Using Rough Sets to study expert behaviour in induction of labour

D.T. Parry*, W.K. Yeap* and N. Pattison**
***Artificial Intelligence Laboratory, University of Otago, New Zealand**
****Department of Obstetrics and Gynaecology, University of Auckland, New Zealand**
dparry@infoscience.otago.ac.nz, yeap_wk@cs.otago.ac.nz, n.pattison@auckland.otago.ac.nz

Abstract

The rate of induction of labour (IOL) is increasing, despite no obvious increase in the incidence of the major indications. However the rate varies widely between different centres and practitioners and this does not seem to be due to variations in patient populations. The IOL decision-making process of six clinicians was recorded and examined using hypothetical scenarios presented on a computer. Several rules were identified from a rough sets analysis of the data. These rules were compared to the actual practise of these clinicians in 1994 Initial tests of these rules show that they may form a suitable set for developing an expert system for the induction of labour.
Keywords: Rough Sets, Obstetrics, Knowledge Acquisition

1 Introduction

Induction of Labour (IOL) is a medical procedure using drugs and minor surgery that causes a pregnant woman to give birth within the next few hours. It is used when there are reasons (indications) that the pregnancy should not continue for much longer. Still, to intervene or not is a question obstetricians are increasingly finding it difficult to answer. Thus, although they have an overriding duty to each mother and baby to take all reasonable steps to avoid tragedy, unnecessary intervention is costly, inconvenient and possibly dangerous. However, the risks of adverse outcome without interventions have reduced in recent years but at the same time, the methods of intervention are becoming safer and easier. Should one recommend intervention or not?

Studies of hospitals in New Zealand [3], and Finland [2], have noted the very large variations in the rate of IOL between hospitals and between clinicians. In many areas the rate of IOL is increasing and it is widely believed that the present rates are excessive[4]. Unfortunately, definitive guidelines to good practice are lacking and this leads to variations in policy within and between institutions. Attempts have been made to compare the decisions of different clinicians by using a 'standard primipara' [1]. However this approach, although valuable in showing differences between institutions, has problems dealing with different populations and does not reveal the reason for these differences. In this project, we investigated how clinicians make the decisions (see section 2.0), how rules are derived from analysing responses to imaginary examples using rough sets (see section 3.0) and how such rules could then be applied to a database of real examples (see section 4.0). Section 5.0 provides the results and section 6.0 some discussion and proposals for future work.

2 Decision making in IOL

2.1 Source of the data

The AMSIS (Auckland Maternity Service Information System) database has been in use at the National Women's Hospital (NWH) Auckland since the early 1990's. A large amount of data for each birth was entered into the database and this includes information on gestation at delivery, reason for IOL and most other relevant clinical information relating to pregnancy and its complications. The rate of IOL has been increasing at NWH for the 10 years previous to the year that we had study data for.

Reason For IOL	Percentage of Mothers
Pre-Eclampsia (GPH)	5.98%
Post Dates	6.02%
IUGR (Small baby)	3.15%
Antibodies	0.16%
Diabetes	0.59%
Haemmorage	0.23%
Unstable Lie	0.10%
Death of Fetus	0.42%
Spont. Rupture of Membranes	2.36%
Reduced Movements	0.24%
Abnormal Heart Rate	0.11%
Abnormal Ultrasound	0.05%
Decreased Liqour	0.38%
Maternal Distress	0.50%
Fetal Abnormality	1.43%
Other	1.20%
Not Induced	77.06%

Table 1: Reasons for inducing labour

We selected all the births at NWH in 1994 (8800 cases) and from it, extracted the reasons given for IOL (Table 1).

