

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE MEDICINA SOCIAL

Francisco Barbosa Júnior

Desenvolvimento e teste de um novo indicador
para avaliação da capacidade preditiva de
classificadores de risco

Francisco Barbosa Júnior

Desenvolvimento e teste de um novo indicador para avaliação da capacidade preditiva de classificadores de risco

Versão corrigida. A versão original encontra-se disponível tanto na Biblioteca da Unidade que aloja o Programa, quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD)

Dissertação apresentada ao Programa de Pós-Graduação em Saúde Pública do Departamento de Medicina Social da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo como parte das exigências para a obtenção do título de Mestre em Ciências

Orientador: Prof. Dr. João Paulo Dias de Souza

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Barbosa-Junior, Francisco

Desenvolvimento e teste de um novo indicador para avaliação da capacidade preditiva de classificadores de risco, 2019.

86 p.

Dissertação de Mestrado, apresentada ao Programa de Pós-Graduação em Saúde Pública do Departamento de Medicina Social da Faculdade de Medicina de Ribeirão Preto/USP.

Área de concentração: Ciclos de vida e saúde.

Orientador: Souza, João Paulo Dias de

1. Partograma, 2. ROC, 3. Análise Gráfica, 4. Acurácia diagnóstica

Nome: Francisco Barbosa Júnior

Título: Desenvolvimento e teste de um novo indicador para avaliação da capacidade preditiva de classificadores de risco

Dissertação apresentada ao Programa de Pós-Graduação em Saúde Pública do Departamento de Medicina Social da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo como parte das exigências para a obtenção do título de Mestre em Ciências

Aprovado em: 16 de maio de 2019

Banca examinadora

Prof. Dr. João Paulo Dias de Souza
Faculdade de Medicina de Ribeirão Preto - USP

Prof^a. Dr^a. Gleici da Silva Castro Perdoná
Faculdade de Medicina de Ribeirão Preto - USP

Prof^a. Dr^a. Carolina Sales Vieira Macedo
Faculdade de Medicina de Ribeirão Preto - USP

Dr. Newton Shydeo Brandão Miyoshi
Centro de Informação e Informática em Saúde

*À Dona Ana Maria, minha eterna professora e meu exemplo de dedicação
E ao Seu Barbosa, meu exemplo de caráter e disciplina*

AGRADECIMENTOS

À Deus, que me dá saúde e me acalma nas tempestades.

Aos meus pais, que me proporcionam todo o apoio, amor, carinho e compreensão.

À Livia, que divide comigo não apenas sonhos, mas também o presente.

Aos meus avós, José (*in memorian*) e Maria (*in memorian*), que, lá do céu estão olhando por mim e vendo que o Francisco vingou.

Ao meu avô Calixto (*in memorian*), que me deu uma calculadora científica e uma balança de boi, assim ele não erraria na previsão do meu futuro entre fazendeiro, administrador ou acadêmico.

A minha avó Abigail, com toda sua força e perfeccionismo.

À Cupulla (Ari, Juh, Victor e Lih) que já são parte do meu ser e estar nesses 11 anos de amizade.

Ao professor João Paulo, que me inspira e acredita que eu possa melhorar a cada dia.

Ao professor Fernando, que me ajudou, confiou em meu trabalho e me ofereceu oportunidades para crescer pessoalmente e profissionalmente.

À professora Gleici, que me fez acreditar que a matemática é factível, que sempre acreditou em mim e que todos podem ser melhores.

Ao professor Domingos, que luta pela universidade, pela liberdade de expressão e pelos seus alunos.

Ao professor Afonso, um exemplo de pessoa que trabalha para o bem do próximo.

À Valentine, que faz com que eu queira fazer do mundo um lugar melhor não apenas para hoje, mas também para o futuro.

Aos meus grandes amigos: Kelly, Gabriel e Fran, Gustavo, Alissa, Alexandre, Bam, Lívis e Leandro, Sofia e Victor, Prof. Rui, Leo, Gio, Ingrid e Pe.Carlino, que podem não estar comigo a todo tempo fisicamente, mas em pensamentos sempre estão no meu coração.

Aos meus queridos tios: Chaves, Lourdes, Joaquina, Inês, Linésia, Leonel (*in memorian*), Leonardo (*in memorian*) e Clarice.

Aos financiadores desse projeto: Fundação Bill e Melinda Gates, Organização Mundial de Saúde (OMS) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

À Universidade de São Paulo, que me mudou como pessoa e me faz mais capaz a cada dia.

“Primum non nocere”
Hipócrates Asclepiades

RESUMO

Barbosa-Junior F. Desenvolvimento e teste de um novo indicador para avaliação da capacidade preditiva de classificadores de risco [dissertação]. Ribeirão Preto: Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto, 2019

A constante evolução da tecnologia em saúde permite diagnósticos clínicos cada vez mais rápidos e precisos, porém eles não são isentos de danos. Conhecer e calcular a capacidade preditiva de classificadores de risco é um passo fundamental no desenvolvimento de políticas de saúde e na criação de protocolos clínicos. O processo de avaliação de um teste diagnóstico normalmente tem início com o preenchimento da tabela de contingência. Testes diagnósticos com elevado número de falso positivos geram ansiedade, exames adicionais e tratamentos desnecessários. Testes diagnósticos com elevado número de falso negativos podem contribuir para demoras desnecessárias no tratamento de uma possível doença, podendo dificultar sua cura. A partir da experiência adquirida com a análise da capacidade preditiva do partograma da OMS, esta tese apresenta o desenvolvimento de um novo indicador de capacidade preditiva e diagnóstica de testes. Justifica-se este novo desenvolvimento pela necessidade de simplificar, integralizar e exibir de forma conjunta os principais parâmetros necessários à avaliação da capacidade preditiva e diagnóstica de testes. A avaliação dos partogramas de 9.995 mulheres da Nigéria e Uganda mostrou que a linha de alerta deste instrumento possui uma baixa capacidade diagnóstica como preditor para desfechos adversos graves no trabalho de parto. A partir desses e outros resultados, em fevereiro de 2018 a OMS recomendou mudanças no partograma. Com isso, foi explorado o uso do gráfico-radar como uma ferramenta para facilitar a comunicação dos parâmetros já existentes utilizados na avaliação da capacidade diagnóstica dos testes. Sete documentos com diretrizes em saúde materna publicados pela OMS entre os anos de 2016 a 2018 foram utilizados nesse processo. Utilizando o cálculo da área inscrita do gráfico-radar apresenta-se também um novo indicador integrativo para avaliação da capacidade preditiva de testes diagnósticos, o Índice de Máxima Acurácia (IMA).

Palavras-chave: Partograma. ROC. Análise Gráfica. Acurácia diagnóstica.

ABSTRACT

Barbosa-Junior F. Development and Testing of a new indicator for assessing the predictive capacity of risk classifiers. [dissertation]. Ribeirão Preto: Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto, 2019

The constant evolution of health technology allows for faster and more accurate clinical diagnoses, but they are not harmless. Knowing and calculating the predictive capacity of risk classifiers is a fundamental step in the development of health policies and in the creation of clinical protocols. The process of evaluating a diagnostic test usually begins with the completion of the contingency table. Diagnostic tests with high false positives generate avoidable distress, additional tests and unnecessary treatments. Diagnostic tests with high false negatives may contribute to unnecessary delays in the treatment of a possible disease. Based on the experience gained with the analysis of predictive capacity of the WHO partograph, this thesis presents the development of a new indicator of predictive and diagnostic capacity of tests. This new development is justified by the need to simplify, integrate and display together the main parameters necessary for the evaluation of the predictive and diagnostic capacity of tests. The assessment of the partographs of 9,995 women from Nigeria and Uganda showed that the alert line of this instrument has a low diagnostic capacity as a predictor for serious adverse outcomes in labor. From these and other results, in February 2018 the WHO recommended changes in the partograph. The use of the radar chart was explored as a tool to facilitate the communication of the existing parameters used in the evaluation of the diagnostic capacity of the tests. Seven documents with guidelines on maternal health published by WHO between 2016 and 2018 were used in this process. Using the calculated area of the radar chart, a new integrative indicator for the predictive capacity of diagnostic tests, the Index of Maximum Accuracy (IMA), is also presented.

Keywords: Partograph. ROC. Graphical Analysis. Diagnostic Accuracy.

LISTA DE SIGLAS

ASC	Área sob a curva
BOLD	Better Outcomes in Labour Difficulty
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CSV	Comma Separated Values
DAG	Desfecho adverso grave
DOR	Odds Ratio Diagnóstico
E	Especificidade
FN	Falso Negativo
FP	Falso Positivo
IMA	Índice de Máxima Acurácia
J	Estatística J (Índice de Youden)
ODS	Objetivos do Desenvolvimento Sustentável
OMS	Organização Mundial de Saúde
RedCAP	Research Electronic Data Capture
RMM	Razão de Mortalidade Materna
ROC	Receiver Operating Characteristics
RRL	Radio Research Laboratory
RVP	Razão de Verossimilhança Positiva
RVN	Razão de Verossimilhança Negativa
Se	Sensibilidade
TDS	Teoria de Detecção de Sinal
TP	Trabalho de Parto
USP	Universidade de São Paulo
VP	Verdadeiro Positivo
VN	Verdadeiro Negativo

SUMÁRIO

APRESENTAÇÃO	11
INTRODUÇÃO	12
JUSTIFICATIVA	15
OBJETIVOS	16
<i>Objetivo geral</i>	<i>16</i>
<i>Objetivos específicos</i>	<i>16</i>
PROJETOS	17
PROJETO 1	17
<i>Antecedentes</i>	<i>17</i>
<i>Objetivos</i>	<i>19</i>
<i>Métodos</i>	<i>19</i>
<i>Resultados</i>	<i>28</i>
<i>Discussão</i>	<i>37</i>
<i>Conclusão</i>	<i>38</i>
<i>Referências</i>	<i>39</i>
PROJETO 2	41
<i>Introdução</i>	<i>41</i>
<i>Métodos</i>	<i>42</i>
<i>Resultados</i>	<i>46</i>
<i>Discussão</i>	<i>50</i>
<i>Conclusão</i>	<i>52</i>
<i>Referências</i>	<i>53</i>
<i>Material Suplementar</i>	<i>55</i>
CONSIDERAÇÕES FINAIS	58
REFERÊNCIAS	59
ANEXOS	60
ANEXO 1. PARTOGRAMA	60
ANEXO 2. FORMULÁRIO DE COLETA DE DADOS	61
ANEXO 3. A ESTATÍSTICA J E O ESPAÇO ROC	72
ANEXO 4. DEFINIÇÃO DOS PREDITORES DO PARTOGRAMA	75
ANEXO 5. ARTIGO PUBLICADO	76

Apresentação

Eu me chamo Francisco Barbosa Junior, sou formado em Informática Biomédica pela Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo. Desde a infância sou um entusiasta da informática, mas, sempre a utilizando como um meio, e não como um fim. Somado a isso, as ciências biológicas e os seres vivos sempre me fascinaram pelos seus desenhos, cores e formas. Assim, a informática e suas facilidades – multimídia, internet – me aproximaram ainda mais dos seres vivos. Por fim, o ser vivo que mais me fascina é o ser humano.

Este trabalho, de certa forma, traz a resposta para uma das minhas perguntas de vida: “Como saber que algum tratamento está ou não adequado quando se trata do ser humano?”. Claro que a resposta para essa pergunta é muito mais ampla, individual e difusa do que como eu a trato aqui, mas, em essência: pode ser de grande valia reconhecer e interpretar de uma forma mais completa e visual como uma ação pode influenciar a vida de um indivíduo.

Assim, minha jornada aqui se divide em duas partes: Uma na qual eu mergulho em um grande projeto a fim de entender melhor o processo de trabalho de parto. A partir desse projeto pude conhecer e compreender melhor o ferramental com o qual pode ser possível avaliar testes diagnósticos para descobrir qual “dá a resposta certa”. As diferentes formas de entendimento deste ferramental podem dar margem a muitas interpretações. “E como eu posso contribuir?”. Sendo assim, na segunda parte deste projeto, estudei melhor o ferramental disponível para propor uma nova forma de visualização e quantificação de resultados de testes diagnósticos.

Introdução

O avanço das ciências médicas tem contribuído para a melhora da qualidade de vida das populações. Entretanto, a partir da década de 1950, os efeitos prejudiciais decorrentes das práticas da medicina (iatrogenia) passaram a ser melhor estudados (1, 2). Em que pese o desenvolvimento científico e tecnológico, ao longo das últimas décadas, a iatrogenia passou a ser uma das principais causas de mortalidade evitável em todo mundo (3). Considerando que a maior parte da terapêutica atual inclui a possibilidade de ocorrência de efeitos adversos é fundamental que a indicação dos tratamentos seja realizada da melhor forma possível (2).

Neste contexto, os testes diagnósticos têm um papel chave, não apenas por auxiliarem na indicação de tratamentos clínicos, mas também por favorecerem a realização de outros testes diagnósticos. Levando em conta os possíveis efeitos prejudiciais dos testes diagnósticos, é essencial que estes testes sejam capazes de revelarem a real condição do paciente (4). Testes diagnósticos com elevado número de falso positivos geram ansiedade no paciente, exames adicionais e tratamentos desnecessários. Testes diagnósticos com elevado número de falso negativos podem contribuir para demoras desnecessárias no tratamento de uma possível doença, podendo dificultar sua cura (5). Desta forma, conhecer e interpretar corretamente a capacidade diagnóstica dos testes é um passo fundamental no desenvolvimento de políticas de saúde ou diretrizes clínicas. Essa tarefa, entretanto, pode ser não trivial (6,7).

O processo de avaliação de um teste diagnóstico normalmente tem início com o preenchimento da Tabela de Contingência (8). Também chamada de matriz de confusão, a Tabela de Contingência é utilizada para calcular as métricas que auxiliam na interpretação de resultados de testes. Nas colunas desta tabela tem-se o resultado real das instâncias testadas e nas linhas, a resultado dado pelo teste utilizado. Dada

uma condição-alvo e um teste, cada indivíduo terá quatro possibilidades de classificação: verdadeiro positivo (VP), falso positivo (FP), falso negativo (FN) e verdadeiro negativo (VN). A Figura 1 apresenta a Tabela de Contingência e seus elementos.

