI, 41 Buchmann, M. (1989). *The Script of Life in Modern Society: Entry into adulthood in a changing world*. Chicago: University of Chicago Press.

Buchmann, M.C. and Kriesi, I. (2011). Transition to adulthood in Europe. *Annual Review of Sociology*, *37*, 481–503.

Bukodi, E., Goldthorpe, J.H., Waller, L. and Kuha, J. (2015). The mobility problem in Britain: New findings from the analysis of birth cohort data. *British Journal of Sociology*, *66*(1), 93–117.

Elder, G.H., Johnson, M.K. and Crosnoe, R. (2003). The emergence and development of life course theory. In: J.T. Mortimer and M.J. Shanahan (eds), *Handbook of the Life Course*. Boston, MA: Springer, 3–19

Gabler, S. (1994). Eine allgemeine Formel zur Anpassung an Randtabellen [A general formula for adapting to margin tables]. In: S. Gabler, J.H.P. Hoffmeyer-Zlotnik and D. Krebs (eds), *Gewichtung in der Umfragepraxis* [*Weighting in Survey Practice*]. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 88–105.

Gilford, R. and Bengtson, V. (1979). Measuring marital satisfaction in three generations: Positive and negative dimensions. *Journal of Marriage and Family*, *41*(2), 387–398. https://doi.org/10.2307/351705

Greenfield, E.A. and Marks, N.F. (2006). Linked lives: Adult children's problems and their parents' psychological and relational well-being. *Journal of Marriage and Family*, *68*(2), 442–454.

Gresits, G. (2018). Focus group fundamental research. In: Zs. Veroszta (ed), *Conceptual Framework. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study*. Working Papers on Population, Family and Welfare, No. 32. Budapest: HDRI, 39–50.

Hagestad, G.O. and Neugarten, B.L. (1985). Age and the life course. In: E. Shanas and R. Binstock (eds), *Handbook of Aging and the Social Sciences.* New York: Van Nostrand-Reinhold.

Horváth, B. and Mihályffy, L. (2008). Hibaszámítás jackknife módszerrel bonyolult felépítésű, kalibrált minták esetén [Error calculation by jackknife method for complex calibrated samples]. *Statisztikai Szemle*, *86*(6), 591–613.

Kapitány, B. (2018). The target population and the sampling. In: Zs. Veroszta (ed.), *Technical Report. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study. Prenatal research, preparational phase*. Working Papers on Population, Family and Welfare, No. 30. Budapest: HDRI, 7–12.

Kopcsó, K. (2018). Scale adaptation. In: Zs. Veroszta (ed.), *Technical Report. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study. Prenatal research, preparational phase*. Working Papers on Population, Family and Welfare, No. 30. Budapest: HDRI, 24–29.

Mayer, K.U. (2004). The paradox of global social change and national path dependencies: Life course patterns in advanced societies. In: M. Kohli and A. Woodward (eds), *Inclusions and Exclusions in European Societies*. London and New York: Routledge, 103–124.

McLeod, J.D. and Almazan, E.P. (2003). Connections between childhood and adulthood. In: J.T. Mortimer and M.J. Shanahan (eds), *Handbook of the Life Course*. Boston, MA: Springer, 391–411.

Mills, M.C. and Tropf, F.C. (2015). The biodemography of fertility: A review and future research frontiers. *KZfSS – Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *67*(1), 397–424.

Rini, C.K., Dunkel-Schetter, C., Wadhwa, P.D. and Sandman, C.A. (1999). Psychological adaptation and birth outcomes: The role of personal resources, stress, and sociocultural context in pregnancy. *Health Psychology*, *18*(4), 333–345.

Rohr, A. (2018). Providing information technology background support. In: Zs. Veroszta (ed.), *Technical report. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study. Prenatal research, preparational phase*. Working Papers on Population, Family and Welfare, No. 30. Budapest: HDRI, 38–40.

Rudas, T. (2006). *Közvélemény-kutatás – Értelmezés és Kritika* [*Public Opinion Research – Interpretation and criticism*]. Budapest: Corvina.

Settersten, R.A. (2003). Age structuring and the rhythm of the life course. In: J.T. Mortimer and M.J. Shanahan (eds), *Handbook of the Life Course*. Boston, MA: Springer, 81–98.

Solon, G., Haider, S.J. and Wooldridge, J.M. (2015). What are we weighting for? *Journal of Human Resources*, *50*(2), 301–316. https://doi.org/10.3368/jhr.50.2.301

Szabó, L. (2018). Testing the questionnaires: The pilot survey. In: Zs. Veroszta (ed.), *Technical report. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study. Prenatal research, preparational phase*. Working Papers on Population, Family and Welfare, No. 30. Budapest: HDRI, 30–37.

Szelényi, B. (2003). Az adatbázis súlyozása [Database weighting]. In: Balázs Kapitány (ed.), *Módszertan és dokumentáció az „Életünk fordulópontjai" című demográfiai követéses vizsgálat első hullámának adatfelvételének ismertetése* [*Methodology and documentation: description of the data collection of the first wave of the demographic tracking study 'Turning points in our lives'*]. Műhelytanulmányok 2. Budapest: KSH Népességtudományi Kutatóintézet, 64–81.

Veroszta, Zs. (ed.) (2018a). Technical Report. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study. Prenatal research, preparational phase. *Working Papers on Population, Family and Welfare*, No. 30. Budapest: HDRI.

