



BMJ Open Identification of child mental health problems by combining electronic health record information from different primary healthcare professionals: a population-based cohort study

Nynke R Koning ¹, Frederike L Büchner ¹, Nathalie A Leeuwenburgh,² Irma JM Paijmans,² DJ Annemarie van Dijk-van Dijk,² Robert RJM Vermeiren,³ Mattijs E Numans,¹ Mathilde Crone¹

To cite: Koning NR, Büchner FL, Leeuwenburgh NA, *et al.* Identification of child mental health problems by combining electronic health record information from different primary healthcare professionals: a population-based cohort study. *BMJ Open* 2022;**12**:e049151. doi:10.1136/bmjopen-2021-049151

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-049151>).

Received 16 January 2021
Accepted 16 December 2021



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

²GGD Hollands Midden, Leiden, The Netherlands

³Curium-LUMC Academisch centrum voor Kinder- en Jeugdpsychiatrie, Leiden, The Netherlands

Correspondence to

Dr Nynke R Koning; n.r.koning@lumc.nl

ABSTRACT

Objectives To investigate the potential value of combining information from electronic health records from Dutch general practitioners (GPs) and preventive youth healthcare professionals (PYHPs) in predicting child mental health problems (MHPs).

Design Population-based retrospective cohort study.

Setting General practice, children who were registered with 76 general practice centres from the Leiden University Medical Centre (LUMC) primary care academic network Extramural LUMC Academic Network in the Leiden area, the Netherlands. For the included children we obtained data regarding a child's healthy development from preventive youth healthcare.

Participants 48 256 children aged 0–19 years old who were registered with participating GPs between 2007 and 2017 and who also had data available from PYHPs from the period 2010–2015. Children with MHPs before 2007 were excluded (n=3415).

Primary outcome First MHPs based on GP data.

Results In 51% of the children who had MHPs according to GPs, PYHPs also had concerns for MHPs. In 31% of the children who had no MHPs according to GPs, PYHPs had recorded concerns for MHPs. Combining their information did not result in better performing prediction models than the models based on GP data alone (c-statistics ranging from 0.62 to 0.64). Important determinants of identification of MHPs by PYHPs 1 year later were concerns from PHYPs about MHPs, borderline or increased problem scores on mental health screening tools, life events, family history of MHPs and an extra visit to preventive youth healthcare.

Conclusions Although the use of combined information from PYHPs and GPs did not improve prediction of MHPs compared with the use of GP data alone, this study showed the feasibility of analysing a combined dataset from different healthcare providers what has the potential to inform future studies aimed at improving child MHP identification.

INTRODUCTION

Worldwide, on average, one in five adults experienced a mental health problem

Strengths and limitations of this study

- Large, population and real life-based cohort study with children aged 0–19 years old.
- Linked information from electronic health records from different providers, namely Dutch general practice and preventive youth healthcare.
- Many determinants of child mental health problems described in the literature could be operationalised from the routine healthcare data sources, in particular from preventive youth healthcare data.
- Real-life data from daily practice causing substantial missing data for several determinants.
- No free-text data available from the electronic health records.

(MHP) within the previous 12 months.¹ With the majority of MHPs originating in childhood and adolescence, early identification of child MHPs is important to be able to provide adequate treatment strategies and enable prevention of adverse outcomes later in life.^{2–5} Over the past years, the use of information from electronic health records (EHRs) for research, proactive care interventions and healthcare innovations has become increasingly popular. These often very large datasets contain an abundance of detailed information on individual members of diverse patient populations and provide opportunities for all kinds of research, including the development of prediction models that can be used in daily clinical care to identify high-risk individuals and subpopulations.^{6,7} When integrated in daily routine, such prediction models might also be able to support professionals in the timely recognition of child MHPs in an efficient manner.

Recently, several studies have investigated possibilities to predict child MHPs with data from EHRs extracted from British and Dutch general practices, respectively, in order to improve MHP recognition.^{8,9} In the latter study, information regarding known social determinants for child MHPs (eg, regarding the child's family and environment) was not available since it is not a standard part of extractable data from primary care EHRs.^{9,10} Combining information gathered by different healthcare professionals might lead to more complete information. In this way, MHP recognition might be facilitated by including social determinants, that in previous research have appeared to be important predictors for child MHPs.¹¹ So, the EHRs extracted from Dutch preventive youth healthcare (PYH) might be a useful additional source of information regarding social determinants for child MHPs in general, as well as for prediction purposes. PYH paediatricians and nurses (preventive youth healthcare professionals, PYHPs) are, together with general practitioners (GPs), the key professionals in (preventive) primary healthcare for Dutch children.¹² In general all children aged 0–19 years visit PYHPs during standardised, regularly scheduled, free of charge check-ups that are aimed to monitor all aspects of a child's healthy development, including social determinants.¹³ During a visit, a child's history is taken and physical examinations are performed. In addition, validated MHP screening instruments such as the Strengths and Difficulties Questionnaire (SDQ) and short indicative questionnaire for psychosocial problems among adolescents (KIVPA) are filled in during the screening visits, and nurses or doctors working in PYH will report concerns for MHPs.¹³ PYHPs can refer children to general practice, secondary care or other paramedical professionals in case of problems.

GPs on the other hand, are the gatekeepers of the Dutch healthcare system. They provide primary healthcare to children and related family members that are enlisted with their practice centres, free of charge, usually reacting on what patients present and care is usually related to acute and chronic diseases. In case of more severe problems, children will be referred to secondary (mental) healthcare. With their own specific knowledge and tasks within the Dutch healthcare system, PYHPs and GPs each register different information on children and their families.¹²

The aim of this retrospective population-based cohort study was to investigate the potential value of combining and analysing the information from EHRs from both general practice and PYH to improve the identification of children with MHPs in daily practice.