		Condição Alvo	
		Presente	Ausente
Resultado do teste	Positivo	VP	FP
	Negativo	FN	VN

VP (Verdadeiro Positivo): Onde a condição alvo é presente e o teste é positivo.
 FP (Falso Positivo): Onde a condição alvo não está presente, mas o teste é positivo.
 FN (Falso Negativo): Onde a condição alvo é presente, porém o teste é negativo.
 VN (Verdadeiro Negativo): Onde a condição alvo não é presente e o teste é negativo.

Figura 1. A Tabela de Contingência

Após a classificação de cada indivíduo da amostra em uma das possibilidades apresentadas, é possível calcular a sensibilidade e a especificidade. A definição e as fórmulas para o cálculo destas métricas são mostradas no Quadro 1. Ainda que a sensibilidade e a especificidade sejam consideradas as principais métricas para avaliação da capacidade de um teste diagnóstico, elas são complementares (9). Analisados individualmente podem ser insuficientes para o entendimento adequado do desempenho de um teste. Com o passar dos anos foram desenvolvidas outras métricas derivadas da tabela de contingência para complementar o entendimento sobre o desempenho dos testes diagnósticos. Razão de verossimilhança positiva, razão de verossimilhança negativa, odds ratio diagnóstico, estatística J, constituem os principais destes parâmetros adicionais. Algumas destas métricas, como Odds Ratio Diagnóstico (DOR) (10) e a Estatística J, também conhecida como Índice de Youden (11), tem o potencial de serem substitutivos das demais métricas por possuírem um caráter integrador da sensibilidade e especificidade. Entretanto, nenhuma destas

costuma ser utilizada de forma isolada e cabe ao indivíduo que está avaliando a capacidade diagnóstica de um teste, interpretar os resultados obtidos com as métricas utilizadas.

Quadro 1. Métricas de avaliação, suas definições e fórmulas

Sensibilidade¹	Detecta quão bem um teste classifica positivamente um indivíduo que está doente $\text{Sensibilidade} = \frac{VP}{VP + FN}$
Especificidade¹	Detecta quão bem um teste classifica negativamente um indivíduo que está sadio. $\text{Especificidade} = \frac{VN}{VN + FP}$
Razão de verossimilhança positiva²	Indica quantas vezes mais chances o paciente doente tem de ter determinada condição do teste comparado com um paciente sadio. $LH+ = \frac{\text{Sensibilidade}}{1 - \text{Especificidade}}$
Razão de Verossimilhança negativa²	Indica quantas vezes menos chances o paciente doente tem de ter determinada condição do teste comparado com um paciente sadio. $LH- = \frac{1 - \text{Sensibilidade}}{\text{Especificidade}}$
Odds Ratio Diagnóstico³	Indicador abrangente e integrador de acurácia diagnóstica do teste. Combina os valores de sensibilidade e especificidade. Tem como limitação principal que um mesmo resultado pode significar alta sensibilidade e baixa especificidade como também baixa sensibilidade e alta especificidade. $ORD = \frac{LH+}{LH-}$
Estatística J⁴	Indicador abrangente e integrador de acurácia diagnóstica do teste, combinando os dados de sensibilidade e especificidade. $J = \text{Sensibilidade} + \text{Especificidade} - 1$
VP (Verdadeiro Positivo): Onde a condição alvo é presente e o teste é positivo. FP (Falso Positivo): Onde a condição alvo não está presente, mas o teste é positivo. FN (Falso Negativo): Onde a condição alvo é presente, porém o teste é negativo. VN (Verdadeiro Negativo): Onde a condição alvo não é presente e o teste é negativo. 1. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. BMJ. 11 de junho de 1994;308(6943):1552. 2. Deeks JJ, Altman DG. Statistics Notes: Diagnostic tests 4: likelihood ratios. BMJ. 17 de julho de 2004;329(7458):168–9. 3. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. novembro de 2003;56(11):1129–35. 4. Youden WJ. Index for rating diagnostic tests. Cancer. Wiley Online Library. 1950. [Internet]. [citado 22 de outubro de 2018]. Disponível em: https://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142%281950%293%3A1%3C32%3A%3AAID-CNCR2820030106%3E3.0.CO%3B2-3	

Justificativa

A adequada avaliação da capacidade preditiva dos testes utilizados para auxiliar a prática médica é essencial para a redução dos possíveis efeitos prejudiciais decorrentes de equívocos de interpretação. Com o intuito de facilitar e sistematizar esta tarefa, uma série de parâmetros são utilizados para realizar a avaliação da capacidade diagnóstica de testes. Estes parâmetros são: sensibilidade, especificidade, razões de verossimilhança, odds ratio diagnóstico e estatística J.

Interpretar e tomar decisões mediante a análise destes parâmetros permanece um desafio para clínicos, desenvolvedores de diretrizes clínicas e de políticas de saúde. Este trabalho apresenta o desenvolvimento de um novo indicador de desempenho na tentativa de simplificar, integralizar e exibir de forma conjunta os principais parâmetros necessários à avaliação da capacidade diagnóstica de testes.

O termo “classificador” é amplamente utilizado em diversas áreas do conhecimento. A classificação pode ser definida como rotular, separar, especificar. No contexto deste trabalho utilizamos a classificação como forma de diferenciar o que não está em equilíbrio, o que não está saudável. Um bom teste diagnóstico é aquele que consegue classificar – na maioria das vezes – se o indivíduo em questão está doente ou sadio. A “capacidade de predição” se refere a habilidade de um determinado instrumento em classificar corretamente um evento antes do seu acontecimento.

Objetivos

Objetivo geral

Facilitar o entendimento e a comunicação da avaliação da capacidade preditiva de classificadores em saúde.

Objetivos específicos

1. Desenvolver expertise na avaliação da capacidade preditiva de classificadores em saúde (Projeto 1);
2. Desenvolver e testar um indicador integrativo da capacidade preditiva de classificadores em saúde (Projeto 2).

Projetos

Projeto 1

Desenvolvimento da expertise na avaliação da capacidade preditiva de classificadores em saúde: o caso do partograma

Antecedentes

Para o desenvolvimento do conhecimento na avaliação da capacidade preditiva de classificadores foi estudada a capacidade preditiva de classificadores presentes em uma ferramenta da Organização Mundial da Saúde denominada “partograma”. O partograma é uma ferramenta utilizada para monitorar a progressão do trabalho de parto. Com base no registro realizado no partograma pode-se realizar uma classificação prospectiva de risco e se derivar intervenções ao longo do curso do trabalho de parto.

Neste trabalho foram avaliados 15 componentes presentes no partograma. A partir do ferramental clássico de avaliação foi possível concluir quais testes eram mais eficientes na predição de desfechos adversos graves relacionados ao trabalho de parto e parto.

Introdução

A mortalidade materna ainda é uma realidade em todo o mundo. Segundo a Organização Mundial de Saúde (OMS), aproximadamente 830 mulheres morrem diariamente de causas relacionadas a gravidez e ao trabalho de parto. Em que pese a queda de 44% da razão de mortalidade materna (RMM) nos últimos 25 anos, os valores continuam expressivos, com 303 mil mortes apenas em 2015 (1). O terceiro Objetivo de Desenvolvimento Sustentável (ODS) (2) possui como uma de suas metas a redução da RMM global para valores menores do que 70 mortes para cada 100 mil nascidos vivos. Além disso, almeja-se que nenhum país poderá ter RMM maior do que duas vezes a média global. Em 2015, ocorreram 239 mortes maternas para cada 100 mil nascidos vivos nos países em desenvolvimento e 12 mortes para cada 100 mil nascidos vivos nos países desenvolvidos. Mundialmente, hipertensão, hemorragia e sepse são os maiores causadores de morte materna (3). Porém, o trabalho de parto e parto são processos naturais em mulheres grávidas saudáveis e possuem baixa frequência de complicações (4,5). Durante o trabalho de parto, monitorar a evolução da parturiente é uma medida utilizada para a identificar precocemente possíveis complicações (6).

O partograma (Anexo 1) é um instrumento para avaliação e classificação de risco da parturiente (7). Sua correta interpretação pode auxiliar à decisão clínica, com o objetivo de maximizar os desfechos favoráveis do trabalho de parto e parto. Neste trabalho é apresentada uma análise do projeto BOLD (*Better Outcomes in Labour Difficulty*) (8). Este projeto da Organização Mundial de Saúde coletou dados de mais de 10 mil mulheres na Nigéria e Uganda. Seu objetivo principal é o estudo e criação de diversas ferramentas para a redução da morbimortalidade no cuidado maternal intraparto.

Curvas de referência e o partograma

As primeiras curvas de trabalho de parto foram desenvolvidas na década de 1950 por Friedman (9). Com base em seu trabalho, na década de 1970, Philpott e Castle desenvolveram uma ferramenta para anotação de dados durante o trabalho de parto chamada partograma (10,11). No partograma é possível anotar a pressão arterial, temperatura corpórea, batimentos cardíacos, número de contrações, entre outros. Seu principal componente é o cervicograma, um quadro onde é anotada a

duração do trabalho de parto no eixo horizontal e o valor de dilatação do colo do útero no eixo vertical. No cervicograma estão as linhas de alerta e de ação, que exercem um papel chave na atual tomada de decisão clínica (12,13). Estas ações podem ser tomadas a partir da identificação de padrões específicos que possam sugerir possíveis riscos à saúde da mãe ou ao bebê. Estão entre as ações mais comuns: intensificação do monitoramento intraparto, transferência para maternidades que atendem casos de alto risco, aceleração do trabalho de parto e cesariana (14).

A progressão normal do trabalho de parto é definida como: a taxa de dilatação do colo do útero igual ou superior a 1cm/h durante a fase ativa do trabalho de parto. Usualmente a fase ativa do trabalho de parto é estabelecida a partir dos quatro centímetros de dilatação do colo do útero (12). A taxa de dilatação de 1cm/h é representada pela linha de alerta. A linha de ação é paralela à linha de alerta, porém com um avanço de quatro horas em relação à primeira. Com o objetivo de reduzir a proporção de mulheres com trabalho de parto prolongado, foi proposto o manejo ativo do trabalho de parto (13). De acordo com esta abordagem, as mulheres com taxa de dilatação menor que 1cm/h deveriam receber intervenções para acelerar o trabalho de parto. Embora o manejo ativo do trabalho de parto possa ser controverso, alguns aspectos desta abordagem são de grande popularidade entre os profissionais que assistem ao parto e podem contribuir para o superuso da ocitocina para acelerar o trabalho de parto. Novos estudos sugerem que a regra do 1cm/h tem suas limitações e tem-se buscado o desenvolvimento de novos padrões como alternativas ao partograma original (15 - 17).

Objetivos

Este trabalho avalia a acurácia diagnóstica da linha de alerta, da linha de ação e de outros parâmetros incluídos no partograma da OMS como preditores de desfechos adversos graves durante o trabalho de parto.

Métodos

O projeto Better Outcomes in Labour Difficulty (BOLD) (8, 18) é um estudo prospectivo de coorte realizado pela Organização Mundial de Saúde (OMS) em 13 maternidades da Nigéria e Uganda. Para serem elegíveis à participação, as maternidades deveriam apresentar: número mínimo de 1.000 partos por ano, cuidado

intraparto fornecido por profissionais qualificados, acesso a cesariana, capacidade para realizar parto vaginal operatório e boas práticas de cuidados intraparto. A maioria dos hospitais selecionados à participação do projeto BOLD participaram anteriormente de um estudo multicêntrico da OMS (19), o que facilitou a dinâmica com os assistentes de pesquisa. É importante ressaltar que o partograma era instrumento de uso padrão em todas as maternidades.

O objetivo principal deste projeto é a redução da morbidade e mortalidade materna e perinatal durante o trabalho de parto e parto. A coleta de dados teve início em dezembro de 2014 e fim em novembro de 2015. Estes dados foram coletados nas maternidades elegíveis, a partir das parturientes que atendiam aos critérios de participação do estudo. Eram elegíveis a participar do estudo: parturientes que deram entrada nas maternidades com fetos vivos durante o primeiro estágio do trabalho de parto e dilatação do colo do útero ≤ 6 cm, independentemente do recebimento de assistência pré-natal. Foram excluídas as mulheres com gravidez múltipla, idade gestacional menor do que 34 semanas e 0 dias, cesariana eletiva e aquelas que estavam incapacitadas de assinar o termo de consentimento. Para a análise final também foram excluídas mulheres que deram à luz a bebês com má formação congênita grave e as que tiveram menos de duas avaliações após o início da fase ativa do trabalho de parto.

Todas as informações e avaliações individuais sobre cada parturiente foram obtidas a partir do formulário de coleta do BOLD (Anexo 2). Não houve padronização na duração do intervalo e na frequência das avaliações. Para avaliação do estado vital do feto no momento da admissão foi utilizado o doppler. As informações necessárias ao preenchimento do formulário de coleta eram provenientes da equipe clínica e dos registros da maternidade. A coleta ocorreu somente durante a internação hospitalar, com início no momento da admissão e término em caso de morte materna, transferência ou alta hospitalar. Não houve nenhum tipo de seguimento pós-alta. No Quadro 2 temos a definição das características do desfecho adverso grave do projeto do BOLD.

Quadro 2. Desfecho adverso grave (Bad Outcome)

O desfecho adverso grave é definido com a ocorrência de um ou mais dos seguintes eventos:	
<i>Com relação ao bebê:</i> - Natimorto - Morte neonatal - Fez uso de anticonvulsivantes - Recebeu ressuscitação cardiopulmonar - Apgar <6 no 5º minuto	<i>Com relação a mãe:</i> - Cesariana por suspeita de ruptura uterina - Ruptura uterina - Morte materna ou disfunção orgânica relacionada a distócia no trabalho de parto

Formulário de coleta de dados

O formulário final de coleta é apresentado com 82 questões (Anexo 2), separadas em seis diferentes seções, sendo elas: admissão hospitalar – parte 1, admissão hospitalar – parte 2, dados intraparto, intervenções intraparto, nascimento e desfecho neonatal, desfecho materno. O protocolo de coleta para cada seção é apresentado abaixo.

A – Admissão hospitalar (Parte 1 - Todas as mulheres)

Nesta seção eram coletados os dados iniciais de cada mulher admitida em uma maternidade participante do estudo. Aqui era verificada a elegibilidade de cada uma (questão 12). Se elegível, a parturiente era convidada a participar do estudo e era obtido o Termo de Consentimento (aprovado previamente pelo comitê de ética de cada país participante do estudo e da OMS) e iniciava-se a seção B.