Veroszta, Zs. (2018b). The steps of fundamental technical work. In: Zs. Veroszta (ed.), Technical Report. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study. Prenatal research, preparational phase. *Working Papers on Population, Family and Welfare,* No. 30. Budapest: HDRI, 18–23.

Veroszta, Zs. (ed.) (2019). *Conceptual Framework. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study*. *Working Papers on Population, Family and Welfare,* No. 32. Budapest: HDRI.

LIST OF WORKING PAPERS

1. László Hablicsek, Pál Péter Tóth: The Role of International Migration in Maintaining the Population Size of Hungary between 2000–2050
2. Maritetta Pongrácz: Birth out of Wedlock
3. Attila Melegh: East/West Exclusions and Discourses on Population in the 20th Century
4. Zsolt Spéder: Fertility and Structural Change in Hungary
5. Sándor Illés: Foreigners in Hungary: Migration from the European Union
6. Magdalena Muszyńska: Family Models in Europe in the Context of Women's Status
7. Attila Melegh, Elena Kondratieva, Perttu Salmenhaare, Annika Forsander, László Hablicsek, Adrienn Hegyesi: Globalisation, Ethnicity and Migration. The Comparison of Finland, Hungary and Russia
8. Zsolt Spéder, Balázs Kapitány: Poverty and Deprivation: Assessing Demographic and Social Structural Factors
9. Etelka Daróczi: Ageing and Health in the Transition Countries of Europe – the Case of Hungary
10. Péter Őri: Demographic Patterns and Transitions in 18–20th Century Hungary. County Pest-Pilis-Solt-Kiskun in the Late 18th and Early 20th Centuries
11. Zsolt Spéder, Balázs Kaptány: Ideational Factors and Parenthood. A Gender- and Parity Specific Analysis in a Post-Communist Society
12. Irén Gödri: The Role of Ethnicity and Social Capital in Immigration to Hungary
13. Attila Melegh, Arland Thornton, Dimiter Philipov, Linda Young-DeMarco: Mapping Societal Developmental Hierarchies in Europe: a Bulgarian Perspective
14. Balázs Kapitány, Zsolt Spéder: Factors Affecting the Realisation of Child-Bearing Intentions in Four European Countries
15. Zsolt Spéder, Balázs Kapitány: Realising Birth Intention in European Comparison – Understanding the Post-Communist Fertility Transition
16. Tamás Faragó: Historical Demography in Hungary: a History of Research
17. Attila Melegh: Net Migration and Historical Development in Southeastern Europe since 1950
18. Róbert I. Gál, Endre Szabó, Lili Vargha: The age profile of invisible transfers: the true size of asymmetry in inter-age reallocations
19. Irén Gödri, Béla Soltész, Boróka Bodacz-Nagy: Immigration or emigration country? Migration trends and their socio-economic background in Hungary: A longer-term historical perspective
20. Péter Őri, Levente Pakot: Residence patterns in nineteenth century Hungary: Evidence from the Hungarian MOSAIC sample.
21. Zsuzsa Blaskó: Studying emigration by extending a large-scale household survey. Methodology, evaluation and descriptive findings
22. Lili Vargha, Róbert Iván Gál, Michelle O. Crosby-Nagy: Household production and consumption over the lifecycle: the National Time Transfer Accounts in 14 European countries
23. Kálmán Joubert, Gyula Gyenis: The Hungarian Longitudinal Growth Study: From birth to the age of 18 years
24. Róbert Iván Gál, Árpád Törzsök: The Savings Gap in Hungary
25. Lajos Bálint: Suicide in the Hungarian Kingdom
26. Róbert Iván Gál, Pieter Vanhuysse, Lili Vargha: Pro-elderly welfare states within pro-child societies: Incorporating family cash and time into intergenerational transfers analysis
27. Zsolt Spéder, Lívia Murinkó, Lívia Sz. Oláh: Sailing close to the wind?
    The effects of third birth policies in post-communist Hungary
28. Róbert Iván Gál, Márton Medgyesi: Financing the lifecycle or mitigating poverty: Redistribution in the Hungarian welfare system by age and income
29. Irén Gödri, Gábor Attila Feleky: Selection of migrants and realization of migration intentions – lessons from a panel study
30. Zsuzsanna Veroszta (ed.): Technical Report. Growing Up in Hungary - Cohort '18 Hungarian Birth Cohort Study. Prenatal Research, Preparational phase.

31. Róbert Iván Gál, Márta Radó: Participation and Postponed Retirement in Central and Eastern Europe
32. Zsuzsanna Veroszta (ed.): Conceptual framework. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study
33. Mikołaj Szołtysek, Bartosz Ogórek: Global and local correlations of Hajnal's historical household formation markers: a geographically weighted approach
34. Laura Szabó, Igor Kiss, Branislav Šprocha, Zsolt Spéder: Fertility of Roma minorities in Central and Eastern Europe
35. Róbert Iván Gál, Márton Medgyesi, Pieter Vanhuysse: The transfer cost of parenthood in Europe
36. Csaba G. Tóth: Multi-population models to handle mortality crises in forecasting mortality: a case study from Hungary
37. Laura Szabó: Mixed-ethnic partnerships and ethnic reproduction among Roma women in Hungary

The above Working Papers can be ordered at the following e-mail address:
workingpapers@demografia.hu

The Working Papers are also available online at:
http://www.demografia.hu/workingpapers