METHODS

Study design, setting and population

To improve the identification of children with MHPs we aimed to predict first recorded MHPs based on general practice data. We, therefore, created a population-based cohort including children aged 0–19 years who were

enlisted with 76 practice centres (107 GPs) that were affiliated with the Extramural LUMC Academic Network of the Leiden University Medical Centre (LUMC), the Netherlands. The participating practices were located in Leiden and surroundings. All patients aged 0–19 years on 31 December 2016 and registered with participating GP centres between 1 January 2007 and 1 January 2017 for at least 1 year were part of our cohort.

The GP data consisted of demographics, consultation dates, symptoms and diagnoses coded according to the WHO/World Organization of Family Doctors (WONCA) accredited International Classification of Primary Care (ICPC), prescribed medication coded according to the Anatomical Therapeutic Chemical (ATC) classification, laboratory test results, as well as descriptive or coded information from referrals and correspondence with other healthcare professionals.^{9,14,15} Patients were excluded in case a MHP was recorded before 1 January 2007 (n=3415).

We used data from two different sources, namely EHR data extracted from GP centres and from PYH centres. The nature and quality of the data are previously described in more detail by Koning and colleagues.^{9,16} For the included children we obtained anonymously extracted data from PYH centres that were part of the Regional Public Health Service Hollands Midden. All PYH electronic healthcare data from the period 2010–2015 and all available summary data from a prior electronic registration system from 2005 to 2010 for children born between 1994 and 2012 were available. The coded data were linked with the assistance of a trusted third party (TTP).¹⁷ The TTP deidentified the general practice EHR data of every child. In order to link the patient data from general practice with the data from PYH, the TTP provided both the Dutch citizen service number and the pseudo patient number from the children to the PYH organisation. The PYH organisation extracted all available data for these children and also deidentified their routine healthcare data with the same pseudo patient numbers. In this way, we received anonymous patient data from PYH and general practice for approximately 50 000 children, which we could combine on the individual patient level with the pseudo patient number.

Outcome

Our outcome was a first recorded child MHP based on GP data, and was defined when at least one of the following was present: a recorded MHP, a referral to child mental healthcare and/or a mental health medication prescription between 1 January 2007 and 1 January 2017 (online supplemental table 1). We defined a recorded MHP when ICPC codes from the P (psychological) chapter or ICPC code T06 ('anorexia nervosa/bulimia') were present, including both mental health symptoms as well as hypothesised and confirmed disorders. Related mental health medication prescriptions were defined as prescriptions coded

with ATC codes N05A, N05B, N05C, N06A, N06BA02, N06BA04, N06BA09, N07BA or N07BB which includes all relevant psychiatric medication (such as antidepressants and medication for attention deficit hyperactivity disorder). Referrals to child mental healthcare were defined as referrals to a psychologist, psychiatry or psychotherapy.⁹

Determinants

Potential determinants were related to the child (eg, gender, developmental characteristics, somatic complaints and comorbidities), the family (eg, parental education, parental divorce and MHPs occurring with other family members), healthcare in general (eg, number of visits or prescriptions) and determinants related to possible MHPs registered in PYH (online supplemental tables 2 and 3). PYHPs register results of validated MHP screening tools such as the SDQ, but they can also record their concerns for MHPs. Concerns for MHPs were defined when either abnormal psychosocial functioning in the child was reported (eg, problems in making contact with others or hyperactive behaviour) during the check-up, and/or when the child received extra healthcare regarding mental health (within PYH or within curative care).

Determinants were selected a priori based on literature regarding determinants for MHPs and an expert panel.^{10–18} Regarding the GP data, every first occurrence of a determinant was taken into account as GPs see patients on an irregular, patient-determined basis. We assumed that in case a determinant was not registered, it was not present as GP data are a problem-based registry.¹⁹

PYHPs see children regularly during standard visits in which specific items are checked and recorded. During the first 4 years of life, around 15 PYH visits are scheduled. Subsequently, in both primary (children age 4–11 years) and secondary school years, (children age 12–19 years), children are generally seen twice.¹³ Regarding PYH determinants that should have been registered by PYHPs, we assumed that in case of missingness the determinants were normal.¹⁹ Since some determinants can change over time, we included either the first (eg, for bullying or school problems) or last (eg, for overweight) registered value at the moment of prediction. For the other determinants, we included the first known registered value (online supplemental table 3). Due to sparseness of the data, we clustered closely related determinants.¹⁶

As determinants for MHPs may vary across childhood and adolescence, we investigated models for the age groups primary school-aged children (age 4–11 years) and secondary school-aged children (age 12–19 years) separately. The same set of determinants was examined in the different age groups, however we required the prevalence of a determinant to be >1% per age group with regard to the clinical usefulness of the determinant.