B – Admissão hospitalar (Parte 2 - Apenas mulheres elegíveis)

Se a parturiente era elegível e consentiu à participação no estudo, a seção B era iniciada. Nesta seção eram coletados dados de caracterização da amostra, entre eles: “Possuí atividade remunerada?”, idade gestacional, condições pré-gravidez (obesidade, hipertensão, diabetes).

C – Dados intraparto

Aqui eram anotados os dados para cada avaliação individual realizada durante o trabalho de parto. Dividido em 18 colunas (C1 a C18), sendo cada coluna para uma avaliação e cada linha para anotação de um parâmetro. Eram coletados, entre outros dados, o número de contrações, duração das contrações, batimento cardíaco fetal e materno, pressão sistólica e diastólica, posição da cabeça do feto, apresentação fetal e estado da bolsa (membrana amniótica).

D – Intervenções intraparto

Nessa seção eram obtidos os dados sobre intervenções médicas realizadas durante o trabalho de parto. Data e horário do início do preparo do colo do útero, indução ao trabalho de parto, justificativa sobre a amniotomia (se executada) e cesariana.

E – Nascimento e desfecho neonatal

Aqui eram coletados: data e hora do nascimento, modo de nascimento (parto vaginal, cesariana), sexo, condições vitais, possíveis morbidades e data da alta hospitalar (ou morte, quando aplicável).

F – Desfecho materno

Os dados maternos com ênfase nas complicações durante o trabalho de parto e parto eram coletados nessa última seção do questionário. Aqui eram anotadas a presença de alguma disfunção orgânica (cardíaca, renal, hepática), hemorragia e qual a condição da alta materna (viva ou morta).

Gerenciamento e limpeza dos dados

Para a coleta e gerenciamento de dados foi utilizada a plataforma *Research Electronic Data Capture* (REDCap) (20), hospedada no sistema computacional em nuvem da USP (21). Criado em 2004 por pesquisadores da Universidade de Vanderbilt, nos EUA, o REDCap está presente em mais de 645.000 projetos em 3275 instituições no mundo todo¹. Projetado para a criação de formulários eletrônicos de

¹ Informação obtida no site www.project-redcap.org em 01/02/2019

coleta de dados em pesquisa, provê recursos para entrada, auditoria, geração de relatórios, importação e exportação de dados.

A entrada de dados no REDCap a partir dos formulários em papel foi realizada em centros estabelecidos na Nigéria e Uganda. Semanalmente, os formulários originais preenchidos na coleta de dados eram enviados aos centros de digitação de cada país. Antes do envio para o centro de digitação, os formulários passavam por conferência visual com os coordenadores locais. Se, já no centro de digitação, algum dado categórico estivesse fora dos padrões estipulados, um aviso era emitido pelo software. Assim, o formulário em questão ficava retido para conferência com a equipe correspondente. Semanalmente eram feitas reuniões com os coordenadores das equipes de cada país junto com a equipe de pesquisa da Faculdade de Medicina da USP de Ribeirão Preto, Brasil.

A coleta de dados do projeto BOLD foi encerrada em novembro de 2015. Após a transcrição das fichas em papel para o sistema eletrônico, iniciou-se o processo de limpeza do banco de dados, que teve a duração de seis meses. Um dos principais processos da limpeza de dados foi a criação e execução de um *script* em R que verificava a progressão positiva dos tempos em cada estado de dilatação. Em resumo, inicialmente o script verificava a primeira anotação temporal presente na Seção C e partir daí o próximo horário sempre deveria ser maior ou igual ao anterior. Na segunda fase o mesmo processo se repetiu para a dilatação. Na terceira fase, os valores eram cruzados para verificação de que não houvessem diferentes dilatações para um mesmo horário, porém com a possibilidade de horários distintos para uma mesma dilatação. Esse processo foi repetido para cada uma das mulheres.

Análise de Dados

Após a limpeza, o banco de dados consolidado foi distribuído para as várias frentes de pesquisa do BOLD no formato .csv (*comma separated values*). Para as análises foram utilizadas a linguagem R (22) e o Microsoft Excel 2014. Inicialmente foram calculadas as frequências simples e proporções, utilizadas para descrever as características da população de estudo. Os dados de progressão do trabalho de parto foram plotados na representação gráfica do partograma. Essa representação se dá em um plano cartesiano onde o eixo X (horizontal) representa o tempo percorrido (valores em horas, com 18 sendo o número máximo considerado) e o eixo Y (vertical) representa a dilatação do colo do útero (4 cm foi considerado o início da fase ativa do

TP, com o tempo a contado a partir dessa dilatação). Como parte da análise, a população do estudo foi estratificada conforme a classificação de Robson (23).

Foram criadas tabelas de contingência para cada um dos 15 classificadores presentes no partograma. A partir das tabelas de contingência foram calculadas as medidas de sensibilidade, especificidade, razões positivas e negativas de verossimilhança, odds ratio diagnóstico e estatística J (24). Com os valores de sensibilidade e especificidade foi possível representar graficamente o desempenho dos classificadores do partograma utilizando o espaço ROC.

O espaço ROC

O espaço ROC (25) compreende um plano cartesiano com dimensão de lado 1 (Figura 2). Seu eixo X representa a especificidade (no formato 1 - Especificidade) e seu eixo Y representa a sensibilidade do teste avaliado. Assim, a partir dos valores de sensibilidade e especificidade é possível plotar um ponto no espaço ROC. Cada ponto no espaço ROC é a representação de um teste e sua avaliação. Portanto, na análise gráfica, o melhor teste possível é aquele no qual o seu ponto mais se aproxima do canto superior esquerdo do espaço.

O ponto extremo (0,1) apresenta 100% de especificidade (onde não há falsos positivos) e 100% de sensibilidade (ou seja, não há falsos negativos). Em contrapartida, o pior teste é aquele no qual seu resultado é aleatório. A aleatoriedade se dá aos testes que possuem 50% de sensibilidade e 50% de especificidade. Estes testes estão situados na linha diagonal ascendente ((0,0),(1,1)) do Espaço ROC. Na Figura 2 o teste A é classificado como perfeito, pois está exatamente na coordenada (0,1) do espaço ROC. Na comparação entre os testes B e C, B é mais sensível que C, porém C é mais específico. Já no teste D tem-se uma classificação aleatória. Com o teste E tem-se um resultado pior do que o aleatório (está abaixo da linha diagonal ascendente) (26).

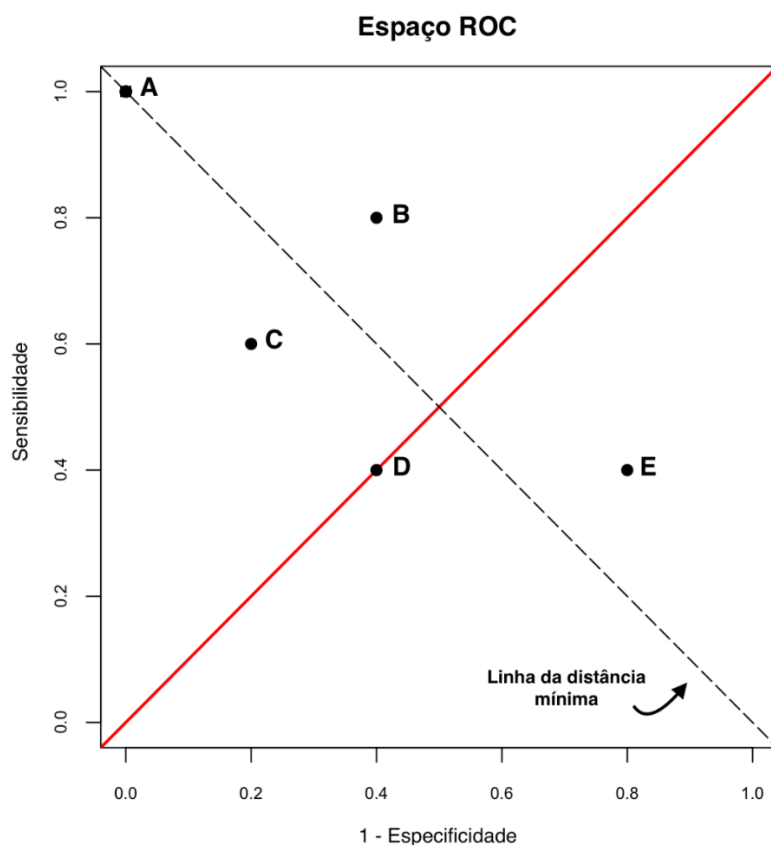


Figura 2. Espaço ROC. Seu eixo X representa a especificidade (no formato de complemento 1-especificidade) e seu eixo Y, a sensibilidade. Cada ponto no espaço ROC representa um resultado de desempenho de um teste.

Tabela 1. Equivalência entre as medidas de desempenho

Sensibilidade	Especificidade	Razão de verossimilhança positiva	Razão de verossimilhança negativa	Odds Ratio diagnóstico	Estatística J	ASCROC
50%	50%	1.00	1.00	1.00	0.00	0.500
60.0%	60.0%	1.50	0.67	2.25	0.20	0.600
70.0%	70.0%	2.33	0.43	5.44	0.40	0.700
80.0%	80.0%	4.00	0.25	16.00	0.60	0.800
90.0%	90.0%	9.00	0.11	81.00	0.80	0.900
90.9%	90.9%	9.99	0.10	99.78	0.82	0.909
95.0%	95.0%	19.00	0.05	361.00	0.90	0.950
99.99%	99.99%	9999.00	0.00	99980001.00	1.00	1.000

ASCROC : Área sob a curva ROC (Receiver Operator Characteristic)

Os testes plotados em um mesmo Espaço ROC podem ser avaliados visualmente ou por meio de manipulações trigonométricas. Temos como exemplo a distância h' , que compreende a distância em linha reta entre o *plot* do teste e a diagonal ascendente (diagonal de não-discriminação) do Espaço ROC. Com a diferença entre as distâncias h (distância do ponto perfeito) e h' obtém-se o valor da estatística J (Anexo 3). Assim como o ORD, a estatística J é um classificador sintético de desempenho. A Tabela 1 apresenta as equivalências entre as medidas de desempenho. A Figura 3 ilustra a curva ROC e o Quadro 3 sumariza suas origens.

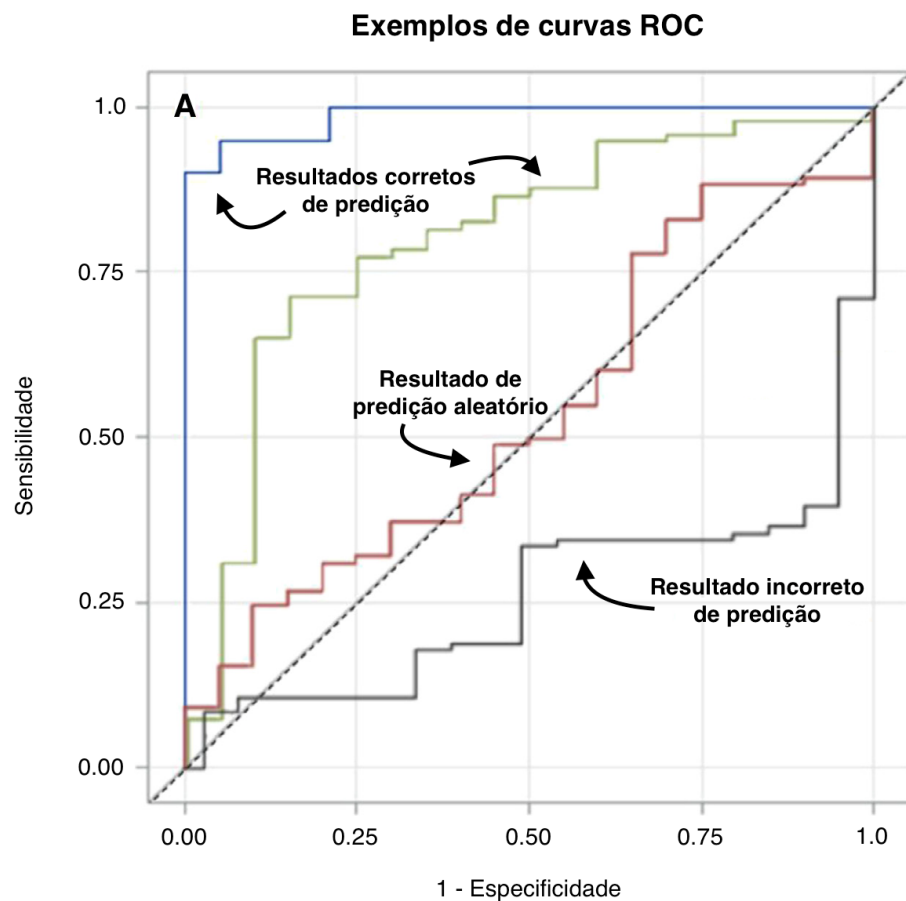


Figura 3. Curvas ROC. Quanto mais próxima a curva estiver do ponto A, maior será o poder de predição do teste avaliado, assim como a sua área sob a curva (ASC). Curvas muito próximas a diagonal de não discriminação sugerem baixa capacidade de predição do teste. Os resultados abaixo da diagonal de não discriminação necessitam de revisão.

Adaptado de: Carter JV, Pan J, Rai SN, Galandiuk S. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery*. 2016 Jun;159(6):1638–45.

Quadro 3. Teoria de detecção de sinal e a curva ROC

Durante a Segunda Guerra Mundial, pesquisadores do Laboratório de Pesquisa em Rádio (RRL - *Radio Research Laboratory*) da Universidade de Harvard investigavam sobre a utilidade prática da detecção de sinais.⁽¹⁾ Em meio a guerra, seria uma enorme vantagem aumentar a capacidade dos operadores de radar em diferenciar corretamente um sinal significativo (um avião) de um ruído (um pássaro). Assim foi desenvolvida a técnica de análise das Características de Operação do Receptor (ROC - *Receiver Operating Characteristics*).^(2,3)

Em seu artigo científico publicado em 1984 na revista *Medical Decision Making*, Lee Lusted descreve a análise ROC, sua história e possíveis aplicações na tomada de decisão clínica.^(1,2,3) Com o passar dos anos, a análise ROC se tornou comum em validações de testes laboratoriais.⁽⁴⁾ Em outubro de 2000, o artigo “Better Decisions Through Science” traz uma reflexão sobre o auxílio matemático na tomada de decisão. Entre outras coisas, esse artigo avalia como essa ferramenta pode acelerar novas descobertas na área da saúde.⁽⁵⁾

1 - Lusted LB. ROC Recollected. *Med Decis Making*. 1984 Jun 1;4(2):131–5.

