Midwifery ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Midwifery

journal homepage: www.elsevier.com/midw

Feasibility and reliability of a newly developed antenatal risk score card in routine care

Mieke J. van Veen, RM, MSc (Researcher Midwife)[a,*], Erwin Birnie, PhD (Senior Researcher Fellow)[b,c], Jashvant Poeran, MD, PhD (Post-Doctoral Associate of Epidemiology)[d], Hanneke W. Torij, MSc (Associate Professor Midwifery & Childbirth)[a], Eric A.P. Steegers, MD, PhD (Professor in Obstetric and Prenatal Medicine)[c], Gouke J. Bonsel, MD, PhD, MPH (Professor in Perinatal Care and Public Health)[c,e]

[a] Rotterdam University of Applied Sciences, Research Centre Innovations in Care, Rotterdam, The Netherlands
[b] Institute of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands
[c] Department of Obstetrics and Gynaecology, Division of Obstetrics & Prenatal Medicine, Erasmus MC, University Medical Centre Rotterdam, The Netherlands
[d] Weill Cornell Medical College, Department of Public Health, Division of Biostatistics and Epidemiology, New York, USA
[e] Department of Public Health, Erasmus MC, University Medical Centre Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 19 February 2014
Received in revised form
30 June 2014
Accepted 1 August 2014

Keywords:
Antenatal risk assessment
Feasibility
Reliability

ABSTRACT

Objective: to study in routine care the feasibility and inter-rater reliability of the Rotterdam Reproductive Risk Reduction risk score card (R4U), a new semi-quantitative score card for use during the antenatal booking visit. The R4U covers clinical and non-clinical psychosocial factors and identifies overall high risk pregnancies, qualifying for intensified antenatal care.
Design: a population-based cross-sectional study (feasibility) and a cohort study (inter-rater reliability).
Setting: feasibility was studied in six midwifery practices and two hospitals; the reliability study was performed in one midwifery practice.
Participants: 1096 pregnant women in the feasibility study and a subsample of 133 participants in the inter-rater reliability study.
Measurements: feasibility was expressed as (a) time needed to complete the R4U and (b) the missing rate at the item and client level. For inter-rater reliability (IRR) an independent, blinded, caregiver completed a re-test R4U during a second visit; inter-rater agreement for each item and all domain sum scores were computed.
Findings: completion of the R4U took 5 minutes or less in 63%; and between 5 and 10 minutes in another 33%. On the participant level 0.2% of women had $>20\%$ missing values (below 4% threshold, $P<0.001$). One of 77 items had a $>10\%$ missing rate. The per item IRR was 100% in 20% of the items, and below the predefined 80% threshold in 13% of the items ($n=9$). The domain sum scores universally differed less than the predetermined $\pm 15\%$ margin.
Key conclusion: the R4U risk score card is a feasible and reliable instrument.
Implication for practice: the R4U is suitable for the assessment of clinical and non-clinical risks during the antenatal booking visit in a heterogeneous urban setting in routine practice.

Introduction

The Netherlands show a high perinatal mortality rate (nine per 1000 births) compared to other European countries (EURO-PERISTAT Project in Collaboration with SCPE, 2010). Eighty-five per cent of the cases of perinatal mortality are associated with the presence of one or more of the following: congenital anomaly, small for gestation age (SGA, birth weight $<$P10 adjusted for gestational age), preterm birth (<37 weeks), or low Apgar score (<7, five minutes after birth) (van der Kooy et al., 2011). Under current guidelines 25% of these conditions (excluding low Apgar score) are not recognised prior to the onset of delivery. Hence, there is room for improvement in the current system of antenatal risk assessment (Bonsel et al., 2010).

Enhanced risk assessment can be considered in all trimesters of pregnancy, e.g. by means of the collection of biomarkers and ultrasound measurements in the second and third trimester

* Correspondence to: Rotterdam University of Applied Sciences, Research Centre Innovations in Care, Room RS 02.130, PO Box 25035, NL-3001 HA Rotterdam, The Netherlands.
E-mail address: m.j.van.veen@hr.nl (M.J. van Veen).

http://dx.doi.org/10.1016/j.midw.2014.08.002
0266-6138/

respectively. Here we focus on improved non-invasive risk assessment in early pregnancy through checklist-wise history taking with a broader than usual scope. Large cohort studies and public health reports revealed an independent contribution of non-clinical perinatal risk factors, particularly in deprived environments where perinatal mortality is high (Kleijer et al., 2005; (Agyemang et al., 2009; Timmermans et al., 2011; Poeran et al., 2013). Low socio-economic status, domestic violence, psychosocial morbidity and, more generally, living in a deprived neighbourhood are examples of evidence-based non-clinical risk factors which add to adverse perinatal outcome (Goedhart et al., 2008; de Graaf et al., 2013). These risks, taken apart of only moderate impact for e.g. SGA or preterm birth, supposedly act through accumulation (Timmermans et al., 2011). Unlike obstetrical and medical risk factors, history taking in routine care only partially covers these non-clinical risk factors, if covering them at all. Their recording also shows high variability. Appropriate coverage of these risks may enhance awareness for cases with a high combined risk load, and induce active prevention in due time.

As part of a regional, currently national public health initiative, aimed at improving birth outcomes in the Dutch urban areas, we developed an antenatal risk score card for universal use in routine care: the so-called 'Rotterdam Reproductive Risk Reduction for You (R4U)' (Denktas et al., 2012). The R4U equally assesses clinical and non-clinical risks, thereby allowing the estimation of a cumulative risk profile. It translates results from, in particular, international birth cohort studies into a pragmatic risk score card for perinatal use, the format itself being derived from the existing WIC program in the United States (WIC: Women, Infant and Children Program). After initial piloting we report in this study the feasibility and reliability of this tool as measured under routine care by midwives and obstetricians.