Statistical analyses

Descriptive statistics, including percentages of missing data for determinants registered by PYHPs, were carried out with SPSS (V.25). We also looked at the overlap between concerns for MHPs in PYH and MHPs based on GP data. To obtain the 1-year risk of a first recorded MHP, we developed a multilevel logistic regression model per age group. First, the data were split according to the children's age; age 4 years, age 5 years and so on. For every age, (time point 0 (T0)) the status of all determinants was updated and the outcome was assessed 1 year later (time point 1 (T1)). We obtained a prediction model per age group by combining the data from the different ages and fitting a logistic regression model including a cluster effect on the patient level with R (V.3.5.3). This to adjust for using different age years of one patient.²⁰

We investigated several models for MHPs in three steps for both age groups: (1) determinants based on GP data, (2) determinants based on GP data and PYH results of validated MHP screening tools and PYHPs' concerns for MHPs, (3) determinants based on all available GP and PYH data (online supplemental table 3). In daily clinical practice, one can imagine that it would be more efficient to share only the minimum amount of relevant information between GP and PYH. By step-wise adding information from PYH in step 2 and 3, we would be able to see the additional value of combining only a small selection of PYH data in relation to all available PYH data. In step 2, we only added data from PYH that represented their findings regarding the mental health development of the child. In step 3, all other data from the PYH regarding the child and parents were added. The PYH determinants in step 2 (a PYH concern for MHPs and/or deviant results for MHP screening tools), were chosen because we hypothesised they would be important determinants for child MHPs based on GP data. As not for all children data from both GP and PYH were available, we also explored models with complete cases, that is, with the patients that had no missing data for the PYH determinants. For all models with outcome first MHP, we excluded children with recorded MHP before T0.

The ability of the model to distinguish between children who are recognised with a first MHP and those who are not (discrimination), was assessed using the c-statistic.²¹ The in-sample calibration of the model was assessed by the calibration plot of actual probabilities vs predicted probabilities. The models were internally validated using bootstrap resampling (500 bootstrap samples) and estimating shrinkage factors. Brier scores were calculated to assess the average prediction error.²²

Patients and public involvement

Due to the nature of the data, patients and public were not directly involved in this study.

RESULTS

Our cohort of GP data consisted of 70 000 children. From 48 256 children (68.9% of those included), data registered in PYH could be individually linked by our TTP. The median follow-up time of children in the GP data was 6.4 years, in the PYH data 3.6 years. Of the children aged 4–11 years, 48.8% was male and 3.0% had an increased SDQ score (table 1). Of the children aged 12–19 years, 46.7% was male and 3.9% had an increased SDQ score. Over half of the determinants supposed to be registered in PYH had more than 50% missing data (table 1, online supplemental table 4).

Prevalence of (concerns for) MHPs

We were able to include data registered at PYH centres for the period 2005–2015, while GP data were available for the period 2007–2016. In the period 2007–2015, 15 823 of 48 256 (32.7%) included children had a first recorded MHP based on GP data and 18 092 of 48 256 (37.5%) children had first recorded concerns for MHP based on PYH data (table 2). In the 15 823 children with MHP according to GPs, 8079 (51%) children had concerns for MHPs according to PYHPs. In 10 013 of 32 433 (30.9%) children in whom the GP did not have a recorded MHP in that period, PYH had registered concerns for MHPs in the same period. In 7744 of 30 164 (25.7%) children in whom PYH did not have concerns for MHPs, GPs had recorded MHPs.

Prediction of a first MHP

Determinants of a first recorded MHP 1 year later based on GP data in the school-aged children were similar in all models (table 3). In the GP data, determinants of a first recorded MHP were somatic complaints, life events and the healthcare use related variables more than two GP visits, one or more medication prescript, one or more laboratory test and one or more referral to or contact with other healthcare professional all measured in the previous year. Low socioeconomic status in children aged 12–19 years, developmental problems and a recorded chronic disease in children aged 4–11 years, were only related to MHP in step 1 when not including data registered in PYH. Male gender was related to an increased likelihood of a recorded MHP compared with female gender in children aged 4–11 years and a lower likelihood in children aged 12–19 years in all models.

PYH determinants (steps 2 and 3) that were associated with an increased risk of first recorded MHPs based on GP data 1 year later in both age groups were concerns for MHP according to PYHPs, elevated problem scores on MHP screening instrument SDQ, extra healthcare visit in PYH, life events and family history of MHP. Protective PYH determinants registered in PYH in both age groups were non-Western ethnicity of one or both parents, child low secondary school level and high technology use (eg, on average over 2 hours of daily screen use, online supplemental table 2). Incontinence, sleeping problems and school problems were positively associated with a

first recorded MHP 1 year later only in age group 4–11 years, while prenatal risk factors such as substance abuse by mother during pregnancy and young parenthood, were associated with a decreased risk of MHPs in this age group.

In age group 12–19 years, an increased problem score on MHP screening instrument KIVPA was positively associated with a first recorded MHP. In this age group, a relatively higher or lower secondary school level of the child and being under treatment outside PYH were negatively associated with a first recorded MHP. All other determinants were not found to be associated with a first recorded MHP.

The prediction of a first MHP 1 year later based on combined data from GPs and (partly) PYH did not result in better performing prediction models than the models based on GP data only, c-statistics ranged between 0.62 and 0.64. Internal validation was good (online supplemental figures 1 and 2).

Complete-case analysis

Complete-case analyses, analyses with patients who had no missing data regarding determinants registered in PYH, were only possible for the models investigating SDQ, KIVPA and concerns for MHPs registered during PYH visits (step 2). These models did not perform better than models with all available data. We could not carry out complete case analyses for the models incorporating all determinants registered at PYH centres (step 3), since too many patients had missing data.

DISCUSSION

Main findings

This population-based cohort study investigated the added value of combining data registered at PYH centres with general practice data for the identification of child MHPs as recorded by GPs. Combining information from PYH and GP centres to predict MHPs based on GP data, did not result in better performing prediction models than the models based on analysis of GP data alone. Determinants derived from PYH registries for the prediction of a first MHP 1 year later were contextual determinants, concerns for MHPs as registered by PYHPs and elevated scores on MHP screening tools. Furthermore, our study showed that in 51% of the children who had a recorded MHP according to GPs between 2007 and 2015, concerns for MHP were also registered by PYHPs in the same period. In nearly a third of the children who had no MHPs according to GPs, PYHPs had recorded concerns for MHPs in the same period. In 25% of the children in whom PYHPs did not register concerns for MHPs, GPs had recorded MHPs in the same period.