2 - Lusted LB: Decision making studies in patient management. *N Eng J Med* 284:416-424, 1971 Medline

3 - Lusted LB: Signal detectability and medical decision-making. *Science* 171:1217-1219, 1971 Crossref, Medline

4 - Henderson, A.R., 1993. Assessing test accuracy and its clinical consequences – a primer for receiver operating characteristic curve analysis. *Ann. Clin. Biochem.* 30, 521±539.

5 - Swets, J.A., Dawes, R.M., Monahan, J., 2000. Better decisions through science. *Scientific American* 283, 82–87.

Resultados

Foram triadas 17.845 mulheres. Dessas, 7.612 não cumpriam os critérios de inclusão. As condições mais frequentes para a não inclusão foram: dilatação do colo do útero igual ou maior a sete centímetros no momento da admissão hospitalar (4.212 mulheres, 55,33%), indicação para cesariana de emergência (1.390 mulheres, 18,26%) e idade gestacional menor do que 34 semanas (1.154 mulheres, 15,16%). Os demais motivos são apresentados na Tabela 2.

Tabela 2. Razões para não inclusão

Condições apresentadas	n(%)	
Morte fetal	374	(4,91%)
Dilatação do colo do útero $\geq$ 7cm no momento da admissão	4212	(55,33%)
Gravidez múltipla	386	(5,07%)
Idade gestacional menor do que 34 semanas	1154	(15,16%)
Cesariana eletiva	673	(8,84%)
Indicação para cesariana de emergência	1390	(18,26%)
Menores de idade não emancipadas desacompanhadas	20	(0,26%)
Mulheres que não foram capazes de consentir a participação na pesquisa devido a qualquer problema de saúde	141	(1,85%)
Sem resposta	6	
Não consentiram a participação no estudo	72	(0,95%)
Total	7612	

Após o encerramento da coleta, havia 10.233 mulheres no banco de dados. Após a limpeza inicial, 238 mulheres foram excluídas pelo motivo de dados essenciais incompletos, como por exemplo a falta de informação sobre o Apgar e a falta de informação sobre a situação de saída do recém-nascido da maternidade. Por fim foram avaliadas 9.995 mulheres, com suas características sociodemográficas descritas na Tabela 3. A taxa de desfechos adversos graves foi de 2,2% (223 mulheres). A taxa de cesariana desse estudo foi de 13,2% (1323 mulheres), sendo que 5,4% (535 mulheres) tinham cesariana prévia. Após a limpeza dos dados, 1.038 registros foram excluídos por inconsistências relacionadas ao tempo de trabalho de parto e a aferição do grau de dilatação do colo do útero. Lembrando que as análises que não dependiam do tempo mantiveram 9.995 mulheres. O fluxograma do projeto BOLD é mostrado na Figura 4.

Tabela 3. Características sociodemográficas

Total de mulheres		9995	
País			
	Nigéria	4964	(49,66%)
	Uganda	5031	(50,34%)
Idade (em anos)			
	<20	320	(3,20%)
	20-24	2321	(23,22%)
	25-29	3744	(37,46%)
	30-34	2506	(25,07%)
	35+	1100	(11,01%)
	Sem resposta	4	
	Média	27,9 anos (± 5,0)	
Estado civil			
	Casada / Mora junto	9753	(97,58%)
	Solteira / Separada / Divorciada / Viúva	239	(2,39%)
	Sem resposta	3	
Escolaridade			
	Analfabetas	150	(1,50%)
	Ensino primário	75	(0,75%)
	Ensino Fundamental Incompleto	300	(3,00%)
	Ensino Fundamental Completo	564	(5,64%)
	Ensino Médio Incompleto	1622	(16,23%)
	Ensino Médio Completo	2623	(26,24%)
	Ensino Superior Incompleto	748	(7,48%)
	Ensino Superior Completo	3789	(37,91%)
	Outros	39	(0,39%)
	Sem resposta	85	
Trabalho remunerado		6657	(66,67%)
	Sem resposta	10	

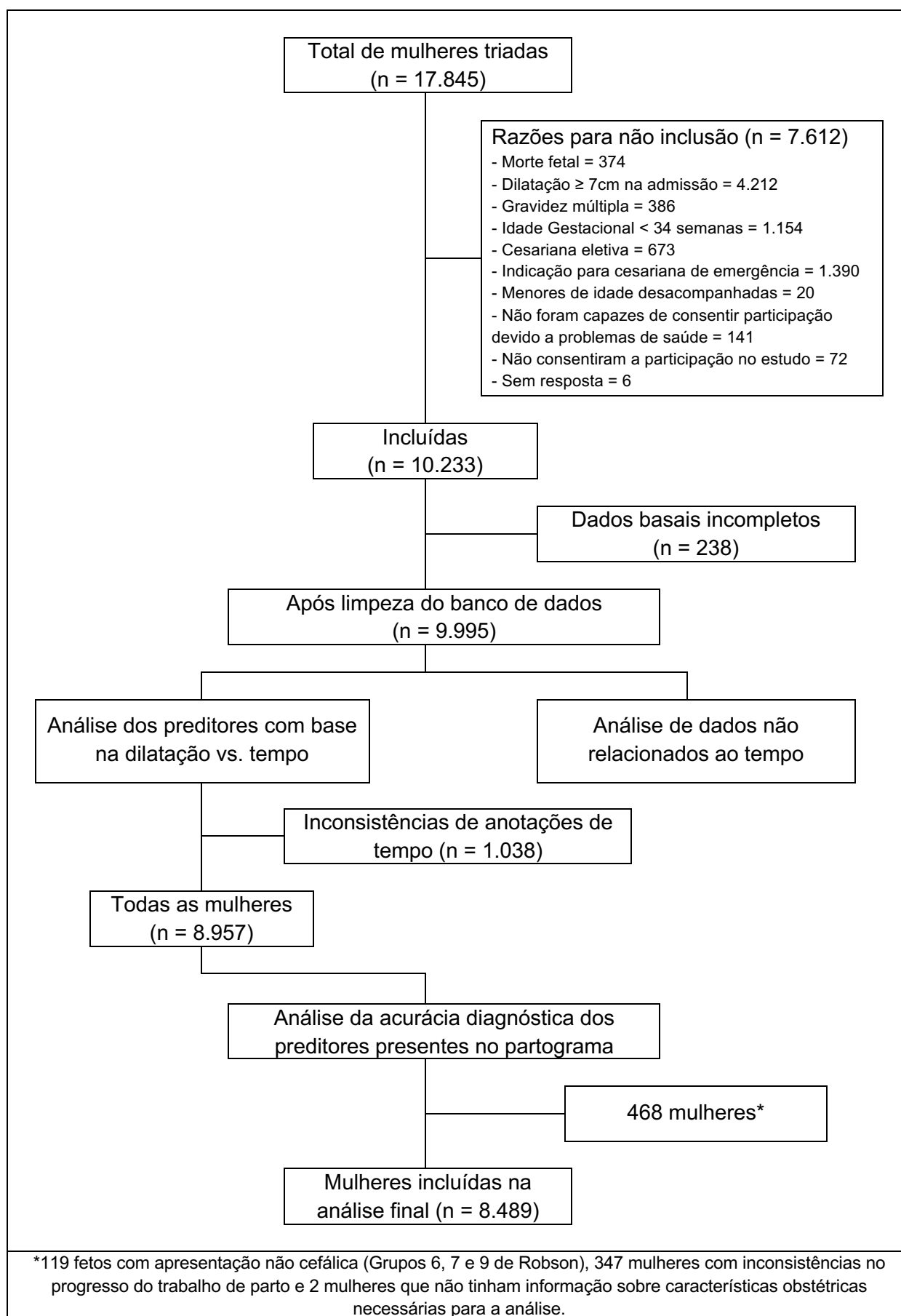


Figura 4. Fluxograma do projeto BOLD

A maioria das mulheres tiveram o trabalho de parto iniciado de forma espontânea e tinham idade gestacional entre 37 e 41 semanas. HIV+ e hipertensão eram as condições de saúde que mais estavam presentes. Na Tabela 4, tem-se Antecedentes Médicos e Reprodutivos.

Tabela 4. Antecedentes Médicos e Reprodutivos

Número de gravidezes			
1-2	5668	(56,98%)	
2+	4280	(43,02%)	
Sem resposta	47		
Número de partos			
0	4076	(40,78%)	
1-2	4241	(42,43%)	
2+	1678	(16,79%)	
Abortos prévios			
0	7568	(75,72%)	
1	2107	(21,08%)	
2+	320	(3,20%)	
Cesarianas prévias			
0	9454	(94,59%)	
1	535	(5,35%)	
2+	6	(0,06%)	
Condições prévias de saúde		n (casos por 1000 gravidezes)	
Hipertensão crônica	68	(6,8)	
Diabetes Mellitus	15	(1,5)	
HIV +	426	(42,6)	
AIDS	10	(1,0)	
Anemia crônica	6	(0,6)	
Obesidade	32	(3,2)	
Doenças cardíacas	5	(0,5)	
Doenças pulmonares	17	(1,7)	
Doenças renais	2	(0,2)	
Outras doenças crônicas	31	(3,1)	

Na Tabela 5 são mostradas as características da gravidez atual. Complicações pré-parto acometeram 1.228 mulheres (12,3%), sendo a Malária (775 mulheres) a maior delas; 667 mulheres (6,7%) não tiveram visitas pré-natal e 5.007 (50,6%) tiveram quatro ou mais visitas.

Tabela 5. Características da gravidez atual

	n (por 1000 gravidezes)	
Complicações		
Placenta prévia	7	(0,7)
Acretismo placentário	0	(0,0)
Descolamento de placenta	2	(0,2)
Outras doenças hemorrágicas	13	(1,3)
Pré-eclampsia (exclui eclampsia)	271	(27,1)
Eclampsia	7	(0,7)
Pielonefrite	9	(0,9)
Malária	775	(77,5)
Ruptura de membranas pré-termo	53	(5,3)
Anemia	18	(1,8)
Diabetes gestacional	6	(0,6)
Outras complicações	149	(14,9)
Pelo menos uma das complicações	1228	(122,9)
Número de consultas Pré-Natal n (%)		
0	667	(6,74%)
1	885	(8,94%)
2-3	3344	(33,77%)
4+	5007	(50,56%)
Sem resposta	92	

A Tabela 6 apresenta a população do estudo de acordo com os Grupos de Robson. O Projeto BOLD só contemplava mulheres com feto único, sendo assim, o grupo 8 (gestação múltipla) não tinha nenhuma parturiente. Fetos com apresentação não cefálica (grupos 6, 7 e 9) foram agrupados e retirados da análise.

Tabela 6. Classificação de acordo com os Grupos de Robson

Grupo	Descrição	n (%)
1	Nulípara, feto único com apresentação cefálica, ≥ 37 semanas, TP espontâneo	2945 (32.89%)
2	Nulípara, feto único com apresentação cefálica, ≥ 37 semanas, TP induzido	385 (4.30%)
3	Múltiparas sem cesariana anterior, feto único com apresentação cefálica, ≥ 37 semanas, TP espontâneo	4160 (46.45%)
4	Múltiparas sem cesariana anterior, feto único com apresentação cefálica, ≥ 37 semanas, TP induzido	379 (4.23%)
5	Múltiparas com pelo menos uma cesariana anterior, feto único com apresentação cefálica, ≥ 37 semanas	434 (4.85%)
6	Nulíparas, feto único com apresentação pélvica	21 (0.23%)
7	Múltiparas, feto único com apresentação pélvica (inclui mulheres com cesariana anterior)	76 (0.85%)
8	Gestação múltipla (inclui mulheres com cesariana anterior)	0 (0.00%)
9	Feto com apresentação transversa (inclui mulheres com cesariana anterior)	22 (0.25%)
10	Feto único com apresentação cefálica, ≤ 36 semanas (inclui mulheres com cesariana anterior)	533 (5.95%)

A representação do partograma com 8.489 mulheres pode ser vista na Figura 5. Cada linha cinza representa o progresso individual de cada uma das mulheres que não tiveram um desfecho adverso grave do parto. Analogamente, cada linha vermelha representa o progresso individual das mulheres que tiveram um desfecho adverso grave. Considerando as mulheres que tiveram pelo menos duas avaliações na fase ativa do trabalho de parto (8.489 mulheres), praticamente metade delas (4.163 mulheres, 49%) cruzaram a linha de alerta.

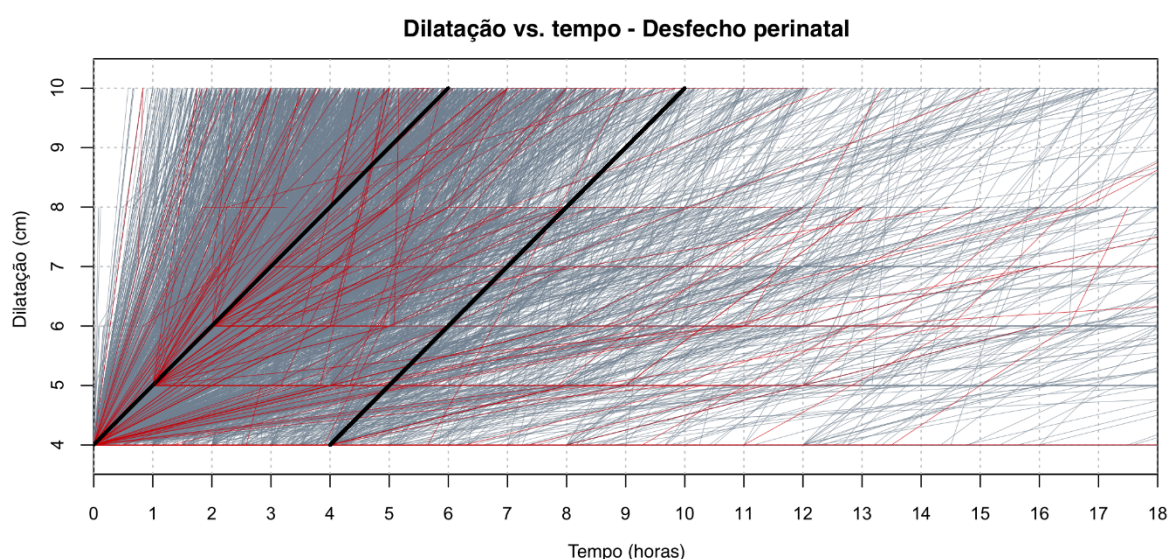


Figura 5. Progressão visual de 8.489 mulheres plotadas no partograma. As linhas de cor cinza correspondem as mulheres sem desfecho adverso grave. As linhas em vermelho correspondem as que tiveram desfecho adverso grave.