Methods

R4U risk score card

Concept

The R4U risk score card consists of 46 non-clinical and 31 clinical items, conveniently grouped into six domains (Table 2). The left hand side of the R4U form (paper version; see Appendix 5) covers four domains: psychosocial and economic, communication and ethnicity, pregnancy onset, and lifestyle. The right hand side of the form encloses the clinical risk items in the medical (e.g. psychiatry, cardiovascular) and the obstetrical domains.

Response is generally dichotomous (yes/no presence of risk); if a measure is continuous, a cut-off point (risk-threshold) is stated. After completion of the card by the caregiver, the positive responses can be summed into a domain score, and subsequently into an overall sum score. This sum score represents the accumulated risk for perinatal morbidity (preterm birth, SGA, congenital anomalies); a setting-defined threshold of 'high risk' can be applied. When computerised, the R4U is an adaptive instrument (i.e., skips the non-relevant items) that automatically generates sum scores.

The R4U rests on the following principles: (a) non-clinical risk factors (see below) are regarded as independent risk factors, some of them amenable for intervention or support; (b) while a single risk may act as cause, inviting for a specific intervention, accumulation of heterogeneous risks is regarded as another pathway to perinatal morbidity (particularly IUGR and prematurity), justifying the computation of summated risk scores even if not all factors are amenable for prevention or treatment (Timmermans et al., 2011); (c) the early detection of a high sum score is regarded as an extra preventive opportunity; it is the total risk profile that defines the urgency of the case and enables the timely choice for particular preventive or curative strategies; (d) the built-in thresholds of high risk are regarded relative, to enable accommodation to the available (preventive) resources; cut-offs at the individual risk level and the sum score level can be adapted; and (e) the assessment of the R4U should have minimal impact the interaction with the client.

Non-clinical risk factors

Examples of non-clinical risk factors are domestic violence, poor education, single parenthood, low income and serious debts, psycho-social problems, having a migrant background, or being a teenager with an unplanned pregnancy. These factors increase the risk for e.g. SGA and premature birth, although biological pathways sometimes are elusive, or only partially understood. They are not systematically covered in current history taking.

Format

The R4U was developed as part of innovative regional projects to improve birth outcomes in urban areas (Denktas et al., 2012). The 77 clinical and non-clinical items were primarily obtained from three sources: (a) results from two ongoing large urban birth cohort studies: Generation R in Rotterdam (Jaddoe et al., 2012) and ABCD in Amsterdam (Vrijkotte et al., 2009); (b) 23 comprehensive international perinatal studies (see Appendix 1); and (c) a targeted literature search for each candidate risk variable to support decisions on its independent relevance and risk threshold (see Appendix 2). The final version of the R4U used in this study was developed through the multidisciplinary input from obstetricians, midwives, social scientists, public health experts, representatives from municipal health service and social services, psychiatrists, teenage support groups and addiction care experts. The detailed development and design of the R4U score card will be described elsewhere.

To increase the feasibility, we supplemented each item with a so called 'script' (see Appendix 3), that contains the literal question on how to pose questions on sensitive topics like domestic violence; this is an optional aid to caregivers. We instructed caregivers to ask the items as open questions (see the script [see Appendix 3]) and to interview in depth in case of unequivocal answers, aiming to reach a dichotomous answer. In case risk factors were identified, caregivers were requested to ask for advice or to refer women to an appropriate follow-up caregiver, as part of usual care.

The format of the R4U score card (available upon request) was adopted from the paper and pencil format of the Women Infant Children (WIC) checklist which has proved valuable in the United States in routine practice in deprived areas (WIC: Women, Infant and Children Program).

The resulting null version of the R4U (items, response, script, score card presentation) was judged by clinicians not engaged in the development, and was piloted among 19 pregnant women at the antenatal booking. From this experience and from a second round of independent consultation of domain experts (notably on psychiatry and nutrition), the format remained unchanged but some items were rephrased. This final version (presented in Appendix 5) was studied for its suitability as routine assessment tool in the current study.

Study design: feasibility and reliability

We examined two psychometric features under routine care conditions: feasibility and reliability (see below for definitions). The feasibility was addressed in a 18-month multicentre cross-sectional study (started November 25, 2010) in six independent

Please cite this article as: van Veen, M.J., et al., Feasibility and reliability of a newly developed antenatal risk score card in routine care. Midwifery (2014), http://dx.doi.org/10.1016/j.midw.2014.08.002

midwifery practices and two hospitals; altogether 20 primary care midwives, 22 obstetrics/gynaecology residents and two nurses were involved. Pregnancy characteristics of the participating midwifery practices resembled the 2011 national averages in terms of nulliparity (44% versus 46%) and maternal age <30 years (55% versus 42%).

The participating hospitals showed lower proportions of nulliparity (32% versus 52%) and a slightly lower maternal age <31 years (46% versus 51%).

The inter-rater reliability (IRR) study was conducted as a single-centre cohort study between January 9, 2012 and June 26, 2012 and was nested within the feasibility study. The IRR study comprised the completion of two R4U score cards, one at booking visit by the regular midwife and one at a repeat visit by an external midwife who was blinded for the first R4U and other information from the booking visit. During the study period all consecutive women were invited to participate. They received verbal and written information by an independent research-midwife (MJvV). For those consenting, the repeat visit was scheduled close to the index visit. To reduce memory effects and interval changes in risk level, the intended interval between the booking visit and the repeat visit was set at two weeks. The clients of the IRR study were rewarded 10 euro to cover travel expenses. For both studies informed consent was not deemed necessary by the Institutional Review Board of the Erasmus MC (permission MEC-2010-332) as an inventory of risk factors in this case was considered as normal antenatal care, and as the focus was standardisation of history taking of otherwise non-disputed clinical and non-clinical factors.