Strengths and limitations

We used a large population-based sample of 70 000 children. For the vast majority of these children, we were able to link data registered in PYH centres at an

Table 1 Baseline characteristics

GP characteristics	Children age 1–3 years N=12,196		Children age 4–11 years N=32 081		Children age 12–19 years N=18 829	
	% (n)	Missings % (n)	% (n)	Missings % (n)	% (n)	Missings % (n)
Male gender	51.0 (6221)	NA	48.8 (15656)	NA	46.7 (8788)	NA
Low Socioeconomic status	4.2 (510)	NA	3.4 (1079)	NA	3.4 (644)	NA
Congenital anomaly	11.0 (1339)	NA	12.2 (3908)	NA	15.2 (2854)	NA
Neoplasms	2.4 (297)	NA	5.6 (1797)	NA	7.4 (1397)	NA
Chronic disease*	48.8 (5950)	NA	46.5 (14912)	NA	41.7 (7845)	NA
Somatic complaints†	25.2 (3078)	NA	39.9 (12793)	NA	53.0 (9986)	NA
Life event	0.4 (51)	NA	1.0 (313)	NA	2.0 (373)	NA
>2 Visits	91.3 (11133)	NA	87.5 (28065)	NA	87.6 (16489)	NA
≥1 Medication prescript	78.5 (9569)	NA	71.2 (22848)	NA	71.9 (13539)	NA
≥1 Laboratory test	14.6 (1782)	NA	24.1 (7742)	NA	36.4 (6848)	NA
≥1 Referral/ correspondence other healthcare prof.	70.1 (8544)	NA	68.0 (21817)	NA	69.4 (13062)	NA
PYH characteristics						
CMHP	4.9 (592)	39.5 (4812)	22.4 (7197)	11.6 (3707)	42.9 (8078)	9.7 (1820)
Ethnicity	8.5 (1036)	54.2 (6613)	7.8 (2498)	60.2 (19303)	3.1 (588)	86.5 (16288)
Developmental problems	4.2 (516)	41.4 (5055)	4.5 (1439)	21.9 (7018)	1.6 (292)	22.3 (4191)
Incontinence	NA	NA	3.3 (1071)	23.3 (7490)	3.2 (600)	16.4 (3097)
School problem	0.3 (31)	98.2 (911982)	3.0 (966)	61.4 (19692)	8.5 (1599)	27.5 (5182)
Secondary school level low	NA	NA	3.9 (1241)	96.1 (30833)	16.6 (3129)	43.5 (8192)
Secondary school level high	NA	NA	0.0 (0)	73.9 (23718)	9.4 (1763)	43.5 (8192)
Secondary school level other	NA	NA	0.0 (6)	73.9 (23718)	0.1 (27)	43.5 (8192)
High technology use	NA	NA	4.8 (1547)	88.8 (28493)	15.4 (2892)	69.7 (13132)
SDQ border area	0.5 (59)	96.8 (11801)	5.7 (1821)	32.3 (10361)	5.6 (1049)	35.2 (6626)
SDQ increased	0.4 (43)	96.8 (11801)	3.0 (966)	32.3 (10361)	3.9 (736)	35.2 (6626)
KIVPA	NA	NA	0.0 (1)	100.0 (32079)	7.6 (1424)	46.1 (8680)
Under treatment	NA	NA	5.1 (1622)	94.9 (30459)	10.9 (2048)	89.1 (16781)
Extra PYH healthcare visit	32.6 (3977)	NA	40.3 (12920)	NA	23.3 (4379)	NA
Life events	3.6 (441)	74.2 (9049)	12.8 (4116)	39.4 (12626)	17.3 (3262)	22.7 (4277)
Family history of MHP	3.8 (463)	59.0 (7200)	4.6 (1467)	31.2 (10006)	2.1 (394)	24.4 (4590)
Prenatal risk factors	31.1 (3794)	44.9 (5478)	13.7 (4383)	47.2 (15153)	2.7 (501)	77.2 (14527)

Determinants from general practice that are displayed in online supplemental table 4 because of no significant results in the analyses or a prevalence of <1%: perinatal morbidity, disabilities, difficult temperament, somatic complaints+chronic disease.

Determinants from PYH that are displayed in online supplemental table 4 because of no significant results in the analyses or a prevalence <1%: premature, neonatal problems, non-spontaneous birth, sleeping problems, eating problems, overweight, underweight, bullying/being bullied, low self-confidence/resilience, self-harm, member hobby/music club, insufficient physical exercise, substance use, energy drink consumption, poorly experience health, total referral in past year, chronic illness parent, risk factor parents, non-traditional family composition, negative balance, little confidence in parenting skills, environmental stressors.

*Chronic disease when present one or more of the following: asthma, eczema, psoriasis, inflammatory bowel disease, epilepsy, diabetes mellitus, cystic fibrosis, rheumatoid arthritis.

†Somatic complaint when present one or more of the following: tension headache, migraine, abdominal pain, constipation, tiredness, musculoskeletal symptoms, dizziness, nausea, hyperventilation syndrome, palpitations, fainting.

