

Transcript - WP3 – Introduction and Core Minimum Dataset – Elizabeth Draper

Slide 1 – Title slide

Hi, my name is Elizabeth Draper and I lead of the data mapping and harmonisation work package (WP3) for the RECAP preterm project. In this session I will introduce you to our team and provide you with an overview of the work plan for this work package as well as describing the development of a core minimum dataset for use in the development of future preterm birth cohorts.

Slide 2 – WP3 – Who are we?

So the data mapping and harmonisation team are based within the Department of Health Sciences at the University of Leicester in the UK. The team is led by 3 investigators: myself, Sam Johnson and Brad Manktelow who encompass expertise in perinatal & paediatric epidemiology, child development and medical statistics. You will meet the 3 research staff on the team, Deb, Helen and Charlotte who will present the main aspects of the data mapping and harmonisation of the preterm birth cohorts.

Slide 3 – Learning outcomes

By the end of this session I hope that you will have an understanding of:

- the approach taken to identify, collate and harmonise data for the RECAP preterm project;
- which birth cohorts are included in the RECAP preterm project in terms of their geographical location, the year the birth cohort was initiated AND the ages at which follow-up assessments and additional data collection was carried out; and finally
- the processes & principles used to develop a core minimum dataset for use in future preterm birth cohorts

Slide 4 - The Aim of WP3

The overarching aim of this work package can be broken down into four areas: to...

- facilitate pooling of individual patient data from very preterm birth cohort studies
- increase statistical power of preterm birth analyses
- enhance international comparison of preterm birth studies by providing a framework for standardising existing data
- inform on-going and future data collection for preterm birth studies

Slide 5 - RECAP Preterm work packages

This slide presents the eleven work packages that constitute the RECAP preterm project. These WPs can be grouped into 4 areas:

- Management and Governance
- Data Platform and Harmonisation
- Studies of very preterm outcomes in childhood and adulthood and
- Dissemination, sustainability and improving participation.

WP3 falls under the data platform and harmonisation aspects of the programme alongside WP4 – the data platform and WP5 – statistical methods for dealing with individual level patient data.

Slide 6 – WP3 workplan

Here the work plan of WP3 is represented as a circular pyramid showing the stepwise process by which the different aspects of the work build on each other.

These will be described by members of the team

Helen will describe the development of the metadata catalogue and inventory of the study documentation from the preterm birth cohorts involved – this forms foundations of WP3.

Deb will then describe the development of the schema used to organise the variables collected by the preterm birth cohorts and how these variables are then mapped to facilitate research projects carried out on the RECAP preterm data platform.

Charlotte will then describe how data harmonisation can then be carried out for the preterm birth cohorts

Within this presentation I will describe the work carried out to develop a prospective minimum core dataset for use in future preterm birth cohorts - represented here as the top of the pyramid.

Slide 7 – The Geographical location of the RECAP Preterm birth cohorts

The next three slides will describe the 23 preterm birth cohorts included within the RECAP Preterm project. This map shows the geographical location of these cohorts and indicates the number of cohorts within each country. Most of the cohorts are based in just one country for example the EXPRESS cohort in Sweden and the EPIPAGE cohorts in France whereas the EPICURE1 study was carried out in the UK and the Republic of Ireland and the EPICE study was carried out across 11 EU countries. The shading on the map shows the number of preterm birth cohort included in RECAP Preterm from the lightest blue indicating one cohort (Poland and RoI) to the darkest blue indicating 4 separate cohorts - in Finland.

Slide 8 – Individual RECAP preterm cohorts by year of birth and period of recruitment

The first of the preterm birth cohorts included in our study was the Finnish HeSVA cohort that was established in the late 1970's and early 1980's, with a further 5 studies being established in the mid 1980's. The remaining preterm birth cohorts have been established since the mid 1990's with some countries developing repeat cohorts to assess the impact of changing policies around viability of extremely preterm births and improvements in perinatal and neonatal care on longer term survival eg. the UK Epicure studies and the French EPIPAGE studies. As shown here recruitment periods also varied between studies.

Slide 9 – RECAP preterm birth cohorts – age at data collection/assessments

The number of follow-up assessments varies between the cohorts (from the establishment of the cohort with one follow-up assessment up to 8 assessments at different ages) and this graph shows the age at each point of data collection and assessment for each of the 23 preterm birth cohorts. All cohorts followed up their children until at least 4 years of age. However many of the more historical cohorts have carried out repeated follow-up assessments and data collections into teenage years and adulthood –with the POPS study from the Netherlands assessing their cohort at the age of 28 years. In RECAP preterm the early studies that have followed up their cohorts to at least 10 years of age and potentially into adulthood are the focus of Work Package 9 and the later cohorts focussed on birth and early childhood are the focus of Work Package 7.

Slide 10 – Scale of the RECAP Preterm task

Given the large number of follow-up assessments carried out by some cohorts the scale of the task to collect together all study documentation was vast. This is summarised here on this slide. So in total WP3 have collected 118 data dictionaries covering around 74 data collection points across the 23 cohorts

Perinatal – 32 data dictionaries – some cohorts have more than one data dictionary to cover the different aspects of data collection in the perinatal and neonatal periods.