A Tabela 7 apresenta a sensibilidade, especificidade, razões de verossimilhança positivas e negativas, odds ratio diagnóstico e estatística J para os principais preditores do partograma: as linhas de alerta e de ação. Na Tabela 8 foi realizada a análise para os outros 13 classificadores presentes no partograma.

Tabela 7. Acurácia diagnóstica da linha de alerta e da linha de ação do partograma para os desfechos adversos graves

Preditor		Desfecho adverso grave		Sensibilidade (95% IC)	Especificidade (95% IC)	Razão de verossimilhança positiva (95% IC)	Razão de verossimilhança negativa (95% IC)	Odds Ratio diagnóstico (95% IC)	Estatística J Índice de Youden (95% IC)
		Presente	Ausente						
Linha de alerta	Cruzou	110	4053	56.7%	51.1%	1.16	0.85	1.37	7.8%
	Não cruzou	84	4242	(49.7-63.5)	(50.1-52.2%)	(1.02-1.32)	(0.72-1.00)	(1.03-1.83)	(0.8-14.9)
Linha de ação	Cruzou	38	1230	19.6%	85.2%	1.32	0.94	1.40	4.8%
	Não cruzou	156	7065	(14.6-25.7)	(84.4-85.9)	(0.99-1.77)	(0.88-1.01)	(0.98-2.01)	(-0.9 – 10.4)

Tabela 8. Acurácia diagnóstica de outros preditores incluídos no partograma para os desfechos adversos graves

Preditor*		Desfechos Adversos Graves		Sensibilidade (95% IC)	Especificidade (95% IC)	Razão de verossimilhança positiva (95% IC)	Razão de verossimilhança negativa (95% IC)	Odds Ratio diagnóstico (95% IC)	Estatística J (95% IC)
		Presente	Ausente						
Mecônio	Presente	62	1220	28.4%	87.1%	2.2	0.82	2.67	15.5%
	Ausente	156	8203	(22.5-34.4)	(86.4-87.7)	(1.77-2.73)	(0.76-0.89)	(1.98-3.61)	(9.5-21.5)
Caput Succedaneum significativo	Presente	31	390	14.2%	95.9%	3.41	0.9	3.81	10.0%
	Ausente	188	9011	(9.5-18.8)	(95.4-96.3)	(2.43-4.79)	(0.85-0.94)	(2.57-5.64)	(5.4-14.6)
Temperatura materna anormal	Presente	24	362	11.9%	95.7%	2.76	0.92	3	7.6%
	Ausente	177	8003	(7.5-16.4)	(95.2-96.1)	(1.87-4.07)	(0.87-0.97)	(1.93-4.65)	(3.1-12.1)

Tabela 8 (cont.). Acurácia diagnóstica de outros preditores incluídos no partograma para os desfechos adversos graves

Preditor*		Desfechos Adversos Graves		Sensibilidade	Especificidade	Razão de verossimilhança positiva	Razão de verossimilhança negativa	Odds Ratio diagnóstico	Estatística J
		Presente	Ausente	(95% IC)	(95% IC)	(95% IC)	(95% IC)	(95% IC)	(95% IC)
Amoldamento significativo	Presente	22	327	10.0% (6.0-14.0)	96.5% (96.2-96.9)	2.88 (1.91-4.33)	0.93 (0.89-0.97)	3.08 (1.96-4.86)	6.5% (2.5-10.5)
	Ausente	198	9075						
Movimentos fetais não observados	Presente	55	1744	25.0% (19.3-30.7)	81.4% (80.6-82.2)	1.35 (1.07-1.70)	0.92 (0.85-0.99)	1.46 (1.07-1.99)	6.4% (0.7-12.2)
	Ausente	165	7649						
Batimento cardíaco fetal anormal	Presente	16	190	7.3% (3.8-10.7)	98.0% (97.7-98.3)	3.63 (2.22-5.93)	0.95 (0.91-0.98)	3.83 (2.26-6.50)	5.3% (1.8 – 8.7)
	Ausente	204	9281						
Descida lenta	Presente	40	1243	22.0% (16.0-28.0)	83.2% (82.4-84.1)	1.31 (0.99-1.73)	0.94 (0.87-1.01)	1.4 (0.98-2.00)	5.2% (-0.9 – 11.3)
	Ausente	142	6176						
Duração sub-ótima das contrações uterinas	Presente	39	1407	18.7% (13.4-23.9)	83.9% (83.1-84.7)	1.16 (0.87-1.55)	0.97 (0.91-1.03)	1.2 (0.84-1.70)	2.6% (-2.8 – 7.9)
	Ausente	170	7341						
Pressão arterial sistólica anormal	Presente	12	318	5.6% (2.5-8.7)	96.4% (96.0-96.8)	1.58 (0.90-2.76)	0.98 (0.95-1.01)	1.61 (0.89-2.92)	2.1% (-1.1 – 5.2)
	Ausente	201	8595						
Posição anormal da cabeça do feto	Presente	4	66	1.9% (0.1-3.8)	99.3% (99.1-99.5)	2.71 (0.99-7.63)	0.99 (0.97-1.00)	2.74 (0.99-7.60)	1.2% (-0.7 – 3.1)
	Ausente	203	9188						
Batimento cardíaco materno anormal	Presente	6	153	2.8% (0.6-5.1)	98.3% (98.0-98.6)	1.66 (0.74-3.72)	0.99 (0.97-1.01)	1.68 (0.73-3.85)	1.1% (-1.1 – 3.4)
	Ausente	206	8835						
Número sub-ótimo de contrações uterinas	Presente	32	1269	15.3% (10.4-20.2)	85.5% (84.8-86.2)	1.06 (0.76-1.46)	0.99 (0.93-1.05)	1.07 (0.73-1.56)	0.8% (-4.1 – 5.7%)
	Ausente	177	7479						
Pressão arterial diastólica anormal	Presente	2	110	0.9% (0.0-2.2)	98.8% (98.5-99.0)	0.76 (0.189-3.06)	1 (0.99-1.016)	0.76 (0.19-3.09)	-0.3% (-1.6 – 1.0)
	Ausente	211	8809						

*O Anexo 4 apresenta as definições utilizadas para cada preditor.

Como principal análise deste trabalho foi utilizado o espaço ROC para visualizar os diferentes preditores presentes no partograma. Na Figura 6 tem-se a análise das linhas de alerta e de ação e na Figura 7, outros 13 preditores presentes no partograma.

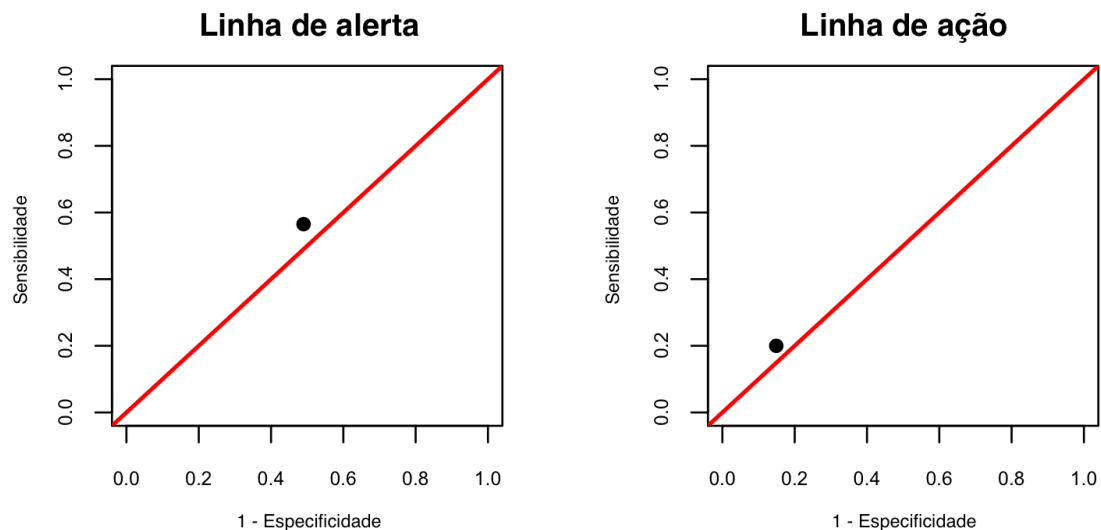


Figura 6. Análise das linhas de alerta e de ação utilizando o Espaço ROC

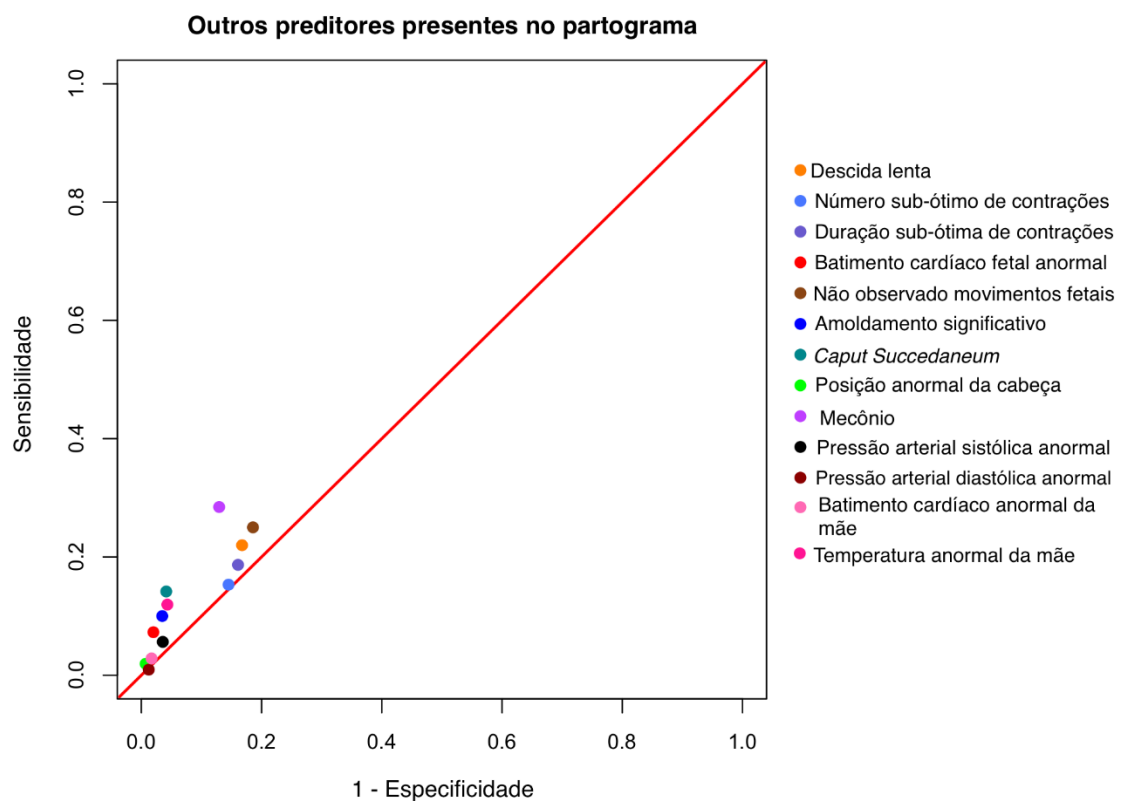


Figura 7. Análise de outros preditores presentes no partograma

Discussão

Os resultados deste trabalho sugerem que os parâmetros presentes no partograma possuem baixa capacidade de predição de desfechos adversos graves. Também é possível observar que o trabalho de parto é altamente variável, sendo assim, é pouco provável que o partograma consiga prever possíveis desfechos adversos graves.

Historicamente, trabalhos de parto com progressão mais lenta que “1 cm/h” tem sido utilizada como gatilho para intervenções no trabalho de parto. Ao reduzir a velocidade de dilatação, a linha de alerta é ultrapassada e muitas vezes são tomadas medidas desnecessárias (6, 13). De acordo com o modelo de cuidado intraparto preconizado pela OMS, mulheres com risco aumentado para desfechos adversos devem ser assistidas com maior frequência. Porém as que são consideradas de baixo risco devem receber a assistência padrão durante o trabalho de parto (6).

Analizando a plotagem das parturientes no partograma, é possível observar a formação de padrões de evolução do trabalho de parto próximos a linha de alerta. Porém, analisando todo o banco de dados, praticamente metade das mulheres cruzaram essa linha, mesmo as que não apresentaram desfechos adversos graves. Somado a isso, mulheres com desfechos adversos graves estão presentes em todo o espectro do instrumento: Antes e após a linha de alerta e após a linha de ação. Considerando o baixo desempenho do preditor, cruzar a linha de alerta não deve ser um problema. Com relação a linha de ação, os números de desempenho sofrem um pequeno aumento, porém desprezíveis.

O projeto BOLD é o maior estudo prospectivo na literatura que correlaciona as curvas de trabalho de parto e seus desfechos. Sendo assim, é plausível acreditar que atualmente não há como obter outro banco de dados que contenha os dados de evolução das parturientes para comparação com a presente análise. Como limitação deste estudo, é necessário levar em conta que qualquer tipo de coleta de dados está sujeita a erros. Dadas as diferenças e as peculiaridades de cada maternidade, monitorar – e principalmente anotar – a evolução do binômio mãe e bebê foi uma tarefa que demandou muita atenção e dedicação dos profissionais envolvidos. Alguns mecanismos foram utilizados para minimizar a heterogeneidade da coleta, como o treinamento dos profissionais e o uso de software gerenciador de dados. Também é importante pontuar que, mesmo com a mínima mudança na rotina padrão, o fato de participar de um estudo pode causar alterações diretas e indiretas nos desfechos

desejados. Isso pode ser explicado pelo efeito de Hawthorne (27), efeito causado por interferências nas ações realizadas por uma pessoa que sabe que está sendo avaliada. Outro ponto importante foi o fornecimento de um sonar doppler para as instituições participantes, que pode ter facilitado o monitoramento fetal e contribuído para o aumento da identificação de anormalidades, impactando no processo de manejo clínico.