CMHP, concern for mental health problem according to preventive youth healthcare; GP, general practice; IBS, irritable bowel syndrome; KIVPA, short indicative questionnaire for psychosocial problems among adolescents; MHP, mental health problem; NA, not applicable, for example, when determinant is not applicable for a specific age (eg, member of hobby/music club in children age 1–3 yearS), or in case no missing (eg, for extra PYH healthcare visit yes/no); PYH, preventive youth healthcare; SDQ, Strengths and Difficulties Questionnaire.

individual level. To our knowledge, this is the first study that investigated the combination of routine healthcare data from different healthcare providers as the basis

for identification of child MHPs in primary care in the Netherlands. As GPs and PYHPs have different positions and complementing roles,¹³ linkage of EHR data from

Table 2 Overlap in MHP according to GPs and preventive youth healthcare professionals' concerns for between 2007 and 2015

		Preventive youth healthcare concerns for MHPs 2007–2015		
		Yes (%)	No (%)	Total (%)
MHP 2007–2015	Yes	8079 (51.0)	7744 (48.9)	15 823 (100)
	No	10 013 (30.9)	22 420 (69.1)	32 433 (100)
	Total	18 092 (37.5)	30 164 (62.5)	48 256 (100)

GPs, general practitioners; MHP, mental health problem.

both sources provides a potentially valuable source of information regarding child development and health. Worldwide, linking EHR data from different sources is becoming increasingly popular. The Scandinavian countries are well equipped for linking information from different, mostly medical, sources.^{23 24} Experiences with linking data from different sources, for example, from the UK, are being published,^{25 26} although results of studies using linked data from different domains are still sparse.

We aimed to incorporate all available information regarding known risk factors for child MHPs and explored whether information exchange would result in better prediction models based on routine healthcare data that could be used in daily practice to improve MHP identification in an efficient manner. By using this population-based cohort, our study gives a more comparable reflection of the whole population than studies that actively recruit patients, studies in which it is known that minority populations (either ethnic or socioeconomically defined) are represented less.²⁷

Although data registered in PYH was available in nearly 70% of the children we originally included, over half of the determinants presumed to be registered in PYH showed more than 50% missing data and the prevalence of determinants like family history of MHPs was lower than expected from literature.²⁸ The electronic system used by PYHPs to record findings from clinical care is technically built in such a manner that important information from previous consultations remains present in the system, for example, birth weight and prematurity. However, in the extracted data for this research, this was not always the case, resulting in substantial missing of data from potential determinants.¹⁶ By design, we could not actively ask patients about specific risk factors. As missingness was likely to be missing not at random, we chose to not use multiple imputation techniques.²⁹ We did however, use normal values in case there should be a values present, as abnormal values will be noted by PYHPs. In addition, our aim was to explore which specific information from PYH could be useful to exchange with GPs to enhance MHP identification. Imputing data missing from our extracted data, eventually potentially used to share with GPs for clinical practice purposes, seemed not justifiable too.

Comparison with previous research

Combining information from PYH and GP centres did not show an added value in our study. However, as expected from literature, MHP screening tools and concerns for MHPs appeared to be determinants usable to recognise MHPs.¹⁰ The quality of our data and not having information available regarding important determinants of MHPs, such as academic achievement might be reasons for the lack of added value we found from adding PYH data in our analyses. In a prospective cohort study in which parents of Dutch children aged 11 years filled out questionnaires, a prediction model for adolescent MHP was derived which showed good discriminatory power (c-statistic 0.75).¹¹ Apart from similar predictors gender, family history of psychopathology, and life events such as parental divorce and moving house, this study also found math achievement at school and maternal educational level to be predictors.¹¹ These predictors were not, or not well reported in our study.

Contrary to previous research, known risk factors for child MHPs high technology use, prenatal risk factors such as substance abuse during pregnancy and non-Western ethnicity were found to be protective for MHPs in this study. One important reason is probably the high amount of missing values for these variables (60%–88%). We conducted additional analyses to see whether our findings could be explained by effects of other factors. It seems that children with a higher technology use, prenatal risk factors and non-Western ethnicity were more likely to have a higher socioeconomic status, which possibly explains the protective effect of these determinants. For example, our study cohort included the beginning period of technology use, a time during which technology use was more often seen in higher educated people, who in general have a lower risk of developing MHPs.

In addition, we have chosen to predict a first recorded MHP 1 year later. This time interval of a 1-year prediction might be of influence. The question might be for which time interval it is possible to predict MHP with sufficient accuracy based on routine healthcare data. A recent case–control study with British GP data predicting a first episode of depression in adolescents showed a better performance (c-statistics approximately 0.71).⁸ However, this study had a cross-sectional design and also included symptoms of depression such as low mood and

Table 3 Results of adjusted logistic regression analysis for the 1-year risk of MHPs

