PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cohort profile: the national, longitudinal NASCITA birth cohort study to investigate the health of Italian children and potential influencing factors
AUTHORS	Pandolfini, Chiara; Clavenna, Antonio; Cartabia, Massimo; Campi, Rita; Bonati, Maurizio; Work Group, NASCITA

VERSION 1 – REVIEW

REVIEWER	LeWinn, Kaja University of California San Francisco, Psychiatry
REVIEW RETURNED	13-Jun-2022

GENERAL COMMENTS	This is an impressive undertaking, and this study will likely make many important contributions to the understanding of child health in Italy. For the purposes of a Cohort Profile Paper, there are some aspects of the study design and primary objectives that remain somewhat unclear. I have made a few suggestions below by section of the paper. Intro: Lines 94-98 seem to drastically under-represent decades of work elucidating the social determinants of child health. Perhaps what is missing is an understanding of how these findings, often in other areas of the world, generalize to the Italian population? Line 111: It's unclear if the goal here is to use the 2020 review of birth cohorts only to justify this study. A pregnancy cohort would be a stronger design by including the pregnancy period, so for the purposes of justifying the novelty of this cohort in the broader context, recent pregnancy cohorts with similar plans to follow participants through birth and beyond should be included. Line 126-127: This statement requires references. It would be helpful to have more detail on why some of the other cohort studies in Europe wouldn't be generalizable to the Italian context. Similarly, it seems that there are several ongoing pregnancy/birth cohort studies in Italy specifically. Is what sets this study apart it's focus on nurturing behavior? Do the other Italian studies rely on pediatricians for recruitment? Study design Since the pediatrician seems to be the source of the representativeness of this study, it would be helpful to include more details about the recruitment of pediatricians. How many

	were approached? How many refused? Was there any association between region/area/pediatrician characteristics and refusals? Was any information collected on those parents who did not consent to participate and were they different than those who did consent? Data collection It is stated that all data are collected by the pediatrician, but it's unclear on how the pediatrician gathered data on some of the primary exposures. Later in the paper, it becomes clear that nurturing care refers to positive health behaviors during pregnancy and early childhood. Defining this term more clearly at the outset would be helpful (it is easy to confuse nurturing with constructs like parenting). It is also unclear how other important contextual data are being collected (e.g., sociodemographics such as family income). Are the parents filling out any surveys? More detail needs to be provided here. Measurement Were all measures for this study newly generated by study pediatricians? Were these assessments validated? More clarity on what specific was assessed in the study would help clarify this issue. In addition, it would be helpful if the authors could comment on why they did not rely on established measures for some of the constructs they are interested in. Results For tables 1 and 2, it would be clearer to put the column with the total first since the p-values are referring to comparisons between the North, Centre, and South. After reviewing the results, the authors seem to be most interested in regional differences in early analyses but this focus isn't motivated in the introduction. Could the authors comment on why these differences are important? Is it related to resource allocation? Or potential regional differences related to health care, exposures, etc.? I am also concerned that some of these findings might be stigmatizing. There are numerous differences by region highlighted without adjustment for potential differences in economic resources (which also seem to differ by region). It seems that many of these difference may be explained by differences in socioeconomic resources. There are many descriptive findings presented. Are any of these findings surprising or differ significantly from the other Italian cohorts mentioned in the intro? I would move table 5 earlier in the paper as it is related to the general characteristics of the cohort and how they relate to the nation. Related, despite the notion that using pediatricians to recruit participants would lead to a largely representative sample, there are quite notable differences in sociodemographic characteristics, with lower SES folks less likely to have been recruited into the study. In the context of this universal health care system, can the authors comment on why that might be? (Also see earlier comment regarding providing more information on characteristics of pediatricians that refused to participate).

	Lines 348-350 seem to overstate the similarity between the cohort and national population.