Conclusão

Considerando o banco de dados do nosso trabalho, há mulheres com desfechos adversos graves em todo o espectro do partograma, e não apenas as que cruzaram a linha de alerta. Praticamente metade de todas elas – e não só as que tiveram desfechos adversos – cruzaram a linha de alerta. Com isso sugere-se que o partograma possui baixa capacidade de predição de desfechos adversos graves e devem ter seu uso e sua forma revistos.

Referências

1. Alkema L, Chou D, Hogan D, Zhang S, Moller A-B, Gemmill A, et al. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Maternal Mortality Estimation Inter-Agency Group. *The Lancet*. 2016 Jan 30;387(10017):462–74.
2. UN (2018) Sustainable Development Goals [Internet]. United Nations Sustainable Development. [citado 4 de fevereiro de 2019]. Disponível em: <https://www.un.org/sustainabledevelopment/>
3. Say L, Chou D, Gemmill A, Tunçalp Ö, Moller A-B, Daniels J, et al. Global causes of maternal death: a WHO systematic analysis. *The Lancet Global Health*. 2014 Jun 1;2(6):e323–33.
4. Snowden JM, Tilden EL, Snyder J, Quigley B, Caughey AB, Cheng YW. Planned Out-of-Hospital Birth and Birth Outcomes. *New England Journal of Medicine*. 2015 Dec 31;373(27):2642–53.
5. de Jonge A, Geerts CC, van der Goes BY, Mol BW, Buitendijk SE, Nijhuis JG. Perinatal mortality and morbidity up to 28 days after birth among 743 070 low-risk planned home and hospital births: a cohort study based on three merged national perinatal databases. *BJOG*. 2015 Apr;122(5):720–8.
6. Maffi I. The Detour of an Obstetric Technology: Active Management of Labor Across Cultures. *Med Anthropol*. 2016;35(1):17-30.
7. Orji E. Evaluating progress of labor in nulliparas and multiparas using the modified WHO partograph. *International Journal of Gynecology & Obstetrics*. 1º de setembro de 2008;102(3):249–52.
8. Souza JP, Oladapo OT, Bohren MA, Mugerwa K, Fawole B, Moscovici L, et al. The development of a Simplified, Effective, Labour Monitoring-to-Action (SELMA) tool for Better Outcomes in Labour Difficulty (BOLD): study protocol. *Reprod Health*. 2015 May 26;12:49.
9. Friedman EA. The graphic analysis of labor. *Am J Obst Gynecol*. 1954, 68(6):1568-1575
10. Philpott RH, Castle WM. Cervicographs in the management of labour in primigravidae. I. The alert line for detecting abnormal labour. *Journal of Obstetrics and Gynaecology of the British Commonwealth*. 1972a; 79(7):592–598.
11. Philpott RH, Castle WM. Cervicographs in the management of labour in primigravidae. II. The action line and treatment of abnormal labour. *Journal of Obstetrics and Gynaecology of the British Commonwealth*. 1972b; 79(7):599–602.
12. World Health Organization (1993a) Preventing prolonged labour: A Practical Guide. The Partograph Part I. Principle and Strategy. WHO, Geneva WHO/FHE/MSM 93.8
13. O'Driscoll K, Stronge JM. Active management of labor. *British medical journal* 1973; 3(5880): 590.
14. World Health Organization (2003) Managing complications in pregnancy and childbirth: a guide for midwives and doctors. IMPAC. WHO/RHR.007. World Health Organization, Geneva.
15. Zhang J, Landy HJ, Branch DW, et al. Contemporary patterns of spontaneous labor with normal neonatal outcomes. *Obstetrics and gynecology* 2010; 116(6): 1281-7.

16. Neal JL, Lowe NK. Physiologic partograph to improve birth safety and outcomes among low-risk, nulliparous women with spontaneous labor onset. *Med Hypotheses*. 2012 Feb;78(2):319-26. doi: 10.1016/j.mehy.2011.11.012. Epub 2011 Dec 3. PubMed PMID: 22138426; PubMed Central PMCID: PMC3254242.
17. Hamilton EF, Warrick PA, Collins K, Smith S, Garite TJ. Assessing first-stage labor progression and its relationship to complications. *Am J Obstet Gynecol*. 2016 Mar;214(3):358.e1-8.
18. Oladapo OT, Souza JP, Bohren MA, Tunçalp Ö, Vogel JP, Fawole B, et al. WHO Better Outcomes in Labour Difficulty (BOLD) project: innovating to improve quality of care around the time of childbirth. *Reprod Health* 2015: doi:10.1186/s12978-015-0027-6
19. Souza JP, Gülmezoglu AM, Vogel J, Carroli G, Lumbiganon P, Qureshi Z, et al. Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): a cross-sectional study. *Lancet*. 2013; 381(9879):1747-55.
20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics*. 2009 Apr;42(2):377.
21. Barbosa-Junior, F. Cloud, o quê? Para quê e por quê? [Internet]. *Revista Saúde na Comunidade*. 2014 Set;1(4):e32 [cited 2017 Sep 06]. Available from: <https://saudenacomunidade.wordpress.com/2014/09/01/resc2014-e32/> .
22. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014 (<http://www.R-project.org>)
23. Robson MS. Classification of caesarean section. *Fetal and Maternal Medicine Review*.2001;12:23–39. doi: 10.1017/S0965539501000122.
24. Šimundić A-M. Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC*. 20 de janeiro de 2009;19(4):203–11.
25. Lusted LB. ROC Recollected. *Med Decis Making*. 1984 Jun 1;4(2):131–5.
26. Flach, P., Wu, S., 2003. Repairing concavities in ROC curves. In: *Proc. 2003 UK Workshop on Computational Intelligence*. University of Bristol, pp. 38–44.
27. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. *BMC Med Res Methodol*. 2007 Jul 3;7:30.

Projeto 2

Desenvolvimento e aplicação de um novo parâmetro para avaliação da capacidade preditiva de testes diagnósticos: o Índice de Máxima Acurácia

Introdução

O avanço das ciências da saúde tem proporcionado grandes contribuições para a melhora da qualidade de vida das populações. Entretanto, embora a tecnologia da saúde atualmente permita um diagnóstico mais precoce e um tratamento mais efetivo de diversas patologias, ela não é isenta de danos. A estimativa global de causas de morte sugere que mais de 140.000 pessoas percam suas vidas anualmente em decorrência de efeitos adversos de tratamentos (2013) (1). Outras estimativas sugerem que o dano possa ser ainda maior; Makary e Daniel (2016) estimaram o número de mortes decorrente de efeitos adversos de procedimentos diagnósticos ou tratamentos, apenas nos Estados Unidos da América, em 251.000 mortes por ano (2).

Considerando o papel chave que os testes diagnósticos desempenham no manejo de doenças e, pelos efeitos adversos potencialmente decorrentes dos tratamentos, é fundamental que estes testes sejam os mais precisos possíveis. Testes com elevado número de falso positivos geram exames adicionais e tratamentos desnecessários. Analogamente, testes com elevado número de falso negativos contribuem para retardar o diagnóstico e o tratamento dos indivíduos, eventualmente reduzindo suas chances de cura (3).

Desta forma, conhecer a capacidade diagnóstica dos testes é um passo fundamental no desenvolvimento de políticas de saúde e na criação de protocolos clínicos. Essa tarefa, entretanto, não é trivial (4). Ainda que a sensibilidade (5) e a especificidade (5) sejam usualmente os principais parâmetros para a avaliação da capacidade diagnóstica de um teste, eles são complementares: analisados individualmente podem ser insuficientes para a interpretação adequada da capacidade diagnóstica de um determinado teste. Assim, ao longo dos anos foram desenvolvidas algumas métricas integradoras, como as razões de verossimilhança (6), odds ratio diagnóstico (7) e a estatística J (8). Estes parâmetros visam avaliar de forma integrada a capacidade diagnóstica de um teste. Contudo, ainda que esses parâmetros integradores agreguem valor à análise de desempenho diagnóstico,

raramente são utilizados de forma isolada (9–12). A tomada de decisão clínica permanece um desafio para clínicos e desenvolvedores de diretrizes clínicas e de políticas de saúde (13–16).

Representações gráficas são utilizadas desde o início de nossa história para transmitir informação (17,18). O gráfico-radar é um tipo de representação utilizada para apresentar diferentes parâmetros de forma integrada e simultânea (19,20). Cada gráfico-radar possui como formato final um polígono de N lados no quais cada vértice representa um parâmetro que se posiciona no espaço de acordo com a sua distância a partir de um ponto central. Na Figura 8 é apresentado um exemplo de gráfico-radar que possui seis parâmetros em avaliação (X, Y, K, J, M e P). No presente artigo, iremos explorar o uso do gráfico-radar como uma ferramenta para facilitar a comunicação dos parâmetros já existentes utilizados na avaliação da capacidade diagnóstica dos testes. Além disso, iremos apresentar um novo indicador integrativo para avaliação da capacidade preditiva de testes diagnósticos, o Índice de Máxima Acurácia (IMA).

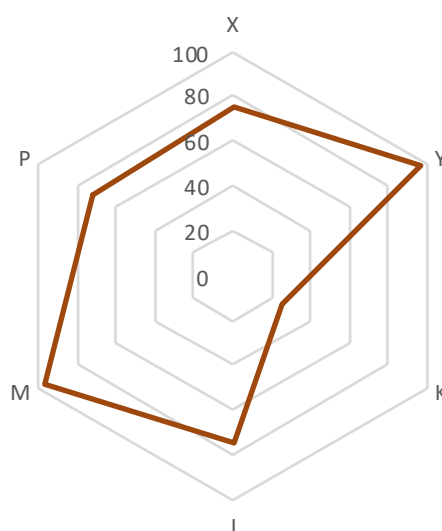


Figura 8. Exemplo de gráfico-radar com seis parâmetros.

Métodos

Trata-se de um estudo exploratório que utilizou o gráfico-radar para comunicar resultados de capacidade diagnóstica de testes e originar um novo indicador integrativo de desempenho. A fonte de dados deste estudo são documentos de diretrizes da Organização Mundial de Saúde (OMS) sobre gravidez e parto publicadas nos anos de 2016, 2017 e 2018. A área temática selecionada como um dos critérios de inclusão deve-se a experiência prévia dos autores na avaliação da capacidade

diagnóstica de testes em Saúde Materna (Anexo 5). As diretrizes são de acesso público e estão disponíveis para *download* no sítio eletrônico da OMS (21). O desenvolvimento e a publicação de diretrizes pela OMS, bem como suas novas edições, seguem rigorosos padrões de qualidade. São utilizados dados de revisões sistemáticas da literatura, seguidas de avaliação crítica da evidência científica e desenvolvimento de consenso internacional. Este processo está descrito no Manual da OMS para Desenvolvimento de Diretrizes (*WHO Handbook for Guideline Development*) (22).

O fluxograma do presente estudo é apresentado na Figura 9. As diretrizes que atendiam aos critérios iniciais de inclusão eram obtidas em formato eletrônico a partir do sítio da OMS e armazenadas para análise posterior. A obtenção destes documentos teve início em fevereiro de 2018 e fim em novembro de 2018. Ao final de cada mês visitava-se o sítio da OMS em busca de novas publicações que atendessem aos critérios de inclusão no estudo. Em caso afirmativo, a nova diretriz, bem como seus possíveis anexos, era obtida e armazenada em formato eletrônico juntamente com as diretrizes já existentes. Ao final do período de inclusão procedeu-se a coleta de dados.

Cada documento foi revisado inicialmente de forma eletrônica e posteriormente de forma manual. Na revisão eletrônica, foi utilizada a funcionalidade de busca do Adobe Acrobat para identificar a ocorrência dos termos “sensibilidade” e “especificidade” em cada documento no formato *portable document file* (PDF). Na revisão manual, cada um dos testes – e seus anexos – apresentados nas diretrizes fora analisado em busca de quadros ou tabelas que apresentassem valores de sensibilidade e especificidade. Em caso negativo, era avaliado se haveria a possibilidade do cálculo de sensibilidade e especificidade a partir do exposto (i.e.: Tabela de contingência). Foram selecionadas todas as diretrizes que avaliaram a capacidade diagnóstica de algum teste e que fosse possível obter os dados de sensibilidade e especificidade. Os valores de sensibilidade e especificidade para cada teste foram tabulados em uma planilha eletrônica utilizando o Microsoft Excel 2016.

O Quadro 4 apresenta os parâmetros comumente utilizados para avaliação de desempenho de testes diagnósticos e seu modo de cálculo. A partir dos valores de sensibilidade e especificidade é possível obter os demais parâmetros, sendo eles: razões de verossimilhanças positiva e negativa, odds ratio diagnóstico e estatística J. Após a tabulação dos valores referentes a cada um dos seis parâmetros para cada

teste incluído no estudo, estes passaram por transformações aritméticas a fim de atender ao formato de visualização do gráfico-radar. Estas transformações estão descritas no Quadro Suplementar S1. Após as transformações, foi utilizada a função de gráficos do Microsoft Excel (2016) para criar as representações.

Em adição a representação gráfica a partir do gráfico-radar, também foi calculada a área do hexágono resultante de cada teste avaliado. O Quadro Suplementar S3 detalha os procedimentos para o cálculo da área do hexágono irregular e da fórmula final do IMA. Em resumo, um hexágono é composto por seis formas triangulares, as quais possuem fórmulas conhecidas para o cálculo da sua área. Com a soma das áreas de cada um dos seis triângulos foi possível obter a área total do hexágono. Supondo um teste que possua máximo desempenho (i.e. sensibilidade igual a 100%, especificidade igual a 100%, razão de verossimilhança positiva ≥ 100 , razão de verossimilhança negativa igual a 0, odds ratio diagnóstico ≥ 100 e estatística J igual a 1) o valor de IMA deverá ser 1. O Índice de Máxima Acurácia é o novo indicador integrativo de desempenho e é calculado da seguinte forma:

$$IMA = \frac{100 * Se * (E + J) + RVP^1 * (1 + E - RVN) + DOR^1 * (1 - RVN + J)}{600}$$

¹Se o valor nestas variáveis for maior do que 100, limitar em 100

Onde:

J: Estatística J

Se: Sensibilidade

E: Especificidade

RVP: Razão de verossimilhança positiva

RVN: Razão de verossimilhança negativa

DOR: Odds ratio diagnóstico

O IMA possui uma variação entre 0 e 1, com o último valor representando a acurácia ideal. Os dados foram analisados no Microsoft Excel 2016, onde foi desenvolvida uma calculadora eletrônica capaz de gerar todo o painel de parâmetros de avaliação de acurácia diagnóstica – juntamente com o gráfico-radar – a partir dos valores de sensibilidade e especificidade. Foi estudada a correlação do IMA com cada um dos parâmetros tradicionalmente utilizados para avaliar a acurácia diagnóstica por meio do índice de correlação de Pearson.