GP covariates	Step 1*				Step 2				Step 3									
	Age 4–11 years 171 577 person years, no of events 8204		Age 12–19 years 98 754 person years, no of events 5947		Age 4–11 years 105 371 person years, no of events 6236		Age 12–19 years 75 106 person years, no of events 3664		Age 4–11 years 105 371 person years, no of events 6236		Age 12–19 years 75 106 person years, no of events 3664							
	Coef	OR	95% CI	Coef	OR	95% CI	Coef	OR	95% CI	Coef	OR	95% CI						
Intercept	3.60			-3.51			-3.62			-3.61			-3.78					
Male gender	0.50	1.65	1.61 to 1.70	-0.19	0.82	0.75 to 0.89	0.47	1.59	1.51 to 1.68	-0.21	0.79	0.74 to 0.85	0.46	1.56	1.48 to 1.64	-0.20	0.77	0.72 to 0.83
Low SES	0.21	1.23	0.23 to 2.23	0.19	1.20	1.06 to 1.34	0.11	1.10	0.95 to 1.27	0.11	1.10	0.91 to 1.32	0.12	1.11	0.95 to 1.28	0.11	1.07	0.89 to 1.28
Congenital anomaly	0.15	1.15	1.09 to 1.21	0.11	1.11	1.03 to 1.18	0.12	1.12	1.03 to 1.20	0.12	1.11	1.02 to 1.21	0.12	1.10	1.02 to 1.19	0.11	1.07	0.99 to 1.17
Developmental problem	0.20	1.21	1.15 to 1.29	0.02	1.01	0.83 to 1.19	0.12	1.12	1.02 to 1.22	0.09	1.08	0.90 to 1.30	0.11	1.09	1.00 to 1.20	0.12	1.08	0.89 to 1.30
Life events	NA	NA	NA	0.58	1.79	1.58 to 1.99	0.49	1.62	1.28 to 2.05	0.48	1.60	1.28 to 2.01	0.44	1.52	1.20 to 1.93	0.41	1.46	1.16 to 1.83
Chronic disease*	0.10	1.10	1.05 to 1.15	0.03	1.02	0.94 to 1.10	0.06	1.05	0.98 to 1.12	-0.01	0.98	0.88 to 1.09	0.07	1.05	0.98 to 1.13	0.02	0.97	0.87 to 1.08
Neoplasms	0.06	1.06	0.97 to 1.15	-0.02	0.97	0.84 to 1.10	0.06	1.05	0.94 to 1.16	0.01	1.00	0.86 to 1.15	0.06	1.04	0.93 to 1.31	0.02	0.98	0.85 to 1.13
Somatic complaints**	0.19	1.20	1.16 to 1.24	0.20	1.21	1.17 to 1.15	0.20	1.21	1.12 to 1.30	0.27	1.29	1.18 to 1.41	0.22	1.21	1.12 to 1.31	0.27	1.27	1.16 to 1.39
>2 Visits	0.23	1.26	1.20 to 1.31	0.21	1.22	1.14 to 1.30	0.20	1.21	1.12 to 1.30	0.15	1.14	1.04 to 1.26	0.19	1.19	1.10 to 1.28	0.14	1.10	1.00 to 1.22
≥1 Medication prescript	0.10	1.10	1.05 to 1.15	0.27	1.30	1.23 to 1.37	0.11	1.10	1.04 to 1.17	0.24	1.26	1.16 to 1.36	0.10	1.08	1.02 to 1.15	0.23	1.22	1.13 to 1.32
≥1 Laboratory test	0.09	1.09	1.02 to 1.16	0.17	1.17	1.09 to 1.25	0.10	1.09	1.01 to 1.19	0.17	1.16	1.07 to 1.27	0.10	1.08	1.00 to 1.17	0.16	1.13	1.03 to 1.23
≥1 Referral/contact other healthcare prof.	0.29	1.33	1.28 to 1.38	0.26	1.28	1.22 to 1.35	0.26	1.28	1.21 to 1.36	0.28	1.31	1.21 to 1.41	0.24	1.24	1.17 to 1.31	0.26	1.26	1.17 to 1.36
PYH covariates																		
CMHHP				0.50	1.64	1.55 to 1.73	0.32	1.36	1.27 to 1.46	0.32	1.36	1.27 to 1.46	0.36	1.41	1.32 to 1.50	0.28	1.27	1.18 to 1.38
Ethnicity													-0.15	0.83	0.74 to 0.94	-0.28	0.71	0.55 to 0.92
Developmental problems													0.17	1.16	1.03 to 1.30	-0.08	0.88	0.64 to 1.20
Incontinence													0.25	1.25	1.11 to 1.41	0.14	1.10	0.90 to 1.34
Negative weight perception													NA	NA	NA	-0.10	0.85	0.66 to 1.11
School problem													0.27	1.28	1.12 to 1.46	0.14	1.10	0.97 to 1.25
Secondary school level low													-0.28	0.73	0.55 to 0.95	-0.20	0.77	0.70 to 0.86
Secondary school level high													NA	NA	NA	-0.11	0.85	0.75 to 0.97
High technology use													-0.34	0.68	0.53 to 0.88	-0.19	0.78	0.68 to 0.89
SDQ border area				0.33	1.38	1.23 to 1.54	0.22	1.24	1.06 to 1.43	0.22	1.24	1.06 to 1.43	0.31	1.34	1.20 to 1.49	0.21	1.19	1.02 to 1.38
SDQ increased				0.51	1.65	1.44 to 1.90	0.43	1.52	1.29 to 1.79	0.43	1.52	1.29 to 1.79	0.45	1.54	1.34 to 1.77	0.41	1.47	1.24 to 1.74
KIVPA													0.56	1.75	1.57 to 1.94	0.54	1.67	1.50 to 1.86
Under treatment													0.05	1.02	0.93 to 1.13	-0.12	0.84	0.74 to 0.94
Extra healthcare visit													0.19	1.18	1.10 to 1.26	0.31	1.32	1.18 to 1.48
Life events													0.12	1.10	1.02 to 1.20	0.19	1.16	1.05 to 1.28
Family history of MHP													0.22	1.22	1.08 to 1.37	0.29	1.30	1.03 to 1.62