2.2 Clinical Explanation

The majority of these indications account for a very small number of inductions. In addition, apart from the first three indications (Pre-Eclampsia (GPH), Post-Dates, IUGR), the criteria for induction are obvious, and the decisions made quickly . However or Pre-Eclampsia Post-Dates and IUGR there is thought to be a wide variation in opinion, so we decided to concentrate on these. Pre-eclampsia also known as gestational proteinurea and hypertension (GPH) is a disease where the mother begins to have high blood pressure and protein in her urine. If left untreated it can lead to eclampsia where the mother will start to have fits and it may eventually lead to death of mother and baby. The cause is uncertain and the treatment is basically to try and reduce the symptoms and get the baby born. We conclude that blood pressure, proteinurea (amount of protein in the urine) and gestation (number of weeks of pregnancy) are the relevant information for the clinicians. Post dates refers to pregnancy that takes longer than normal (40 weeks). Over 42 weeks, there is a strong consensus that the baby should be born to avoid complications. Between 40 and 42 weeks there is a great deal of debate and controversy about the correct time for induction. Gestation is the important indicator here. Intra Uterine Growth Retardation (IUGR) means that the baby's growth is slower than it should be inside the womb. It has many possible causes but most of them involve a reduced blood flow to the baby and immediate delivery is the best solution (since the baby is not getting bigger, there is no advantage to it being left in the womb). Fetal growth, biophysical profile and gestation are the important information for monitoring this condition.

2.3 Relevant Data

From the above analysis, we concluded that the most important information when deciding to induce or not is information on: blood pressure, proteinurea, gestation, Fetal growth, and biophysical profile. The induction's carried out for indications other than Pre-Eclampsia (GPH), Post-Dates, and IUGR (7.79% of the 1994 patients) are ignored in this study. Table 2 shows the units of these parameters and the range of possible values for them.

3 Method of Investigation

To investigate the IOL decision-making process by clinicians, experts in the field were asked to create 20 difficult or borderline cases i.e. those where the decision is not clear-cut and in which either decision would be within the range of acceptable clinical practice. All these cases have at least one abnormality present i.e. one of the five key parameters identified in Table 2 will be outside its normal range. We constructed a model of the IOL decision making process (Fig. 1) to assist us in the design of the scenarios.

Information Required	**Units**	**Range (best to worst)**
Blood pressure	mmHg systolic/diastolic	70/50 to 220/150, 110/70 to 130/90 normal.
Proteinurea	dipstick units	nil to ++++.
Gestation	weeks	normal (38-41) below 38, less is worse, above 41 the greater number of weeks is worse.
Fetal growth	empirical	Good, progressive but on fifth percentile, no growth for two weeks.
Biophysical profile	empirical	8/8 CTG reactive to 0/8.

Table II: Information identified as relevant to IOL decision-making

is inadequate for them to make a decision. Other information (such as name of patient and non-significant past history) could also be accessed to provide a more realistic feel to the experiment but they were not relevant to this study and were not included in the analysis. Six clinicians took part in the experiment and each was given 15 out of the 20 cases at random. They were encouraged to make a decision for all cases within seven minutes. The computer recorded the information they revealed and their decision. Each subject was tested with an experimenter sitting next to him/her, and each was allowed a practice run with dummy data that was not recorded.

4 Analysis

4.1 Assumptions

In Table 2 we have indicated how the values of the parameters relate to the clinical state of the Mother-to-be. For example, a woman with a blood pressure of 140/100 is in a worse state than one with a blood pressure of 120/80. We postulated that if a clinician was prepared to induce a patient with one set of parameter values in our test scenario then he or she would induce all women with the same or worse values (i.e. >=). Similarly if the decision is not to induce then we would expect that the subject would not induce all women with the same or better values (i.e. ≤). However, it is possible that the clinician has inadvertently