Quadro 4. Principais parâmetros indicadores de desempenho de testes diagnósticos e valores de referência para interpretação

Tabela de contingência	Resultado do teste	Condição Alvo		Valores de Referência*
	Positivo Negativo	Presente VP FN	Ausente FP VN	
Sensibilidade ¹	Detecta a probabilidade um teste classificar positivamente um indivíduo que está doente. $Sensibilidade = \frac{VP}{VP + FN}$			> 0,80
Especificidade ¹	Detecta a probabilidade de um teste classificar negativamente um indivíduo que está sadio. $Especificidade = \frac{VN}{VN + FP}$			> 0,80
Razão de verossimilhança positiva ²	Indica a proporção entre a probabilidade de um paciente doente e um paciente sadio terem um teste positivo. $LH+ = \frac{Sensibilidade}{1 - Especificidade}$			Associado com a presença da doença: > 1 Forte evidência de associação com a doença: > 10
Razão de verossimilhança negativa ²	Indica a proporção entre a probabilidade de um paciente doente e um paciente sadio terem um teste negativo. $LH- = \frac{1 - Sensibilidade}{Especificidade}$			Associado com a ausência da doença: < 0,2
Odds Ratio Diagnóstico ³	Indicador abrangente e integrador de acurácia diagnóstica do teste. Combina os valores de sensibilidade e especificidade. Tem como limitação principal que um mesmo resultado pode significar alta sensibilidade e baixa especificidade como também baixa sensibilidade e alta especificidade. $ORD = \frac{LH+}{LH-}$			Boa associação: >10 Ótima associação: >100
Estatística J ⁴	Indicador abrangente e integrador de acurácia diagnóstica do teste, combinando os dados de sensibilidade e especificidade. $J = Sensibilidade + Especificidade - 1$			> 0,7

* Os valores de referência dependem do uso de cada teste.

VP (Verdadeiro Positivo): Onde a condição alvo é presente e o teste é positivo.

FP (Falso Positivo): Onde a condição alvo não está presente, mas o teste é positivo.

FN (Falso Negativo): Onde a condição alvo é presente, porém o teste é negativo.

VN (Verdadeiro Negativo): Onde a condição alvo não é presente e o teste é negativo.

1. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. BMJ. 11 de junho de 1994;308(6943):1552.

2. Deeks JJ, Altman DG. Statistics Notes: Diagnostic tests 4: likelihood ratios. BMJ. 17 de julho de 2004;329(7458):168–9.

3. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. novembro de 2003;56(11):1129–35.

4. Index for rating diagnostic tests - Youden - 1950 - Cancer - Wiley Online Library [Internet]. [citado 22 de outubro de 2018]. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/1097-014>

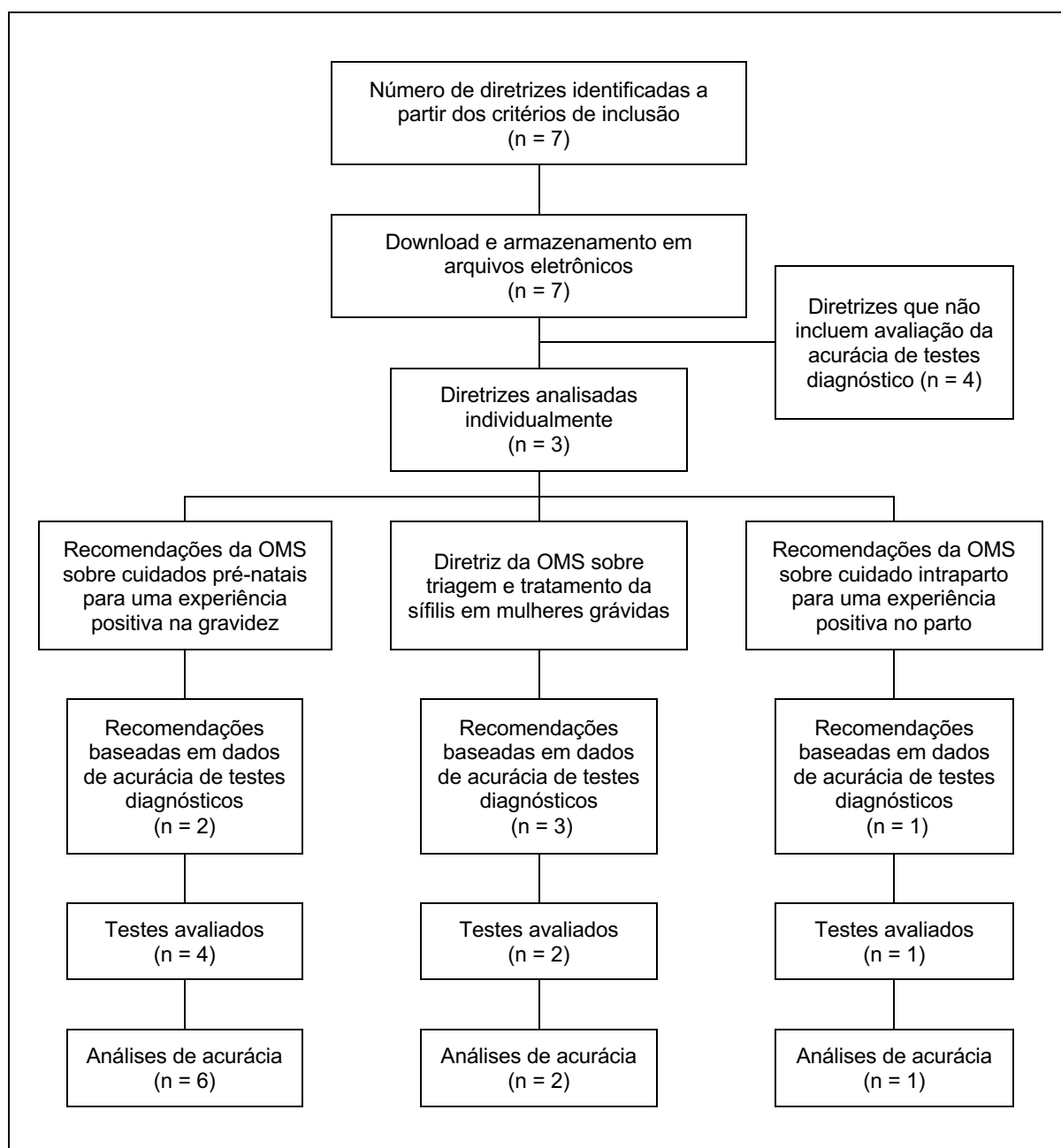


Figura 9. Fluxograma da coleta de dados a partir de diretrizes sobre saúde materna

Resultados

Um total de sete documentos com diretrizes em saúde materna foram publicados pela OMS entre os anos de 2016 a 2018. Destes, três continham recomendações baseadas em análises de capacidade diagnóstica. As características pormenorizadas sobre cada diretriz (número de revisões sistemáticas e estudos independentes utilizados, número de recomendações criadas e o número de recomendações baseadas em dados de testes de capacidade diagnóstica) estão descritas no Quadro Suplementar S2.

O documento de 2016 “*Recomendações da OMS sobre cuidados pré-natais para uma experiência positiva na gravidez*” contém duas recomendações baseadas em análises de capacidade diagnóstica. Quatro testes foram avaliados, sendo dois deles para detecção de anemia (i.e. HemoCue® e Escala de coloração de hemoglobulina da OMS) e outros dois para detecção de bacteriúria assintomática (i.e. Coloração de Gram e Fitas para teste urinário). No caso da anemia, os testes foram avaliados na capacidade de detecção de anemia ($Hb < 110 \text{ g/L}$) e anemia grave ($Hb < 60 \text{ g/L}$). Publicada em 2017, a “*Diretriz da OMS sobre triagem e tratamento da sífilis em mulheres grávidas*” contém oito recomendações, sendo três delas baseadas em análises de capacidade diagnóstica. Foram avaliados dois testes: Teste Rápido para Sífilis e Teste Plasmático Rápido para Sífilis. Por fim, em 2018, a publicação “*Recomendações da OMS sobre cuidado intraparto para uma experiência positiva no parto*” inclui uma recomendação baseada em análises de acurácia diagnóstica. Nesta recomendação é avaliada a capacidade diagnóstica da linha de alerta do partograma em desfechos adversos graves durante o trabalho de parto.

A Tabela 9 apresenta a avaliação dos testes diagnósticos incluídos no estudo. As sensibilidades dos testes incluídos variam de 0,40 (I.C.: 0,38 – 0,42) a 0,86 (I.C.: 0,80 – 0,91). As especificidades variam de 0,47 (I.C.: 0,41 – 0,53) a 0,99 (I.C.: 0,98 – 1,00). Como parâmetro integrador, os valores da estatística J variam entre -0,02 a 0,83. Com exceção da linha de alerta do partograma – classificado como teste não recomendado a partir de fevereiro de 2018 – todos os testes são recomendados para uso em contextos específicos. A Figura 10 apresenta nove gráficos-radar que ilustram cada um dos testes presentes na Tabela 9. Visualmente, os gráficos-radar C e E (que correspondem ao hemoglobinômetro e a coloração de Gram, respectivamente) aparentam ter as maiores áreas inscritas, assim como os maiores valores de IMA. O gráfico-radar I possui a menor área visual e também o mais baixo IMA entre os testes avaliados. Nos testes avaliados, o IMA variou entre 0,040 (gráfico I) e 0,777 (gráfico C). Na Tabela 10 é calculado o Índice de Correlação de Pearson entre o Índice de Máxima Acurácia e os parâmetros presentes no gráfico-radar. Sua variação foi entre 0,645 e 0,950.

Tabela 9. Avaliação dos testes diagnósticos publicados nas diretrizes da OMS entre os anos de 2016 a 2018

Análise	Teste (desfecho)	Sensibilidade (IC 95%)	Especificidade (IC 95%)	Razão de Verossimilhança Positiva	Razão de Verossimilhança Negativa	Odds Ratio Diagnóstico	Estatística J	Posição da diretriz
A	Hemoglobinômetro(HemoCue®)	0,85	0,80	4,25	0,19	22,67	0,65	Teste recomendado em contexto específico
	Anemia (Hb < 110 g/L)	(0,79 – 0,90)	(0,76–0,83)					
B	Escala de coloração de hemoglobulina da OMS	0,75	0,47	1,42	0,53	2,66	0,22	Teste recomendado em contexto específico
	Anemia (Hb < 110 g/L)	(0,71–0,80)	(0,41–0,53)					
C	Hemoglobinômetro (HemoCue®)	0,83	0,99	83,00	0,17	483,35	0,82	Teste recomendado em contexto específico
	Anemia grave (Hb < 60 g/L)	(0,44–0,97)	(0,98–1,00)					
D	Escala de coloração de hemoglobulina da OMS	0,50	0,98	25,00	0,51	49,00	0,48	Teste recomendado em contexto específico
	Anemia grave (Hb < 60 g/L)	(0,15–0,85)	(0,97–0,99)					
E	Coloração de Gram	0,86	0,97	28,67	0,14	198,62	0,83	Teste recomendado em contexto específico
	(bacteriúria assintomática)	(0,80–0,91)	(0,93–0,99)					
F	Fita para teste urinário	0,73	0,89	6,64	0,30	21,88	0,62	Teste recomendado em contexto específico
	(bacteriúria assintomática)	(0,59–0,83)	(0,79–0,94)					
G	Teste rápido para Sífilis	0,83	0,96	20,75	0,18	117,18	0,79	Teste recomendado em contexto específico
	(Sífilis)	(0,58–0,98)	(0,89–1,00)					
H	Teste Plasmático Rápido para Sífilis	0,75	0,97	25,00	0,26	97,00	0,72	Teste recomendado em contexto específico
	(Sífilis)	(0,54–0,88)	(0,96–0,99)					
I	Linha de alerta do partograma	0,40	0,58	0,95	1,03	0,92	-0,02	Teste não recomendado
	(desfechos graves no trabalho de parto)	(0,38–0,42)	(0,57 – 0,58)					