Continued



Table 3 Continued

	Step 1*				Step 2				Step 3						
	Age 4–11 years 171 577 person years, no of events 8204	Age 12–19 years 98 754 person years, no of events 5947	Age 4–11 years 105 371 person years, no of events 6236	Age 12–19 years 75 106 person years, no of events 3664	Age 4–11 years 105 371 person years, no of events 6236	Age 12–19 years 75 106 person years, no of events 3664	Age 4–11 years 105 371 person years, no of events 6236	Age 12–19 years 75 106 person years, no of events 3664	Coef	OR	95% CI	Coef	OR	95% CI	
GP covariates	Coef	OR	95% CI	Coef	OR	95% CI	Coef	OR	95% CI	Coef	OR	95% CI	Coef	OR	95% CI
Prenatal risk factors															
Shrinkage factor B=500	0.99			0.99			0.98			0.97			0.95		
C-statistic corrected	0.62			0.63			0.64			0.63			0.63		
Brier	0.05			0.05			0.05			0.06			0.05		

Determinants from general practice that are not displayed in the table because of no significant results in the analyses: perinatal morbidity, disabilities, difficult temperament, somatic complaints+chronic disease. Determinants from PYH that are not displayed in the table because of no significant results in the analyses: premature, neonatal problems, non-spontaneous birth, sleeping, eating, overweight, underweight, bullying/being bullied, low self-confidence/resilience, self-harm, member hobby/music club, insufficient physical exercise, substance use, energy drink consumption, poorly experience health, total referral in past year, chronic illness parent, risk factor parents, non-traditional family, composition, negative balance, little confidence in parenting skills, environmental stressors.

*The first two columns age 4–11 years and age 12–19 years are regarding the full population of 70 000 children and are shown for comparison as they were published before.⁹ data regarding the remaining columns are calculated with the children with both PYH and GP data.
Coef, coefficient; GP, general practitioner; KIVPA, short indicative questionnaire for psychosocial problems among adolescents; MHP, mental health problem; PYH, preventive youth healthcare; SDQ, Strengths and difficulties questionnaire; SES, socioeconomic status.

anxiety as possible predictors. In our study, MHP symptoms with specific ICPC codes such as ‘feeling depressed’ were included as outcome, as according to our expert panel GPs are more cautious to label children with an actual mental health disorder ICPC code.⁹ Moreover, we aimed to improve the early recognition of child MHP and the inclusion of MHP symptoms as outcome might enable early identification. Furthermore, previous studies investigating the diagnostic properties of screening tools have shown that screening tools have added value in the (longer-term) identification of child MHP, but are not able to identify all children correctly.^{30 31}

Implications

Our study showed that several determinants registered in PYH, for example, increased problem scores on routinely used MH screening tools and registered concerns for MHPs, were identified as risk factors for MHPs. Although this information did not substantially improve the prediction models, it could still be useful information for GPs in daily practice. Especially as our study reported that nearly a third of the children for whom concerns for MHP were registered by PYHPs, had no registered MHPs in the GP data. The purpose of combining data from both professions was to explore the benefits of information exchange between PYHPs and GPs. From qualitative research it is known that Dutch GP’s currently in general have no structural interactions with PYHPs other than occasional referral letters and that both professionals feel the need of better information exchange.³² The standard exchange of for example the results of MHP screening instruments might therefore be useful for GPs. Future studies should investigate whether this type of information is indeed what GPs need and the practical implications of structural information exchange.

FUTURE STUDIES

This study used among others coded information regarding symptoms and diagnoses from general practice. Due to privacy reasons, in this case we had no access to free text notes in which GPs and PYHPs would describe the subjective patient’s story and symptoms. These notes could typically contain important information regarding social determinants for MHPs, such as functioning at school, family environment and life events.¹⁰ Machine learning techniques and in particular natural language processing techniques have shown promising results with EHR data including free text.³³ We recommend future studies to apply these techniques and also to investigate what the views of clinical professionals are regarding the use of the often called ‘black box’ models developed with these techniques in daily practice.

In addition, this study explored the development of prediction models for MHP recognition based on GP data, to support clinical MHP recognition. By doing so, we only predicted problems that were recognised by GPs, not the problems that were not recognised. More adequately

performing models based on EHR data might be developed making use of linkage with other domains like secondary care, thus enriching data to improve confirmation of diagnoses. Once this has been achieved, it should be investigated whether the resulting algorithms indeed improve recognition of MHPs in children that currently remain unrecognised by healthcare professionals.

CONCLUSION

This population-based cohort study investigated the added value of combining information from Dutch PYHPs and GPs in the identification of children with MHPs. Although the use of combined information did not improve prediction of MHPs compared with the use of GP data alone, this study showed the feasibility of analysing a combined dataset from different healthcare providers what has the potential to inform future studies aimed at improving child MHP identification.

Contributors NK conceptualised and designed the study, carried out the analyses, drafted the initial manuscript and revised the manuscript. FB, MC and MEN conceptualised and designed the study, had access to the data and supervised the analyses, critically reviewed and revised the manuscript for important intellectual content. MEN is the guarantor of this study. IP, NL and AvD-vD conceptualised the study, provided the data, participated in the interpretation of the data, reviewed and revised the manuscript. RRJMV conceptualised the study, participated in the interpretation of the data, reviewed and revised the manuscript for important intellectual content. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding Financial support by ZonMw, the Netherlands, Organisation for Health Research and Development (grant 839110012).

Disclaimer ZonMw did not have any role in study design, the collection, analysis and interpretation of data, the writing of the report and the decision to submit the paper for publication. The authors are not dependent of the funders and had full access to all of the data in the study. The authors can take responsibility for the integrity of the data and the accuracy of the data analysis.

Competing interests All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and declare no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years, no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication Not applicable.