REVIEWER	Dodgeon, Brian University College London Institute of Education, Social Science
REVIEW RETURNED	30-Jul-2022

GENERAL COMMENTS	Well done. I enjoyed reading this cohort profile. In general I felt the cohort study was well described, supplying enough detail for potential researchers to approach the study as a research resource. I therefore recommend acceptance subject to the minor revisions detailed below. The revisions mainly take the form of correcting typographical errors in the text or tables, or minor issues regarding English usage. Line 128 I recommend changing the first line of the paragraph to 'Various cohort studies have been carried out in Italy,...' which is the more conventional way of describing resources such as this. Line 137 Typo error: 'recruited them at in eight...' should be 'recruited them in eight...' Line 229 I recommend changing the sentence to 'The majority of mothers and fathers were born in Italy...' Although the word 'majority' is singular, the mothers and fathers who are referred to are many in number, and so I'd say the common English usage is to use the plural form of the verb. Line 233 (similar to above) I recommend 'but the majority were employed...' Line 287 This line mis-quotes the figure in Table 3, which is '2474 (49.0%) were female'. Line 302 This line mis-quotes the figure in Table 4, which is 'median of 24.9 days...' Line 304 (similar to lines 229 & 233) I recommend plural usage: '...over one-third (35.9%) were receiving formula milk and two-thirds (64.1%) were receiving a mixed-feeding regimen.' Line 307 In Table 4, the first line is wrongly labelled 'Mean age, days, SD', when it is clear from the figures that 25.7 is (correctly) the mean but 24.9 is the median, not the standard deviation. Line 307 Also in Table 4, the percentages for the three different categories under Breast feeding are misleadingly inconsistent. In the columns 'North', 'Centre' and 'South' the three respective percentages for 'Exclusive', 'Mixed' and 'Artificial' add up to 100%, whereas in the column 'Total n(%)' the figure of 63.6% represents the 'Exclusive' number as a percentage of all cohort members, but the two following figures represent the 'Mixed' and 'Artificial' numbers expressed as percentages of 'the remainder' (ie of all those not exclusively breastfed). So I recommend the % figures of 64.1 and 35.9 be replaced by 23.3 and 13.1 respectively. Line 369 Recommend inserting the word 'at', to read '...in particular when looking at southern Italy.'

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Kaja LeWinn, University of California San Francisco, University of California San Francisco
Comments to the Author:
This is an impressive undertaking, and this study will likely make many important contributions to the understanding of child health in Italy. For the purposes of a Cohort Profile Paper, there are some aspects of the study design and primary objectives that remain somewhat unclear. I have made a few suggestions below by section of the paper.

Intro:

Lines 94-98 seem to drastically under-represent decades of work elucidating the social determinants of child health. Perhaps what is missing is an understanding of how these findings, often in other areas of the world, generalize to the Italian population?
Ok, we have rephrased the part about all the research already carried out so it does not seem under-represented, and have added general information about the data on Italy and where it stands compared to European data (first paragraph of introduction).

Line 111: It's unclear if the goal here is to use the 2020 review of birth cohorts only to justify this study. A pregnancy cohort would be a stronger design by including the pregnancy period, so for the purposes of justifying the novelty of this cohort in the broader context, recent pregnancy cohorts with similar plans to follow participants through birth and beyond should be included.
We are not sure what the reviewer is asking for here. We have cited all different types of cohorts, both pregnancy and birth cohorts, without trying to justify ours. Our view, in any case, is that pregnancy cohorts tend to follow the women more closely, and from the point of view of gynaecologists, while birth cohorts tend to focus more on the children.

Line 126-127: This statement requires references.
We have added a phrase after this statement, expanding the point ("Many of these inequalities are well-known and relate to regional differences…."), and have added a reference.

It would be helpful to have more detail on why some of the other cohort studies in Europe wouldn't be generalizable to the Italian context. Similarly, it seems that there are several ongoing pregnancy/birth cohort studies in Italy specifically. Is what sets this study apart it's focus on nurturing behavior? Do the other Italian studies rely on pediatricians for recruitment?
We had mentioned why NASCITA was different from other cohorts in Europe and in Italy in separate areas of the introduction: (In the third paragraph of the introduction we had stated that a 2020 review of European cohorts found that few, relatively recent, cohorts focused on family context (nurturing care) and its impact. In the sixth paragraph of the introduction we had also stated that "Like many other cohorts, it addresses multiple research questions.[29,30] The central role of the paediatrician in data collection on overall health and growth and in interaction with patients is a unique quality of NASCITA, however, and is possible in Italy because primary care for children is guaranteed by the family paediatrician, unlike in many other countries.")
Given the reviewer's comments, this may not have been clear enough, so we have added a phrase (seventh paragraph of introduction): "The synergy between the paediatrician's central role and this cohort study's focus on nurturing care make NASCITA different from most Italian and European cohort studies."