In addition there are 47 data dictionaries that cover follow-up to 10 years of age and a further 39 data dictionaries covering follow-up from 10 years of age into adulthood.

In total we have dealt with more than 23,000 Variables : with around 4,500 relating to the perinatal/neonatal period and approximately 18,500 relating to short and longer term follow up.

Full details of how WP3 have dealt with all this metadata will be presented in the remaining WP3 presentations.

Slide 11 – WP3 work plan

The rest of my talk will now focus on the top circle of the pyramid – the development of a minimum core data set for future preterm birth cohort studies.

Slide 12 – Development of core dataset for ongoing and future studies

We started this process by carrying out systematic review of core outcome datasets for preterm births and their long term follow-up using a number of search engines including Medline and PubMed and supplemented our findings with variables from the data dictionaries collected for the RECAP preterm cohorts from which we identified common variables. This was supplemented with variables identified for use in the WP7, 8 and 9 studies, discussions with the RECAP Steering Committee, data from the DELPHI panel developed for WP7 which identified priorities for health professionals, families of VPT children and the graduates of VPT birth as well as the Recommendation of the Adults Born Preterm International Collaboration (APIC). From these sources we developed an initial data set that we used for our DELPHI process.

The **Delphi method** is a **process** used to arrive at a group opinion or decision by surveying a panel of experts. Experts respond to several rounds of questionnaires, and the responses are aggregated and shared with the group after each round. I am not going to describe it in detail here as this will be covered in the presentations for WP7.

Slide 13 – Core dataset – identifying variables

So we established a number of principles for deciding which variables should be included in the minimum core dataset – the key issue being to ensure that this is a truly minimum core dataset – one that includes those variables that ALL preterm birth cohorts should collect as a minimum.

Rationalisation of the variables was conceptual rather than practical and was based on 3 elements:

Firstly the topic – ensuring the focus was on preterm births rather than the more generic variables for any core birth dataset. Similarly we did not want to focus on specific health conditions that are not necessarily related to preterm birth.

Our second consideration was how frequently variables were identified in the relevant literature and within the RECAP project data dictionaries ... and

Finally we wanted to ensure that were as up to date as possible so reviewed variables relating to recently developed treatments of importance to preterm births.

Slide 14 – Core dataset – schema modules covered

We used the schema modules (that will be described later by Deb) to identify and organise the variables that were sent out for the DELPHI process. To prevent overload of DELPHI participants and to allow for a more focussed selection of potential DEPHI participants we carried out 2 separate processes – one for those modules focussed on the perinatal period (Antenatal & Birth, Neonatal care and Mortality & End of Life) and then a second DELPHI process for the collection of data from the follow-up modules: neurodevelopment, lifestyle, physical & mental health etc.

Slide 15 – Core dataset – identifying variables

So far we have completed the first DEPHI – rounds 1-3: antenatal & birth, mortality and neonatal care modules and this Venn Diagram shows the overlap between the different sources we identified for the perinatal variables indicating the total number of variables from each source. So for example, looking at the purple circle collating information from preterm specific paper recommendations there were 94 variables in total (23+29+39+3) with overlaps between the different sources: 39 variables were identified from all 3 major sources – preterm specific, WP7,8,9 core variables, and non specific preterm core outcome sets and recommendations.

Slide 16 – Core dataset – identifying variables cont...

This was a sizeable task – but we followed a clear process in order to decide which variables were included within the first round of the DELPHI.

Firstly we reviewed the WP7/8/9 core variables lists and included all variables that were identified in more than one WP list. Then two recent papers recommending core dataset and linked to the RECAP preterm project were reviewed (Kajantie et al (2020) and Doyle et al (2014)) and all variables recommended in these included. If variables were included in more than one paper identifying preterm birth specific core outcome sets or recommendations – then they were included in the DELPHI. Finally we reviewed the RECAP cohort variables at domain level, splitting RECAP cohorts into those pre-2000 and post-2000 for rationalising on timescale. If a variable was included in >60% of RECAP cohorts post 2000, variables were added to cover this domain if not already included.

Slide 17 – Core dataset – DELPHI process

So to date we have carried out the first three round DELPHI process as shown here. 63 variables were included in this process from across the 3 modules (antenatal & birth, neonatal care and mortality). In round 1 we had responses from 69 participants who suggested a further 52 variables across the 3 modules. In round 2 48 participants voted on the proposed & revised variables and in the final round 43 participants gave a final rating on the proposed variables. This work is now being written up for publication and the variables fitting the criteria for consensus will form the perinatal variables in our suggested RECAP preterm Core Minimum Dataset. We are happy to circulate this paper when published to any of you who are interested. Contact details are in my final slide.

Slide 18 – References

These references provide further information about core outcome datasets and their development

Slide 19– Further reading / web site links

For further reading about initiatives relating to the development of core dataset I have listed a few papers and links to website on this slide.

Slide 20 – Thank you

Thank you – if you want any further information please feel free to contact myself of the WP3 team using the emails on the slide.