Ethics approval The Ethics Committee of the Leiden University Medical Centre issued a waiver of consent (G16.018).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Due to the nature of the included data, the individual patient data will not be made available to others.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is

properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Nynke R Koning <http://orcid.org/0000-0002-2532-1180>

Frederike L Büchner <http://orcid.org/0000-0001-8977-5344>

REFERENCES

- Steel Z, Marnane C, Iranpour C, *et al*. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. *Int J Epidemiol* 2014;43:476-93.
- Kim-Cohen J, Caspi A, Moffitt TE, *et al*. Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Arch Gen Psychiatry* 2003;60:709-17.
- Kessler RC, Berglund P, Demler O, *et al*. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National comorbidity survey replication. *Arch Gen Psychiatry* 2005;62:593-602.
- James AC, James G, Cowdrey FA. Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane database of systematic reviews* 2013;6.
- Weitzman CC, Leventhal JM. Screening for behavioral health problems in primary care. *Curr Opin Pediatr* 2006;18:641-8.
- Rothman B, Leonard JC, Vigoda MM. Future of electronic health records: implications for decision support. *Mt Sinai J Med* 2012;79:757-68.
- Goldstein BA, Navar AM, Pencina MJ, *et al*. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc* 2017;24:198-208.
- Nichols L, Ryan R, Connor C, *et al*. Derivation of a prediction model for a diagnosis of depression in young adults: a matched case-control study using electronic primary care records. *Early Interv Psychiatry* 2018;12:444-55.
- Koning NR, Büchner FL, Vermeiren RRJM, *et al*. Identification of children at risk for mental health problems in primary care-development of a prediction model with routine health care data. *EClinicalMedicine* 2019;15:89-97.
- de Wolff MS, Theunissen MHC, van Rooijen K. JGZ-richtlijn psychosociale problemen. *JGZ Tijdschrift voor jeugdgezondheidszorg* 2017;49:90-2.
- Burger H, Boks MP, Hartman CA, *et al*. Risk score for predicting adolescent mental health problems among children using parental report only: the trails study. *Acad Pediatr* 2014;14:589-96.
- vES BE, Timmermans AE. *Handreiking samenwerking huisarts jeugdgezondheidszorg*, 2008.
- Dunnink G, Lijs-Spek W. Activiteiten basistakenpakket jeugdgezondheidszorg 0-19 jaar per contactmoment. *RIVM rapport* 2008.
- Lamberts H, Wood M. *ICPC, international classification of primary care*. USA: Oxford University Press, 1987.
- Methodology WCCfDS. *ATC index with DDDs. WHO collaborating centre for drug statistics methodology*. Oslo, Norway, 2002.
- Koning NR, Van den Berg AW, Choi SYA. The usefulness of electronic health records from preventive youth healthcare in the recognition of child mental health problems. *Front Public Health* 2020.
- STIZON - Stichting Informatievoorziening voor Zorg en Onderzoek. Available: <https://www.stizon.nl/> [Accessed 01 Aug 2018].
- Koning NR, Büchner FL, Verbiest MEA, *et al*. Factors associated with the identification of child mental health problems in primary care-a systematic review. *Eur J Gen Pract* 2019;25:116-27.
- Wells BJ, Chagin KM, Nowacki AS, *et al*. Strategies for handling missing data in electronic health record derived data. *EGEMS* 2013;1:7.
- Bouwmeester W, Twisk JWR, Kappen TH, *et al*. Prediction models for clustered data: comparison of a random intercept and standard regression model. *BMC Med Res Methodol* 2013;13:19.
- Harrell Jr FE. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, 2015.
- Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 1950;78:1-3.
- Schmidt M, Schmidt SAJ, Sandegaard JL, *et al*. The Danish national patient registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015;7:449.



- 24 Nyberg F, Franzén S, Lindh M, *et al.* Swedish Covid-19 Investigation for Future Insights - A Population Epidemiology Approach Using Register Linkage (SCIFI-PEARL). *Clin Epidemiol* 2021;13:649–59.
- 25 Carson L, Jewell A, Downs J, *et al.* Multisite data linkage projects in mental health research. *Lancet Psychiatry* 2020;7:e61.
- 26 Downs JM, Ford T, Stewart R, *et al.* An approach to linking education, social care and electronic health records for children and young people in South London: a linkage study of child and adolescent mental health service data. *BMJ Open* 2019;9:e024355.
- 27 Ford JG, Howerton MW, Lai GY, *et al.* Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. *Cancer* 2008;112:228–42.
- 28 McGrath JJ, Wray NR, Pedersen CB, *et al.* The association between family history of mental disorders and general cognitive ability. *Transl Psychiatry* 2014;4:e412–e.
- 29 Donders ART, van der Heijden GJMG, Stijnen T, *et al.* Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006;59:1087–91.
- 30 Lavigne JV, Meyers KM, Feldman M. Systematic review: classification accuracy of behavioral screening measures for use in integrated primary care settings. *J Pediatr Psychol* 2016;41:1091–109.
- 31 Nielsen LG, Rimvall MK, Clemmensen L, *et al.* The predictive validity of the strengths and difficulties questionnaire in preschool age to identify mental disorders in preadolescence. *PLoS One* 2019;14:e0217707.
- 32 Koning N, Büchner F, Numans M. Collaboration between general practitioners and preventive youth health physicians: room for improvement. *Ned Tijdschr Geneesk* 2018;162.
- 33 Edgcomb JB, Zima B. Machine learning, natural language processing, and the electronic health record: innovations in mental health services research. *Psychiatr Serv* 2019;70:346–9.