Study design

Since the pediatrician seems to be the source of the representativeness of this study, it would be helpful to include more details about the recruitment of pediatricians. How many were approached? How many refused? Was there any association between region/area/pediatrician characteristics and refusals? Was any information collected on those parents who did not consent to participate and were they different than those who did consent?
We do not know how many paediatricians refused participations because recruitment passed also through the paediatricians, especially the reference paediatricians in each area. We wanted to achieve a set, minimum number of paediatricians per cluster (as specified in the protocol, to which we now refer in the article for additional details) in order to have representativity at the national geographic level.
We have, in any case, added information on the recruitment of the paediatricians (second paragraph, Cohort description section), and have (as stated above) cited the protocol for additional details.

Data collection
It is stated that all data are collected by the pediatrician, but it's unclear on how the pediatrician gathered data on some of the primary exposures. Later in the paper, it becomes clear that nurturing care refers to positive health behaviors during pregnancy and early childhood. Defining this term more clearly at the outset would be helpful (it is easy to confuse nurturing with constructs like parenting). It is also unclear how other important contextual data are being collected (e.g., sociodemographics such as family income). Are the parents filling out any surveys? More detail needs to be provided here.
OK, we have added information on how the paediatrician gathered data, have moved the definition of nurturing care up in the text, and have explained that it is the paediatricians who collect data from the parents (paragraphs 4 and 5 of the Cohort description section).

Measurement
Were all measures for this study newly generated by study pediatricians? Were these assessments validated? More clarity on what specific was assessed in the study would help clarify this issue. In addition, it would be helpful if the authors could comment on why they did not rely on established measures for some of the constructs they are interested in.
For each outcome we have validated tools. In the text we only cited an example. In future studies focusing on certain outcomes we will specify the relevant tools.
OK, we have better specified (paragraph five of Cohort description section) that most data collected are already collected by paediatricians using routine data collection forms checklists and that additional data collected were integrated with validated tools for recording data of particular interest…"

Results

For tables 1 and 2, it would be clearer to put the column with the total first since the p-values are referring to comparisons between the North, Centre, and South.
OK, done.

After reviewing the results, the authors seem to be most interested in regional differences in early analyses but this focus isn't motivated in the introduction. Could the authors comment on why these differences are important? Is it related to resource allocation? Or potential regional differences related to health care, exposures, etc.? I am also concerned that some of these findings might be stigmatizing. There are numerous differences by region highlighted without adjustment for potential

differences in economic resources (which also seem to differ by region). It seems that many of these difference may be explained by differences in socioeconomic resources.
We understand that the findings might seem stigmatizing, however, we are reporting on a situation that has been well-known for years, and that is in need of effective interventions.
To make this more clear we have added statements concerning the regional differences that already exist in Italy and that need to be addressed. (Last phrase, paragraph 5 of Introduction: “Many of these inequalities are well-known and relate to regional differences, with southern Italy often resulting at the greatest disadvantage due to cultural and economic factors”).
We also added the phrase (Last phrase of Introduction): “In particular, geographic differences were analysed since they represent a variable known for influencing healthcare and behaviour at different levels.”
We do not have data on the families’ economic status because the ethics committee did not allow us to collect such data. We have, however, added a section concerning the possible influence of other factors to address the issue (last paragraph of Findings to date section “Geographical differences have been already documented by previous studies…..”).
There are many descriptive findings presented. Are any of these findings surprising or differ significantly from the other Italian cohorts mentioned in the intro?
We have added (first phrase of Strengths and limitations section) a phrase that states that the NASCITA study “confirms what is known about national level and regional level differences in health care practices and behaviours from other studies, such as smoking and alcohol data, but it is also the first study capable of assessing, at the national and regional levels, other variables such as pertussis vaccination in pregnancy”.