Participants

All pregnant women were invited at their booking visit to participate in the feasibility study, except for known multiple gestations. Inclusion criteria of the IRR study were informed consent at the booking visit and sufficient Dutch or English language proficiency.

Data management and data collection

All participating caregivers received a short introduction on the use and scoring of the R4U and were individually trained by a member of the study team at three consecutive booking visits. As part of the study monitoring, the research-midwife (MJvV) visited the sites and each time recorded remarks of the professionals as well as clients' remarks as reported by the professionals. These remarks were grouped into categories ex post.

Outcome measures

Feasibility was expressed as (a) time needed to complete the R4U; and (b) missing rates on item and on client levels. Time needed to complete was used as a direct measure of feasibility; i.e. the more time is needed to complete the R4U (relative to its yield), the less likely it is that the R4U will be adopted in routine midwifery or obstetric care. Time needed to complete was also used as an indirect measure of validity, as appropriate history taking of multirisk cases inevitably involves more time. Completion time was considered acceptable if 95% of the R4Us were completed in <15 minutes. We compared the time to complete the R4U in the first 15 clients (R4Us) versus the remaining clients (R4Us) to test for a learning effect.

Missing rate on the item level was considered acceptable if no single item had a missing rate exceeding 10%. Missing rate on participant level was considered acceptable if no more than 4% of clients had in excess of 20% missing items. Note that this criterion refers to a paper and pencil version. The inter-rater reliability was expressed as, at the item level, the agreement between the first and second midwife (technically 'accuracy'); it was considered good if agreement was at least 80% for at least 80% of the items. For six R4U items inter-rater agreement was not applicable (e.g. gestational age at the time of the current visit, weight/BMI as these items are likely to change between visits).

We added risk factors (unweighted) to obtain sum scores: one overall and six domains sum scores and the overall non-clinical and clinical sum scores of the R4U. We defined acceptable agreement of any sum score to be present if the repeat sum score did not differ by $\pm 15\%$ from the first sum score. If agreement satisfied all the pre-set criteria, we regard this 'excellent' in view of the high thresholds applied.

A short post hoc debriefing interview with the midwives was scheduled prior to analysis.

Statistical analysis and sample size

For description of the baseline characteristics we used conventional statistics. To test whether completion time >15 minutes actually was within defined limits, we calculated the upper confidence limit of the observed proportion of completion times in excess of >15 minutes which should be $<5\%$ (binomial test). The distribution of completion times in midwifery practices was then compared to hospitals (Fisher's exact test). The presence of a learning effect was tested in midwifery practices and hospitals (Fisher's exact test).

Missing rates at the item and the client level were compared with the predefined limits (binomial test). Whether the proportion of score cards with 20% or more missing items exceeded the predefined maximum of 4%, was checked through testing whether the upper limit of the 95% confidence interval (binomial distribution) of the frequency of score cards with $>20\%$ missings indeed included 4%, which should not be the case. A $P<0.01$ indicates that even 99% of the distribution was below this chosen threshold. The inter-rater agreement on the item level and domain level was similarly checked with the binomial test. Sum scores of midwifery practice clients versus hospital clients were compared with the Mann–Whitney U test.

In all comparisons we tested with alpha 0.05 (95% confidence intervals). For all comparisons of feasibility criteria with predefined limits, a sample size of $n<250$ clients sufficed. As little was known on the performance of the instrument under unsupported routine care and as some items reflected rare events, we aimed at 1000 clients, which also allowed for subgroup analysis if required.

The sample size of the IRR study was based on the 41 items of the non-clinical domains as these were expected to show lowest reliability. From the pilot study we expected 4/41 (10%) of items to be discordant, and we aimed at excluding discordance to exceed 8/41 (20%) discordant items; with $n=110$ this test has a power of 0.80 with two sided alpha 0.05.

Findings

Baseline characteristics

Table 1 shows clients' characteristics by midwifery practices and hospitals. Overall midwifery clients had a more 'healthy' profile: being younger, living less often in a deprived neighbourhood, and less often from non-Western origin. The participating clients in the IRR study showed little difference with the non-participating booking visit clients.

Please cite this article as: van Veen, M.J., et al., Feasibility and reliability of a newly developed antenatal risk score card in routine care. Midwifery (2014), http://dx.doi.org/10.1016/j.midw.2014.08.002

Table 1
Characteristics of the clients in the feasibility study by type of caregiver ($n=1096$) and the inter-rater reliability study, $n=133$.

	Feasibility study*				Inter-rater reliability study† Midwifery practice			
	Midwifery practices		Hospitals		Participant clients		Non-participant clients	
	$n=694$	%	$n=402$	%	$n=133$	%	$n=151$	%
Age group								
12– < 20 years	32	5	11	3	3	2	6	4
20– < 30 years	349	50	171	43	64	48	74	49
30– < 40 years	293	42	191	48	65	47	68	45
≥ 40 years	14	2	28	7	3	2	3	2
Parity								
Nulliparous	302	44	128	32	90	68	69	46
Neighbourhood								
Deprived	188	27	133	33	60	45	72	52
Social economic status								
SES < P20 (deprived)	479	69	255	63	99	74	113	75
SES P20–P80	172	25	85	21	33	25	36	24
SES > P80 (wealthy)	38	5	59	15	1	1	2	1
Ethnicity								
Non-Western	113	16	93	23	23	17	51	34

* Missing: age, $n=7$, parity, $n=2$; SES, $n=8$.
† Missing: age, $n=1$.