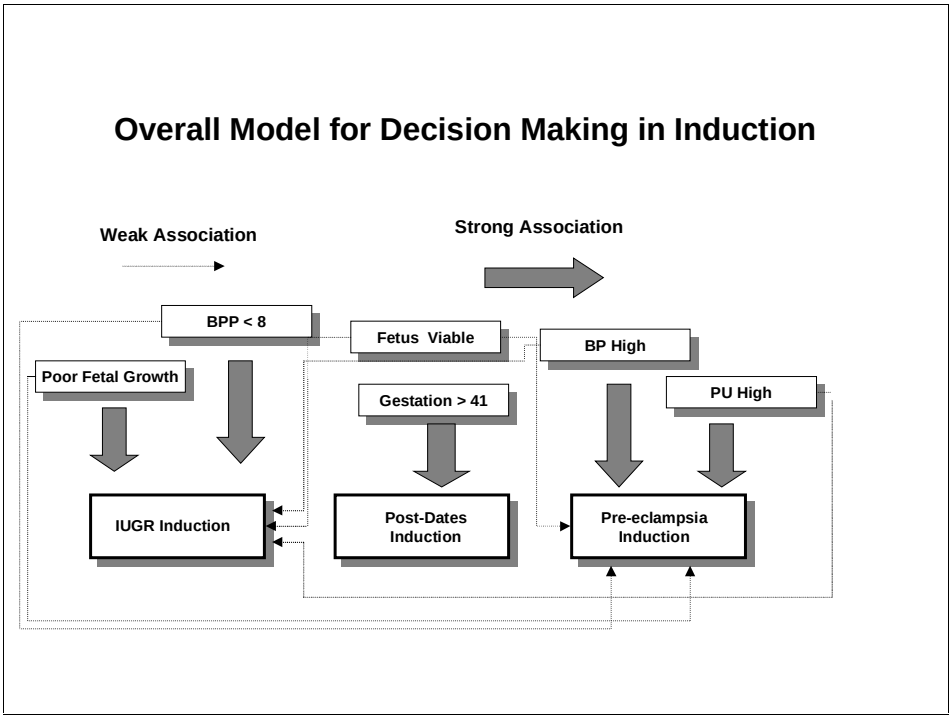

Figure 1 The model used for constructing Scenarios

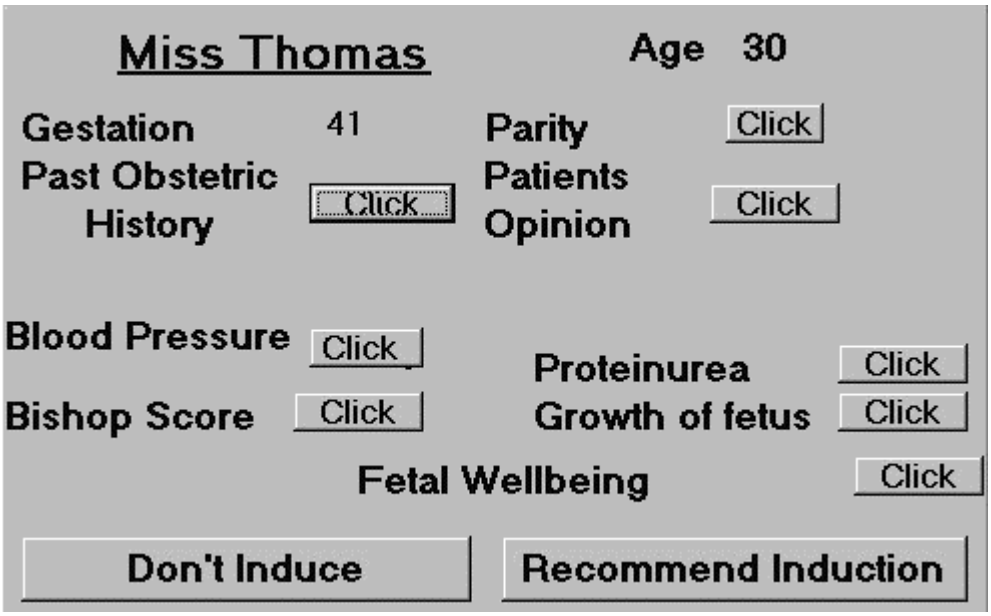

Figure 2: The User Screen

revealed data that is not altering their decision - for example the blood pressure may be revealed but although high (i.e. 'worse' than normal) it is the lack of fetal growth and the gestation that impels the clinician to induce. We needed some way to correct for this possibility.