I would move table 5 earlier in the paper as it is related to the general characteristics of the cohort and how they relate to the nation. Related, despite the notion that using pediatricians to recruit participants would lead to a largely representative sample, there are quite notable differences in sociodemographic characteristics, with lower SES folks less likely to have been recruited into the study. In the context of this universal health care system, can the authors comment on why that might be? (Also see earlier comment regarding providing more information on characteristics of pediatricians that refused to participate).
We feel that Table 5 should follow the other tables because they describe the population and its characteristics in detail, while Table 5 is a general comparison of the data, but also includes some of the details and so would anticipate only part of the more detailed characteristics, possibly confusing the reader.
Concerning the sociodemographic characteristics, we have added a section at the end of the Findings to date section, just before the Strengths and limitations section, explaining that several factors can be associated with these differences.

Lines 348-350 seem to overstate the similarity between the cohort and national population.
We feel this is realistic. However, we have modified the phrase slightly, adding “generally” to the statement and specifying “with the exception of the lower prevalence of foreign-born mothers, as reported above”. It now reads: “Newborns participating in the NASCITA cohort, and their families, are generally representative of the Italian population in terms of geographical distribution [REF] and sociodemographic characteristics, with the exception of the lower prevalence of foreign-born mothers, as reported above.”.

We thank you for your comments.

Reviewer: 2

Mr. Brian Dodgeon, University College London Institute of Education
Comments to the Author:
Well done. I enjoyed reading this cohort profile. In general I felt the cohort study was well described, supplying enough detail for potential researchers to approach the study as a research resource. I therefore recommend acceptance subject to the minor revisions detailed below. The revisions mainly take the form of correcting typographical errors in the text or tables, or minor issues regarding English usage.

Line 128 I recommend changing the first line of the paragraph to 'Various cohort studies have been carried out in Italy,...' which is the more conventional way of describing resources such as this.
OK, done.

Line 137 Typo error: 'recruited them at in eight...' should be 'recruited them in eight...'
OK, done.

Line 229 I recommend changing the sentence to 'The majority of mothers and fathers were born in Italy...' Although the word 'majority' is singular, the mothers and fathers who are referred to are many in number, and so I'd say the common English usage is to use the plural form of the verb.
OK, done.

Line 233 (similar to above) I recommend 'but the majority were employed...'
OK, done.

Line 287 This line mis-quotes the figure in Table 3, which is '2474 (49.0%) were female'.
OK, done.

Line 302 This line mis-quotes the figure in Table 4, which is 'median of 24.9 days...'
OK, done.

Line 304 (similar to lines 229 & 233) I recommend plural usage: '...over one-third (35.9%) were receiving formula milk and two-thirds (64.1%) were receiving a mixed-feeding regimen.'
OK, done.

Line 307 In Table 4, the first line is wrongly labelled 'Mean age, days, SD', when it is clear from the figures that 25.7 is (correctly) the mean but 24.9 is the median, not the standard deviation.
OK, thank you. In any case, we have removed the line in the table, since the data is present in the text.

Line 307 Also in Table 4, the percentages for the three different categories under Breast feeding are misleadingly inconsistent. In the columns 'North', 'Centre' and 'South' the three respective percentages for 'Exclusive', 'Mixed' and 'Artificial' add up to 100%, whereas in the column 'Total n(%)' the figure of 63.6% represents the 'Exclusive' number as a percentage of all cohort members, but the two following figures represent the 'Mixed' and 'Artificial' numbers expressed as percentages of 'the remainder' (ie of all those not exclusively breastfed). So I recommend the % figures of 64.1 and 35.9 be replaced by 23.3 and 13.1 respectively.
OK, done.
Line 369 Recommend inserting the word 'at', to read '...in particular when looking at southern Italy.'
OK, done.

Thank you for all your comments.

*** ***

VERSION 2 – REVIEW

REVIEWER	LeWinn, Kaja University of California San Francisco, Psychiatry
REVIEW RETURNED	07-Oct-2022

GENERAL COMMENTS	I appreciate the thorough responses to reviewer comments, particularly the added context in the introduction, which is very helpful in motivating the importance of the study. I also appreciate the added clarity provided on recruitment and the protocol and additional context provided around regional differences. One note, the last line in the introduction which refers to the geographical differences implies that geography is the root cause of differences in behavior and healthcare. I might just change this sentence slightly to focus on associations rather than using the work "influence" which implies causality (e.g., as region of residence is known to be strongly associated with both healthcare quality and behavior in Italy, geographic differences in outcomes were analyzed).