Time to complete the R4U

The time to complete the R4U was < 5 minutes in 63% of cases, and 5–10 minutes in 33% of cases. More than 15 minutes to complete the R4U was needed in 1.2% of women, which satisfies the pre-set < 5% threshold ($P<0.001$). Midwifery practices generally had shorter completion times (see Fig. 1); in midwifery practices 0.3% of the R4Us needed more than 15 minutes versus 2.7% in the hospitals ($P<0.001$).

Fig. 1 shows that the time to complete the R4U is somewhat less from the 16th client onward, compared to the first 15 R4Us. The first 15 R4Us of each midwifery practice were completed within 5 minutes in 51% of the R4Us versus in 77% from the 16th client onwards ($P<0.001$). In the hospitals these proportions were 13% versus 45%, respectively ($P<0.005$). Regarding the relationship between number of risks and completion time: If no or only one risk factor was present, completion within 5 minutes was 81%; the proportion completing within 5 minutes gradually decreased to 35% if the number of risk factors was 14–20. Still 80% was less than 10 minutes (detailed data available upon request).

Missing rate at item level

Table 2 shows that 28/77 (36%) items in midwifery practices and 15/77 (19%) items in the hospitals had no missing values ($P=0.03$). The items with missing values between 0.1% and 2.0% occurred in 58% of the items recorded in midwifery practices and 5% of the items had missing values between 2.1% and 10.0%. In the hospitals, these proportions were 49% and 29%. Overall, one item 'no immunity Rubella' (1.3% of the items) had more than 10% missing values, significantly exceeding the 0% threshold ($P<0.005$). This item had the highest percentage of missing values in the hospital group (21%).

Missing item rate at client level

The median number of missing values at the client level (that is, per participant) was one for both the midwifery practices and hospitals; the respective means were 0.97 and 1.85 ($P<0.001$). In midwifery practices, 49% of the clients had no missing values compared to 30% of clients in hospitals ($P<0.001$). Two out of the 1096 (0.2%) R4Us showed more than 20% missing values, which satisfied the predefined limit of < 4% ($P<0.001$).

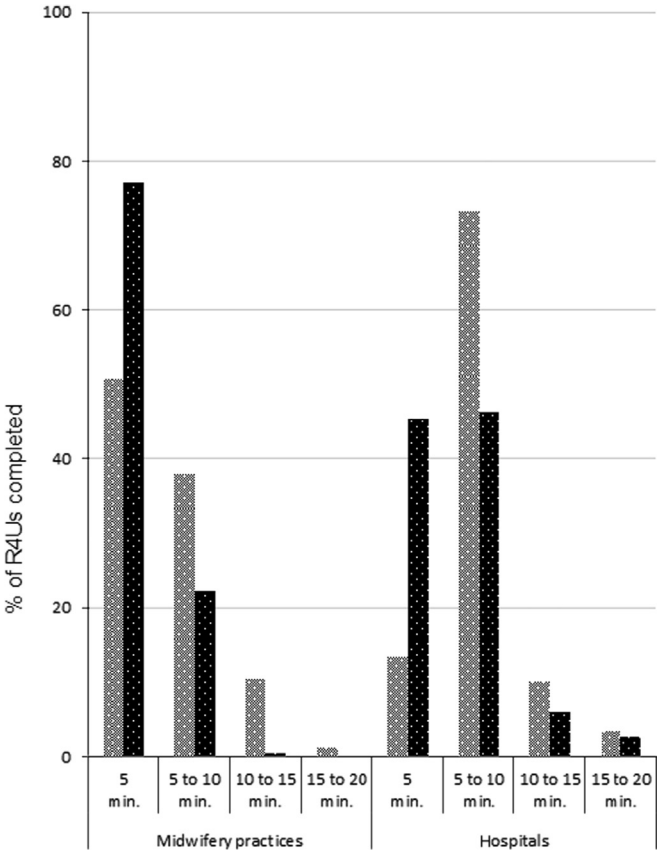

Fig. 1. Time needed to complete the R4U in midwifery practices and hospitals, and between the first 15 clients and from the 16th client onward ($n=1042$). ▩: First 15 clients, ■: 16th clients onward.

IRR: inter-rater agreement

The time interval between booking and the repeat visit ranged from one day to three weeks (3.8% unknown) in 81% of the R4Us; the median interval was 10 days. The inter-rater agreement for all items was 100% for 14/71 items, between 80% and 99% for 48/71 items and < 80% (threshold) for 9/71 items. The 9/71 (13%) all item result did not satisfy the predefined < 20% threshold ($P=0.076$) (Table 2).

The inter-rater agreement for the non-clinical items was 100% for 5/41 items (12%), between 80% and 99% for 32/41 items (78%) and < 80% for 4/41 items (10%). The confidence interval of 4/41 (10%) included the predefined 20% threshold ($P=0.066$). Table 3 shows for each R4U domain the inter-rater difference of the sum scores, being between 0 and 5. The difference in sum scores was < 15% for all domains.

Accumulation of risk factors: sum scores

Fig. 2 shows the accumulation of risk factors of the non-clinical and clinical domains of the R4U, for midwifery practices and hospitals separately, showing a systematic excess risk in the hospital population. The proportion of risk-free clients was highest in the pregnancy onset domain (77%) and lowest in the medical domain (19%).

Please cite this article as: van Veen, M.J., et al., Feasibility and reliability of a newly developed antenatal risk score card in routine care. Midwifery (2014), http://dx.doi.org/10.1016/j.midw.2014.08.002

Table 2
Prevalence of adverse risk items and proportions of missing values at the item level by type of caregiver ($n=1096$); and inter-rater agreement, $n=133$.