4.2 Rough Sets

The Rough sets technique was developed in the early 1980's [5] and has been used for a large number of machine learning applications that involve knowledge discovery from databases [6]. In particular, rough sets have already been used in the field of obstetrics to identify pregnancies that may end prematurely [8].

Rough sets theory is based around the analysis of a decision table, which contains a set of attribute -value pairs, and a conclusion. The technique is well described in [6], and involves creating a reduct of the possible rules from the rules that could be derived from a particular decision table, by removing rule that are indistinguishable from others in their effect. A number of techniques have been used for rule creation using the rough sets technique (for example [8]).

Our technique involves constructing a set of rules from a decision table (see section 4.3) and calculating the minimum and maximum rate in which these rules are satisfied (see section 5). The minimum rate is obtained by including all the cases which are actually induced and which the rules will also suggest induction. The maximum rate is obtained by

Blood Pressure	Proteinurea	BPP	Fetal Growth	Gestation	Decision
Not Revealed	Not Revealed	Not Revealed	No Growth	36	Induce
140/95	++	Not Revealed	Not Revealed	37	Induce
120/85	Not Revealed	8/8	On 5th percentile	35	Don't Induce

Table III: Example of a Decision Table

The Induce rules generated are:
a) Induce IF Gestation Worse than or equal to 36
b) Induce IF Fetal Growth Worse than or equal to No Growth
c) Induce IF Gestation Worse than or equal to 36 AND Fetal Growth Worse than or equal to No Growth
d) Induce IF Blood Pressure Worse than or equal to 140/95
e) Induce IF Proteinurea worse than or equal to ++
f) Induce IF Gestation Worse than or equal to 37
g) Induce IF Blood Pressure worse than or equal to 140/95 AND Proteinurea worse than or equal to ++.
h) Induce IF Blood Pressure Worse than or equal to 140/95 AND Gestation worse than 37
i) Induce IF Proteinurea worse than or equal to ++ AND Gestation Worse than 37
j) Induce IF Blood Pressure Worse than or equal to 140/95 AND Proteinurea worse than or equal to ++ AND Gestation Worse than 37

The don't induce rules:
k) Don't Induce IF Blood Pressure Better than or equal to 120/80
l) Don't Induce IF Proteinurea Better than or equal to ++
m) Don't Induce IF Gestation Better than or equal to 35
n) Don't Induce IF Blood Pressure better than or equal to 120/80 AND Proteinurea Better than or equal to ++
o) Don't Induce IF Blood Pressure Better than or equal to 120/80 AND Gestation Better than or equal to 35
p) Don't Induce IF Proteinurea Better than or equal to ++ AND Gestation Better than or equal to 35
q) Don't Induce IF Proteinurea Better than or equal to ++ And Gestation Better than or equal to 35 AND Blood Pressure Better than or equal to 120/80.

Figure 3: Example of the Development of rules

including all the cases which are actually induced or which the rules will suggest induction. This gives us an estimation of the range of possible values of the outcome. A good set of rules is one that will produce a range, which is comparable with the actual rate (i.e. the rate of the clinicians themselves).

4.3 Rule Production

Table 3 shows some of the data obtained from the experiment. From it, the program obtained two decision tables for each clinician, one consisting of those with an 'induce' response and the other with a 'Don't induce' response (these tables are not shown). Rather than attempt to create a reduct from indistinguishable rules at this point we generated all the possible rules from the decision table and then removed those that contradicted each other from the 'induce' and 'don't induce' set. By this means we hoped to prevent the production of spurious rules, while still producing a reasonable number of rules from a small decision table.