	Prevalence of adverse risk %		Missing at item level* %		Inter-rater agreement* %
	Midwifery practices $n=694$ clients	Hospitals $n=402$ clients	Midwifery practices $n=694$ clients	Hospitals $n=402$ clients	Midwifery practice $n=133$ clients
PSYCHOSOCIAL AND ECONOMIC DOMAIN					
Social situation					
Single mother	6.9	9.5	0.0	0.5	95
Relational problems > 3 months	5.0	4.5	1.4	0.7	95
No social support	2.4	5.0	0.3	0.5	96
Only 1–2 persons for social support	5.2	8.2	0.4	1.0	90
Domestic violence	1.4	2.0	1.0	0.7	95
Youth health care intervention last two years	2.4	4.7	1.5	2.1	99
Work and income					
Unemployed (> 3 months)	24.6	25.9	0.5	0.7	84
Exposure to standing work	24.2	19.9	1.0	2.4	90
Working hours > 32 hours	12.2	18.4	1.5	1.4	87
Net family income < 1000 euro	9.8	10.4	1.5	1.0	98
Irredeemable financial debts	6.6	9.7	2.0	0.7	96
Partner unemployed	12.1	8.5	0.5	2.1	94
Educational level					
Low educational level or illiterate	2.6	7.5	0.5	3.4	99
Living conditions					
Housing problems	5.0	4.7	0.5	1.7	94
Deprived neighbourhood	26.9	27.6	0.5	**16.9**	88
COMMUNICATION AND ETHNICITY DOMAIN					
Ethnicity					
Creole-Surinamese	6.2	5.2	1.0	2.1	97
Hindu-Surinamese	3.7	6.0	1.0	2.1	99
Antillean-Aruban	3.9	6.0	0.5	1.4	99
African	3.0	5.5	0.5	1.7	96
Eastern European	6.1	2.7	0.5	2.1	98
Other non-Western	16.3	23.1	0.5	2.4	83
Language/communication with client					
Language barrier	7.2	5.5	0.0	1.4	97
Only communication with translator	2.9	2.7	0.5	1.4	100
Mentally retarded	0.1	0.5	0.0	2.4	100
PREGNANCY ONSET					
General					
No health insurance	0.9	2.7	0.5	0.0	100
Family planning/age					
Unwanted pregnancy	5.3	6.2	0.5	0.0	96
Assisted reproduction	3.5	11.7	0.0	0.3	99
Teenage pregnancy	3.6	2.0	0.5	0.0	99
Advanced maternal age (> 40 years)	2.0	6.7	0.0	0.0	100
Obstetric					
Start antenatal care after 14 weeks	4.5	6.7	1.0	2.1	N/A
Start antenatal care after 24 weeks	1.0	4.0	1.0	2.1	N/A
LIFESTYLE DOMAIN					
Substance abuse					
Preconceptional smoking	25.4	22.9	0.0	0.3	90
Smoking first trimester	13.5	18.2	0.0	5.2	98
Smoking second trimester	1.6	4.7	1.0	8.3	**27**(99)1;†
Preconceptional alcohol use	28.8	7.0	0.5	0.3	**58**
Alcohol abuse first trimester	2.7	2.0	0.0	5.9	97
Alcohol abuse second trimester	0.0	0.5	1.5	5.5	**27**(100)1;†
Preconceptional drug use	2.6	2.2	0.5	0.3	96
Drugs abuse first trimester	1.0	1.2	0.0	6.2	99
Drugs abuse second trimester	0.1	1.0	1.0	6.6	**26**(100)1;†
Nutrition					
Vegetarian, vegan or macrobiotic diet	3.3	2.0	1.0	0.3	100
Insufficient intake of vegetables (not daily)	15.0	15.9	1.5	0.7	84
Insufficient intake of fruit (less than two daily)	13.8	15.2	1.5	0.7	92
Weight					
BMI < 20	11.7	5.7	1.0	3.4	N/A
BMI 30–40	13.1	19.7	2.0	4.1	N/A
BMI > 40	1.0	4.7	1.5	4.1	N/A
MEDICAL DOMAIN					
Diseases					
Chronic maternal illness	4.6	19.4	0.0	1.4	95
Consultation clinician (last year)	16.4	26.4	0.5	1.7	80
Any surgery	39.8	60.0	0.5	1.0	**74**

Please cite this article as: van Veen, M.J., et al., Feasibility and reliability of a newly developed antenatal risk score card in routine care. Midwifery (2014), http://dx.doi.org/10.1016/j.midw.2014.08.002

Table 2 (*continued*)

	Prevalence of adverse risk %		Missing at item level* %		Inter-rater agreement* %
	Midwifery practices $n=694$ clients	Hospitals $n=402$ clients	Midwifery practices $n=694$ clients	Hospitals $n=402$ clients	Midwifery practice $n=133$ clients
Medication					
Prescribed medication	12.2	29.6	0.5	0.0	80
Over-the-counter drugs	10.5	26.9	0.5	1.4	**52**
No preconceptional folic acid use	45.2	46.3	1.5	0.3	**74**
Infectious diseases					
Last 12 months (treated) sexually transmitted disease	1.9	2.2	0.0	0.7	99
At risk of sexually transmitted disease	1.3	0.7	0.5	2.4	**70**
At risk of Toxoplasmosis	10.7	6.7	0.0	1.7	93
No immunity Rubella	5.0	4.0	7.4	**21.4**	N/A
Psychiatric					
History of psychiatric admission or positive family history (first degree relative)	4.8	3.7	0.5	0.3	98
Ever used psychiatric medication	4.2	7.5	0.5	0.0	97
Current psychiatric problems	3.2	3.0	0.0	0.0	99
OBSTETRIC DOMAIN					
History					
Time to pregnancy >1 year	8.1	20.1	0.0	0.3	**69**
Nulliparous	43.5	31.8	0.0	0.3	82
Recurrent miscarriage (two or more)	6.8	12.7	0.0	0.0	99
Preterm birth (<37 weeks)	2.3	8.0	0.0	0.3	100
Small for gestational age ($<$P10)	1.6	5.7	0.0	0.0	100
Major congenital anomalies	1.2	1.0	0.0	0.3	100
Still birth (22 weeks – 7 days pp)	0.4	3.7	0.0	0.3	100
Shoulder dystocia	0.7	0.5	0.0	0.0	100
Instrumental childbirth	6.5	7.2	0.0	0.0	97
Primary caesarean section	1.9	6.5	0.0	0.0	99
Secondary caesarean section	3.3	11.4	0.0	0.7	98
Gestational diabetes	0.6	5.5	0.0	0.3	100
Placental abruption	0.3	0.5	0.0	0.0	100
Manual placenta removal/hemorrrhage	2.9	5.2	0.0	0.0	96
(Pre)eclampsia or HELLP	0.7	4.7	0.0	0.7	100
Hemoglobinopathy	0.4	0.5	0.5	0.0	99
Jehovah's witness	0.4	0.7	1.9	0.2	100
Family					
Congenital anomaly in first or second degree relative	5.6	10.9	2.0	3.1	93

* In bold print: missing rates exceeding the predefined limit.

† These items had three response modes: Yes, No and N/A; apparently differed in using N/A where rater 1 used 'No' when actually 'N/A' was the correct reply. The first number is stated, the second number between brackets provides the IRR after recoding of the obvious error of rater 1.

Table 3
Inter-rater reliability: absolute differences in domain scores between two raters, $N=133$. Differences range between 0 (no difference) and 5 (maximum).

Absolute difference	Psychosocial and economic %	Communication and ethnicity %	Pregnancy onset %	Lifestyle	Medical %	Obstetric %
0	55.6	84.2	93.2	46.6	24.8	46.6
1	34.6	15.8	6.8	41.4	35.4	42.1
2	6.8	–	–	11.3	28.6	6.8
3	2.3	–	–	0.8	6.8	3.8
4	0.0	–	–	–	3.8	0.8
5	0.8	–	–	–	0.8	–
P-value*	<0.001	<0.001	<0.001	<0.001	0.163	0.001

* For each column (domain) is tested whether the proportion of absolute differences is lower than a predefined threshold (15%). A too large difference was defined relative to the distribution of sum scores of a domain (see Methods).

The median total sum sore of the R4U, combining the scores of all domains (unweighted), was 6; being 5 in midwifery practices and 7 in hospitals ($P<0.001$). Only 0.8% of the clients were risk free. There were more low risk clients (arbitrarily defined as having a total sum score 0–3) in midwifery practices than in hospitals: 24% versus 11%, $P<0.005$ (see Appendix 4).

Discussion

In this study a newly developed antenatal score card that allows a quick uniform screening of 71 clinical and non-medical antenatal risks, showed excellent feasibility and good to excellent inter-rater reliability in an unselected urban population. These

Please cite this article as: van Veen, M.J., et al., Feasibility and reliability of a newly developed antenatal risk score card in routine care. Midwifery (2014), http://dx.doi.org/10.1016/j.midw.2014.08.002