From each 'induce' table we generated a set of rules of the form:

'Induce if <parameter name> is worse than or equal to <parameter value>'

For each revealed parameter and each possible conjunction of them. We then did the same for 'Don't induce' table but in this case the rules are of the form:

'Don't induce if a <parameter name> is better than or equal to <parameter value>'

The rules obtained from table III are shown in Figure 3. The analysis uses the 'Induce' set but the 'Don't induce' set is examined to see if any of the rules contradict the rules in the 'Induce' set and these contradicted rules are removed from the final rule - set. This is done by removing all 'Induce' rules where there is a 'Don't Induce' rule with the same parameters that have equal or worse values than the original rule. In the example given, rule (m) conflicts with both rule (a) and (f) because 35 weeks is worse than 36 or 37, (l) conflicts with (e) because both have Proteinurea ++, (p) conflicts with (i) because Proteinurea ++ is the same for both and 37 weeks is better than 35. After removing these rules, we are left with rules (b), (c), (d), (g), (h) and (j). The extent to which rule reduction took place for each subject is shown in table IV.

4.4 Rule Creation

An example of the rules produced for one clinician is shown in appendix 1. We tested a total of six clinicians and the number of rules produced is shown in table IV.

4.5 Rule Application

We wished to compare the 'actual' IOL Rate of each subject for the indications we studied (IUGR, Pre-eclampsia, and Post-Dates) with the rate that would have occurred if the

Subject	Number of "Induce" rules			Number of "Don't Induce" Rules
	Before Reduction	After Reduction	Percentage reduction	
1	62	52	16.13%	162
2	8	8	0.00%	156
3	92	69	25.00%	77
4	97	87	10.31%	108
5	37	32	13.51%	52
6	140	112	20.00%	64

Table IV: Number of Rules generated and the Rule reduction

Subject	Maximum Rate of IOL	Minimum Rate of IOL	Actual IOL Rate 1994
1	31.34%	2.34%	23.3%
2	43.12%	6.01%	13.3%
3	29.40%	3.57%	17.3%
4	36.92%	2.96%	19.4%
5	34.62%	4.65%	12.5%
6	33.43%	2.52%	21.0%

Table V: Results

rules that we had generated had been used. Each of the subjects had been responsible for the management of a number of patients (48-434 mean 205 total 1231) at NWH in 1994, and this, along with the induction decision for each patient, is recorded on the AMSIS Database. We extracted the relevant parameters for each of these patients under their care. Those patients induced for reasons other than our indications were removed from the patient database. We then calculated the 'actual' rate of induction for the relevant indications in 1994 for each subject. The rules obtained above are then applied to the relevant patient database.

5 Results

The "Induce" rules were applied to the 1994 dataset and the results are shown in table V. All the subjects found the experiment interesting and all agreed that the scenarios presented were realistic and most of them were not trivially easy to decide upon. The time limit of seven minutes was generally found to be too short and all the subjects were allowed longer to complete the set, although this time only extended to 20 mins or so.

The subjects in this study all had a higher rate of IOL than the general rate in NWH. This would be expected because all the clinicians were specialists who deal with patients who are more likely to have clinical problems than the general population. None of the subjects had an actual induction rate in 1994 outside the range of the predicted rate, where all the rates refer to patients induced for the three major indications. However the predicted ranges are much wider than would be acceptable for use in an expert system

6 Discussion and Future Work

6.1 Discussion

Knowledge acquisition is often described as a bottleneck in the development of expert systems. It is especially difficult where there is genuine disagreement between experts in the field and also complex and unstated relationships between the variables that are used in the decision. The field of IOL is particularly difficult to study because of the wide variation in severity of presentation, and the differing population groups that individual clinicians serve. A scenario-based system allows the clinician to make decisions in a way that is similar to their normal practice, rather than having to declare their knowledge in a knowledge engineering sense. We believe that this technique allows normally undeclared rules to be discovered. An added advantage here is that every expert has the opportunity to make decisions on the same patient group, so that there is no bias due to differing populations.