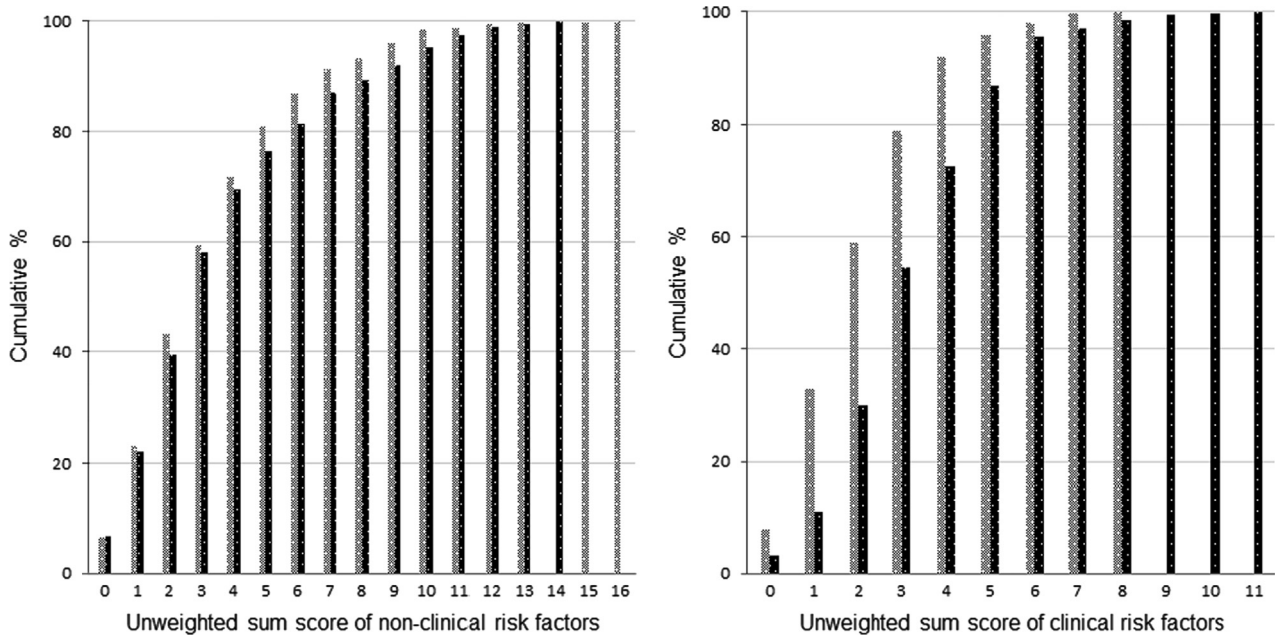

Fig. 2. Accumulation of risk factors: total of non-clinical risk factors (psychosocial and economic, communication and ethnicity, pregnancy onset and lifestyle domains; left panel) and clinical risk factors (medical and obstetric domains; right panel) in clients in midwifery practices and hospitals ($n = 1096$). ▒: Midwifery practices, ■: hospitals.

results were obtained in eight different care settings and did not show feasibility differences under routine care conditions. On average the sum score of all risks was slightly lower in midwifery practices than in hospitals, which should be expected as initial high risk patients have their first visit in the hospital. The rather small difference is in agreement with existing reports that initial risk guided choice of setting is limited (Bonsel et al., 2010). One should be aware that time needed to complete the R4U score card is a measure of feasibility but does not reflect the clinical value of the R4U. The systematic screening most likely will increase time needed to complete the first stage of antenatal screening in complex cases, but we consider this a positive result to be preferred to the detection of important social and psychological risks at a later stage.

Overall, missing rates were low. The higher missing rates in hospitals may be explained by less available time to hospital caregivers, in particular if set against the higher risk level which requires more communication time. Information on Rubella immunity was often missing due to the absence of a standardised lab protocol. The high missing rates of the item 'living in deprived neighbourhood' in the hospitals can be explained by the administrative failure to transfer the client's postal code on the R4U form, which is the source information required to fill out this item. All history taking items, however, appeared easy to complete.

Despite satisfactory overall and domain agreement, the IRR study showed that some items had low agreement, particularly alcohol abuse in preconception period. This may be due to differential perseverance of caregivers to check presence of such risks. Post-hoc debriefing of the midwives involved in the reliability study revealed that differential history taking on some items was present ('over-the-counter drugs' and 'time to pregnancy >1 year'). The moderate agreement of the items 'having had surgery' and 'risk for sexually transmitted disease' appeared to be the result of selective recording of one of the caregivers to record the risk only if she/he assumed an independent clear effect on obstetric outcome. We assume these results are representative for practice variation in antenatal history taking.

The differences in domain sum scores were lower than the predefined threshold. The medical domain showed lower agreement compared to the other domains, which was unexpected in view of the assumed standardised history taking. Apparently, inter-professional differences exist in how medical risks are questioned and recorded (Davis et al., 2010; Loosveldt and Beullens, 2013). We expect these differences to be a general phenomenon, in view of existing practice variation e.g. in referral during pregnancy. For comparison, unfortunately no systematic data exist on the current reliability of antenatal history taking. The national perinatal registry indicates that even the recording of the key risk factor 'previous caesarean section' as part of history taking is fairly incomplete (PRN, 2011, The Netherlands Perinatal Register). To improve accuracy, we adapted the script for future use.

In case of divergent risk assessment in the reliability study we could not verify whose assessment was correct, due to our routine practice approach and the associated limitations imposed by the Medical Ethical Board. Now that the R4U has been shown to be feasible and reliable, validity will be tested through the ability to predict SGA, prematurity, and other adverse outcomes (follow up of our study population, data from an on-going national RCT including R4U use) (TSG, 2012, Tijdschrift voor Gezondheidswetenschappen). While we refrained from a validity study at this stage in view of the primary research questions on feasibility and reliability, some indirect evidence on R4U validity was present: the observed difference between score profiles in hospitals versus midwifery practices is in agreement with the expectation. Moreover, the distribution of risk scores appears similar to those reported by a previous birth cohort study in the same region, which introduced the cumulative risk concept and showed its predictive validity for e.g. SGA and prematurity (Timmermans et al., 2011). For example, our population showed a similar slightly right-skewed normal distribution of the sum scores. The normal shape of the distribution of the overall risk score underlines that a dichotomy of being at 'low' or 'high' risk (or 'normal' versus 'pathological' pregnancy) does not reflect nature in early pregnancy. Risk assessment therefore requires the combined expertise of midwife, gynaecologist, and frequently public health workers too (Posthumus et al., 2013).

Overall, both pregnant women as well as professionals reacted positively to the routine use of the R4U score card in perinatal care. The majority of women did not spontaneously express any problem with the items asked, including the sensitive items.

Please cite this article as: van Veen, M.J., et al., Feasibility and reliability of a newly developed antenatal risk score card in routine care. Midwifery (2014), http://dx.doi.org/10.1016/j.midw.2014.08.002

Few of them explicitly expressed relief that the sensitive items did not apply to them. A minority of women expressed a little embarrassment answering the sensitive questions, or asked for explanation why these items were asked, but none of them actually refused to answer. This may be partly due to the fact that reluctance is usually less than expected among those with serious problems, and partly because the Dutch public is informed on the duty of perinatal caregivers to ask for domestic violence in all women. Further qualitative as well as comparative research on the sensitive questions is addressed elsewhere (Quispel et al., 2014). Professionals initially expressed the following negative feedback: reluctance to ask the sensitive items; fear of lengthy history taking; difficulty with specific items (e.g. teenage pregnancy: what age threshold?, stressful job: what means 'stressful'?). Positive feedback, which increased over time included: the feeling of having a complete picture, the conciseness of items of the social (non-medical) domains; the efficient yes/no response mode; and the supportive script. Actually, the introduction of standardised history taking using the R4U also changed care practice. For example, referral and follow-up care pathways were informally standardised as well as a response to standardised history taking; and caregivers' awareness and alertness was enhanced, especially regarding the identification and impact of non-medical risk factors also in the context of delivery.