The rough sets technique is often used for knowledge discovery from databases but it can be used in any situation where a decision table can be constructed. It has an advantage in that the decision table can produce a set of comprehensible rules. The fact that this technique produces a lower and upper approximation of the true value allows a degree of uncertainty to be represented, which was relatively large in this case.

The rules that are generated by clinicians could be applied to a database of 'standard' mothers, or one that reflects their patient populations to obtain a standard rate of IOL. A combination of these rules may be of use in drawing up guidelines for the use of IOL, and for the development of an expert system.

6.2 Future Work

We are currently investigating other means of deriving a set of rules from the data obtained using the rough sets method. We are especially interested in using domain knowledge to study those rules which comply with the textbook view of the physiology and pathology of labour, and those which do not. We also wish to apply a number of other machine learning and rule extraction techniques to a training set of the AMSIS database for 1995 and compare the result with the rules produced by the scenario system. The rules produced by all these methods can then be used more objectively to compare the induction rates between clinicians at NWH.

References

[1] Cleary, R., Beard, R.W., Chapple, J., Coles, J., Griffen, M., Joffe, M., and Welch, A. "The standard primipara as a basis for inter-unit comparisons of maternity care". *British Journal of Obstetrics and Gynaecology* **103**:223-229, 1996.

[2] Jarvelin, M.R., Hartikainen Sorri, A.L., and Rantakallio, P. "Labour induction policy in hospitals of different levels of specialisation." *British Journal of Obstetrics and Gynaecology.* **100**:310-315, 1993.

[3] Johnson, N. and Ansell, D. "Variation in Caesarean and Instrumental Delivery Rates in New Zealand Hospitals." *Australian and New Zealand. Journal of Obstetrics and Gynaecology.* **35**(1):6-11, 1995.

[4] O'Connor, R.A." Induction of labour - not how but why?" *British Journal of Hospital Medicine* **52**(11):559-563, 1994.

[5] Pawlak, Z. "Rough Sets." *International Journal of Computing and Information Science.* **11**:341-356, 1982.

[6] Pawlak, Z., Grzymala-Busse, J., Slowinski, R., and Ziarko, W. "Rough Sets." *Communications of the ACM* **38**(11):89-95, 1995.

[7] Ziarko, W. and Shan, N. "A Method for Computing All Maximally General Rules In Attribute-Value Systems." *Computational Intelligence* **12**:223-234, 1996.

[8] Woolery, L.K. and Grzymala-Busse, J." Machine learning for development of an expert system to predict premature birth." *Biomedical Sciences Instrumentation.* **31**:29-34, 1995.

Appendix 1

Rules for induction derived for one of the subjects

Key

DBP= Diastolic Blood Pressure (mmHg)
BPP=Biophysical Profile
PU= Proteinurea
FG= Fetal Growth
Gest = Gestation (weeks)