In view of the favourable results, we improved the R4U for current use. We rephrased the items with high missing rates, low inter-rater agreement, and we removed the option 'not applicable'. Future validity research will verify the independent relevance of items, and the added value of weighted summation of items to improve prediction (Timmermans et al., 2011).

In conclusion our practice-based study demonstrated that systematic retrieval of both clinical and non-clinical data in antenatal care is feasible to the potential advantage of perinatal care in general and in deprived urban areas in particular.

Conflict of interest statement

None of the authors had any conflicts of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.midw.2014.08.002.

References

Agyemang, C., Vrijkotte, T.G., Droomers, M., van der Wal, M.F., Bonsel, G.J., Stronks, K., 2009. The effect of neighbourhood income and deprivation on pregnancy outcomes in Amsterdam, The Netherlands. J. Epidemiol. Community Health 63, 755–760.

Bonsel, G.J., Birnie, E., Denktas, S., Poeran, J., Steegers, E.A., 2010. Patterns in Perinatal Death, Pregnancy and Birth 2010.Erasmus MC/The Netherlands Organisation for Health Research and Development, Rotterdam. Lijnen in de Perinatale Sterfte 2010, Signalementstudie Zwangerschap en Geboorte (in Dutch).

Davis, R.E., Couper, M.P., Janz, N.K., Caldwell, C.H., Resnicow, K., 2010. Interviewer effects in public health surveys. Health Educ. Res. 25, 14–26.

de Graaf, J.P., Ravelli, A.C., de Haan, M.A., Steegers, E.A., Bonsel, G.J., 2013. Living in deprived urban districts increases perinatal health inequalities. J. Matern.-Fetal Neonatal Med. 26, 473–481.

Denktas, S., Bonsel, G.J., Van der Weg, E.J., et al., 2012. An urban perinatal health programme of strategies to improve perinatal health. Matern. Child Health J. 16, 1553–1558.

EURO-PERISTAT Project in Collaboration with SCPE, E.a.E., 2010. European Perinatal Health Report. Better Statistics for Better Health for Pregnant Women and their Babies in 2010. Available at: ⟨http://www.europeristat.com/reports/european-perinatal-health-report- 2010.html⟩ (accessed 8 January 2013).

Goedhart, G., van Eijsden, M., van der Wal, M.F., Bonsel, G.J., 2008. Ethnic differences in preterm birth and its subtypes: the effect of a cumulative risk profile. Br. J. Obstet. Gynaecol. 115, 710–719.

Jaddoe, V.W., van Duijn, C.M., Franco, O.H., et al., 2012. The Generation R Study: design and cohort update 2012. Eur. J. Epidemiol. 27, 739–756.

Kleijer, M.E., Dekker, G.A., Heard, A.R., 2005. Risk factors for intrauterine growth restriction in a socio-economically disadvantaged region. J. Matern.-Fetal Neonatal Med. 18, 23–30.

Loosveldt, G., Beullens, K., 2013. The impact of respondents and interviewers on interview speed in face-to-face interviews. Soc. Sci. Res. 42, 1422–1430.

Poeran, J., Maas, A.F., Birnie, E., Denktas, S., Steegers, E.A., Bonsel, G.J., 2013. Social deprivation and adverse perinatal outcomes among Western and non-Western pregnant women in a Dutch urban population. Soc. Sci. Med. 83, 42–49.

Posthumus, A.G., Scholmerich, V.L., Waelput, A.J., et al., 2013. Bridging between professionals in perinatal care: towards shared care in the Netherlands. Matern. Child Health J. 17, 1981–1989.

PRN, 2011. Perinatal Care in the Nederlands, Annual Reports. Stichting Perinatale Registratie Nederland. Available at: ⟨http://www.perinatreg.nl/uploads/150/137/PRN_jaarboek_2011_23122013_PDF.pdf⟩ (accessed 20 December 2013).

Quispel, C., van Veen, M.J., Zuijderhoudt, C., et al., 2014. Patient versus professional based psychosocial risk factor screening for adverse pregnancy outcomes. Matern. Child Health J..

Timmermans, S., Bonsel, G.J., Steegers-Theunissen, R.P., et al., 2011. Individual accumulation of heterogeneous risks explains perinatal inequalities within deprived neighbourhoods. Eur. J. Epidemiol. 26, 165–180.

TSG, 2012. Zorg rond zwangerschap en geboorte. Perinatal care and birth. Tijdschr. Gezondheidswetenschappen 90, 458–460.

van der Kooy, J., Poeran, J., de Graaf, J.P., et al., 2011. Planned home compared with planned hospital births in the Netherlands: intrapartum and early neonatal death in low-risk pregnancies. Obstet. Gynecol. 118, 1037–1046.

Vrijkotte, T.G., van der Wal, M.F., van Eijsden, M., Bonsel, G.J., 2009. First-trimester working conditions and birthweight: a prospective cohort study. Am. J. Public Health 99, 1409–1416.

WIC, 2013. Women, Infant and Children Program. Available at: ⟨http://www.fns.usda.gov/wic⟩ (accessed 20 December 2013, accessed 8 January 2013).

Please cite this article as: van Veen, M.J., et al., Feasibility and reliability of a newly developed antenatal risk score card in routine care. Midwifery (2014), http://dx.doi.org/10.1016/j.midw.2014.08.002