1. Induce if DBP>=60and BPP worse than or equal to 6/8 and PU worse than or equal to Nil and FG worse than or equal to on fifth centile.
2. Induce if PU worse than or equal to Nil and FG worse than or equal to on fifth centile and Gest >=42
3. Induce if DBP>=60and BPP worse than or equal to 6/8 and FG worse than or equal to on fifth centile and Gest >=42
4. Induce if BPP worse than or equal to 6/8 and FG worse than or equal to on fifth centile and Gest >=42
5. Induce if DBP>=60and FG worse than or equal to on fifth centile and Gest >=42
6. Induce if FG worse than or equal to on fifth centile and Gest >=42
7. Induce if DBP>=60and BPP worse than or equal to 6/8 and PU worse than or equal to Nil and Gest >=42
8. Induce if BPP worse than or equal to 6/8 and PU worse than or equal to Nil and Gest >=42
9. Induce if DBP>=65
10. Induce if DBP>=60and BPP worse than or equal to 6/8
11. Induce if PU worse than or equal to Nil and Gest >=42
12. Induce if DBP>=65and PU worse than or equal to +and FG worse than or equal to on fifth centile and Gest >=40
13. Induce if BPP worse than or equal to 6/8 and Gest >=42
14. Induce if DBP>=60and PU worse than or equal to Nil and FG worse than or equal to on fifth centile and Gest >=42
15. Induce if DBP>=60and PU worse than or equal to Nil and Gest >=42
16. Induce if DBP>=65and BPP worse than or equal to 8/8 and Gest >=40
17. Induce if DBP>=65and BPP worse than or equal to 8/8 and PU worse than or equal to +and FG worse than or equal to on fifth centile and Gest >=40
18. Induce if DBP>=65and PU worse than or equal to +
19. Induce if DBP>=60and Gest >=42
20. Induce if PU worse than or equal to Nil and FG worse than or equal to on fifth centile
21. Induce if DBP>=60and BPP worse than or equal to 6/8 and PU worse than or equal to Nil
22. Induce if DBP>=60and FG worse than or equal to on fifth centile
23. Induce if DBP>=60and BPP worse than or equal to 6/8 and FG worse than or equal to on fifth centile
24. Induce if DBP>=60and PU worse than or equal to Nil and FG worse than or equal to on fifth centile
25. Induce if BPP worse than or equal to 6/8 and PU worse than or equal to Nil and FG worse than or equal to on fifth centile
26. Induce if DBP>=65and BPP worse than or equal to 8/8
27. Induce if BPP worse than or equal to 8/8 and FG worse than or equal to on fifth centile and Gest >=40
28. Induce if Gest >=40
29. Induce if DBP>=65and Gest >=40
30. Induce if BPP worse than or equal to 8/8 and Gest >=40
31. Induce if DBP>=60and PU worse than or equal to Nil
32. Induce if PU worse than or equal to +and Gest >=40
33. Induce if DBP>=65and BPP worse than or equal to 8/8 and PU worse than or equal to +and FG worse than or equal to on fifth centile
34. Induce if FG worse than or equal to on fifth centile and Gest >=40
35. Induce if DBP>=65and BPP worse than or equal to 8/8 and PU worse than or equal to +and Gest >=40
36. Induce if DBP>=65and BPP worse than or equal to 8/8 and FG worse than or equal to on fifth centile and Gest >=40
37. Induce if PU worse than or equal to +and FG worse than or equal to on fifth centile and Gest >=40
38. Induce if BPP worse than or equal to 8/8 and PU worse than or equal to +and FG worse than or equal to on fifth centile and Gest >=40
39. Induce if DBP>=60
40. Induce if BPP worse than or equal to 6/8 and PU worse than or equal to Nil and FG worse than or equal to on fifth centile and Gest >=42

41. Induce if BPP worse than or equal to 8/8 and PU worse than or equal to +and Gest >=40
42. Induce if DBP>=65and PU worse than or equal to +and Gest >=40
43. Induce if DBP>=60and BPP worse than or equal to 6/8 and PU worse than or equal to Nil and FG worse than or equal to on fifth centile and Gest >=42
44. Induce if DBP>=65and FG worse than or equal to on fifth centile and Gest >=40
45. Induce if DBP>=60and BPP worse than or equal to 6/8 and Gest >=42
46. Induce if BPP worse than or equal to 8/8 and PU worse than or equal to +and FG worse than or equal to on fifth centile
47. Induce if DBP>=65and BPP worse than or equal to 8/8 and PU worse than or equal to +
48. Induce if DBP>=65and FG worse than or equal to on fifth centile
49. Induce if PU worse than or equal to +and FG worse than or equal to on fifth centile
50. Induce if DBP>=65and BPP worse than or equal to 8/8 and FG worse than or equal to on fifth centile
51. Induce if BPP worse than or equal to 8/8 and FG worse than or equal to on fifth centile
52. Induce if DBP>=65and PU worse than or equal to +and FG worse than or equal to on fifth centile,