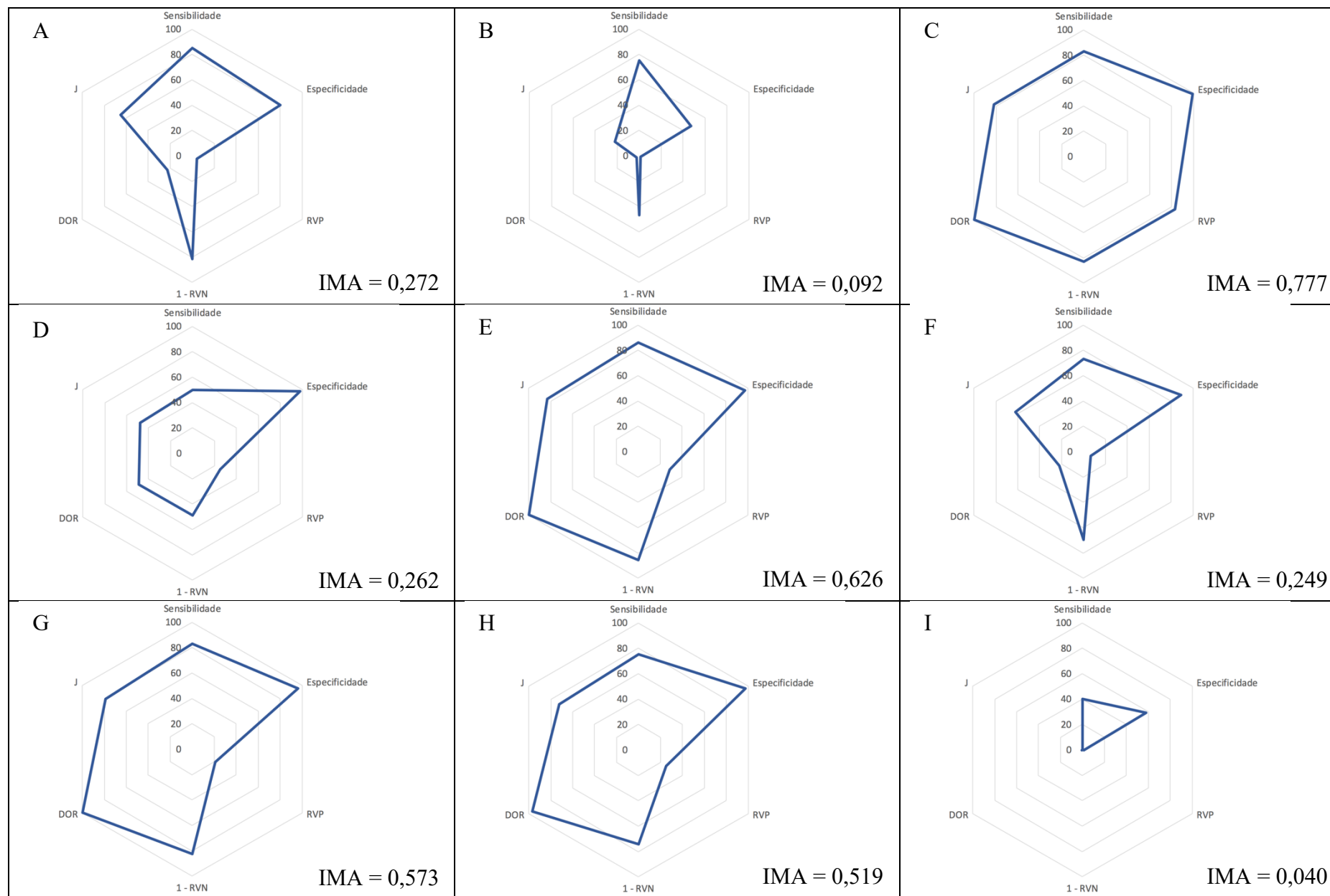


Figura 10. Representação da avaliação de testes diagnósticos utilizando gráficos-radar

Tabela 10. Correlação entre os parâmetros tradicionais de avaliação de desempenho e o IMA

	Coefficiente de Correlação (Pearson)	Interpretação ¹
Sensibilidade	0,645	Correlação moderada
Especificidade	0,788	Correlação forte
RVP	0,824	Correlação forte
1 – RVN	0,771	Correlação forte
DOR	0,950	Correlação muito forte
Estatística J	0,882	Correlação forte

¹Interpretação do coeficiente de correlação [Internet]. [citado 24 de março de 2019]. Disponível em: <http://leg.ufpr.br/~silvia/CE003/node74.html>

Discussão

Este estudo utilizou o gráfico-radar para facilitar a comunicação dos vários parâmetros utilizados na avaliação de desempenho de testes diagnósticos. O cálculo da área do gráfico-radar possibilitou o desenvolvimento de um novo indicador de acurácia diagnóstica – o IMA. A correlação deste novo indicador com os valores de sensibilidade, especificidade, razões de verossimilhança, odds ratio diagnóstico e estatística J foi predominantemente forte, o que sugere que o IMA pode ser uma boa ferramenta para avaliação de desempenho de testes diagnósticos.

Ao longo dos anos, os métodos necessários à interpretação de dados tem passado por várias transformações (17). A evolução da tecnologia de informação (TI) tem um papel essencial nestas mudanças, assim como também para a criação e teste de novos métodos (24). As diferentes formas de visualização gráfica podem trazer novas interpretações para padrões já estabelecidos na literatura (16). O gráfico-radar tem sido proposto como uma alternativa para facilitar a comunicação de informação complexa. Ele pode permitir aos leitores maior rapidez no entendimento e maior clareza na interpretação da informação apresentada (19). Este parece ser o caso da aplicação do gráfico-radar na avaliação da capacidade diagnóstica de testes, pois é possível obter um direcionamento rápido sobre seu desempenho apenas pela visualização gráfica – se comparado com a avaliação dos resultados numéricos (Tabela 9 vs. Figura 10).

Na Figura 10, o gráfico-radar C (hemoglobínômetro) aparenta possuir a maior área inscrita entre os testes avaliados, o que sugere um melhor desempenho. Isso pode ser confirmado pelo valor de IMA – 0,777 – o maior entre eles. Por outro lado, o gráfico-radar I (linha de alerta do partograma) aparenta ter a menor área visual, bem como o menor valor do IMA entre os testes avaliados. Na Tabela 10 é possível concluir que a correlação de Pearson entre o valor de IMA com os parâmetros de avaliação é predominantemente forte, com destaque aos parâmetros integradores odds ratio diagnóstico e estatística J. Isso faz sentido considerando a característica integradora do IMA. Quando confrontados com a análise numérica da Tabela 9, é possível chegar as mesmas interpretações.

Nessa nova proposta de visualização de parâmetros destaca-se a facilidade de comparação entre resultados de testes diagnósticos. O gráfico-radar como apresentado também não depende de ilustração em cores para sua completa interpretação, podendo ser apresentado de forma monocromática. Algumas limitações foram observadas: As transformações matemáticas utilizadas na ferramenta proposta podem não refletir o resultado de determinados testes com exatidão. Por exemplo, os parâmetros com valores negativos podem não ser representados corretamente, pois foi convencionado na ferramenta gráfica o valor zero como limite inferior. Somado a isso, é importante ressaltar que apenas a comparação visual entre diferentes resultados de testes diagnósticos não deve ser o único método de escolha utilizado.

Um aspecto que deve sempre ser lembrado na hora de escolher um teste diagnóstico é a sua finalidade de uso. Por exemplo, um teste sorológico para detectar a infecção pelo HIV pode requerer diferentes qualidades, dependendo de como e onde ele será utilizado. Em um banco de sangue, para a triagem de doadores, será requerida máxima sensibilidade para evitar a transfusão de sangue contaminado, mesmo que às custas da especificidade. Por outro lado, em um serviço de atendimento populacional, o médico requer um teste mais equilibrado, sendo útil a avaliação gráfica e dos valores de IMA.

Embora seja um facilitador da comunicação, a criação de diferentes formas de visualização não é a solução definitiva para as dificuldades de interpretação de dados. É necessário que os usuários deste e de outros métodos possuam capacitação para sua correta interpretação (14). Novos estudos avaliando a contribuição do gráfico-

radar e do IMA na comunicação dos resultados de análise de acurácia diagnóstica são recomendados.

Conclusão

O Índice de Máxima Acurácia se mostrou um facilitador da comunicação dos resultados de avaliação de acurácia diagnóstica. Sua correta utilização pode facilitar o trabalho dos desenvolvedores de diretrizes clínicas e de políticas de saúde.

Referências

1. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*. 10 de janeiro de 2015;385(9963):117–71.
2. Makary MA, Daniel M. Medical error—the third leading cause of death in the US. *BMJ*. 3 de maio de 2016;353:i2139.
3. Borges LSR. Diagnostic Accuracy Measures in Cardiovascular Research. *Int J Cardiovasc Sci* [Internet]. 2016 [citado 31 de agosto de 2018]; Disponível em: <http://www.gnresearch.org/doi/10.5935/2359-4802.20160030>
4. Oliveira GM de, Camargo FT, Gonçalves EC, Duarte CVN, Guimarães CA. Revisão sistemática da acurácia dos testes diagnósticos: uma revisão narrativa. *Rev Colégio Bras Cir*. abril de 2010;37(2):153–6.
5. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. *BMJ*. 11 de junho de 1994;308(6943):1552.
6. Deeks JJ, Altman DG. Statistics Notes: Diagnostic tests 4: likelihood ratios. *BMJ*. 17 de julho de 2004;329(7458):168–9.
7. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 1º de novembro de 2003;56(11):1129–35.
8. Index for rating diagnostic tests - Youden - 1950 - Cancer - Wiley Online Library [Internet]. [citado 22 de outubro de 2018]. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142%281950%293%3A1%3C32%3A%3AAID-CNCR2820030106%3E3.0.CO%3B2-3>
9. Schisterman EF, Perkins N. Confidence Intervals for the Youden Index and Corresponding Optimal Cut-Point. *Commun Stat - Simul Comput*. 4 de maio de 2007;36(3):549–63.
10. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and Optimal Cut-Point Estimated from Observations Affected by a Lower Limit of Detection. *Biom J Biom Z*. junho de 2008;50(3):419–30.
11. Yin J, Tian L. Joint confidence region estimation for area under ROC curve and Youden index. *Stat Med*. 2014;33(6):985–1000.
12. Dujardin B, Van den Ende J, Van Gompel A, Unger J-P, Van der Stuyft P. Likelihood Ratios: A Real Improvement for Clinical Decision Making? *Eur J Epidemiol*. 1994;10(1):29–36.
13. Thornton JG, Lilford RJ, Johnson N. Decision analysis in medicine. *BMJ*. 25 de abril de 1992;304(6834):1099–103.
14. Reid MC, Lane DA, Feinstein AR. Academic Calculations versus Clinical Judgments: Practicing Physicians' Use of Quantitative Measures of Test Accuracy 1. *Am J Med*. 1º de abril de 1998;104(4):374–80.

15. Souza JP, Gülmezoglu AM, Vogel J, Carroli G, Lumbiganon P, Qureshi Z, et al. Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): a cross-sectional study. *The Lancet*. 18 de maio de 2013;381(9879):1747–55.
16. Souza J, Oladapo O, Fawole B, Mugerwa K, Reis R, Barbosa-Junior F, et al. Cervical dilatation over time is a poor predictor of severe adverse birth outcomes: a diagnostic accuracy study. *BJOG Int J Obstet Gynaecol* [Internet]. 17 de abril de 2018 [citado 21 de maio de 2018]; Disponível em: <http://doi.wiley.com/10.1111/1471-0528.15205>
17. Friendly M. Milestones in the history of thematic cartography, statistical graphics, and data visualization. :79.
18. Whiting PF, Sterne JA, Westwood ME, Bachmann LM, Harbord R, Egger M, et al. Graphical presentation of diagnostic information. *BMC Med Res Methodol* [Internet]. dezembro de 2008 [citado 21 de maio de 2018];8(1). Disponível em: <http://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-8-20>
19. Fox CE, Tirlapur SA, Gülmezoglu AM, Souza JP, Khan KS, WITH THE EBM-CONNECT COLLABORATION. Assimilating evidence quality at a glance using graphic display: research synthesis on labor induction: Graphic display for induction of labor evidence. *Acta Obstet Gynecol Scand*. agosto de 2012;91(8):885–92.
20. Find Content Gaps Using Radar Charts [Internet]. Content Strategy Workshops. [citado 29 de janeiro de 2019]. Disponível em: <http://www.content-workshops.com/toolbox/2015/2/find-content-gaps-using-radar-charts>
21. WHO | WHO guidelines: maternal, reproductive and women's health [Internet]. WHO. [citado 29 de janeiro de 2019]. Disponível em: http://www.who.int/publications/guidelines/reproductive_health/en/
22. World Health Organization. WHO handbook for guideline development [Internet]. 2014 [citado 29 de janeiro de 2019]. Disponível em: http://apps.who.int/iris/bitstream/10665/145714/1/9789241548960_eng.pdf
23. Gopalakrishna G, Mustafa RA, Davenport C, Scholten RJPM, Hyde C, Brozek J, et al. Applying Grading of Recommendations Assessment, Development and Evaluation (GRADE) to diagnostic tests was challenging but doable. *J Clin Epidemiol*. 1º de julho de 2014;67(7):760–8.
24. Bauer DT, Guerlain S, Brown PJ. The design and evaluation of a graphical display for laboratory data. *J Am Med Inform Assoc JAMIA*. agosto de 2010;17(4):416–24.

Material Suplementar

Quadro Suplementar S1. Transformações para utilizar o gráfico-radar na avaliação de testes diagnósticos

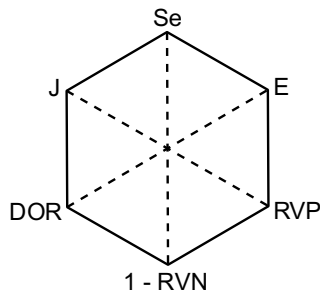
Eixo	Ex. Valor Bruto	Transformação	Valor no gráfico	Ex. Valor Final Visualizado (após transformação)
Sensibilidade	0.7456	Em porcentagem, com 2 casas após a virgula	74.56	74.56%
Especificidade	0.8345	Em porcentagem, com 2 casas após a virgula	83.45	83.45%
RVP	254	Varia de 0 a 100. Valores maiores que 100 ficam igual a 100	100	100%
RVN	0.06	Valor complementar de 1, multiplicado por 100. Se negativo o limite é zero	94	94%
DOR	456	Varia de 0 a 100. Valores maiores que 100 ficam igual a 100	100	100%
Estatística J	0.45	Multiplicado por 100. Se valor negativo, para em zero.	45	45%

Quadro Suplementar S2. Diretrizes em saúde materna publicadas pela OMS entre 2016 e 2018

[illegible]

Quadro Suplementar S3. Fórmula do Índice de Máxima Acurácia

O valor do IMA se refere a porcentagem da área máxima possível de um hexágono de lado igual a 100.



Área do triângulo

$$A = \frac{B * C}{2} \text{Sen}\hat{A}$$

Hexágono: 6 lados

$$360^\circ \div 6 = 60^\circ$$

$$\hat{A} = 60^\circ$$

$$\text{Sen}60^\circ = 0,866$$

1 – Soma-se a área dos seis triângulos que formam o hexágono	$\left(\frac{Se * E}{2} \text{Sen}60\right) + \left(\frac{E * RVP}{2} \text{Sen}60\right) + \left(\frac{RVP * (1 - RVN)}{2} \text{Sen}60\right) + \left(\frac{(1 - RVN) * DOR}{2} \text{Sen}60\right) + \left(\frac{DOR * J}{2} \text{Sen}60\right) + \left(\frac{J * Se}{2} \text{Sen}60\right)$
2 – Manipulação aritmética	$\frac{\left((Se * E) + (E * RVP) + (RVP * (1 - RVN)) + ((1 - RVN) * DOR) + (DOR * J) + (J * Se)\right) * \text{Sen}60}{2}$
3 – A área máxima do hexágono com lado 100	$A^{m\acute{a}x} = \frac{\left((100 * 100) + (100 * 100) + (100 * (100)) + ((100) * 100)\right) * \text{Sen}60}{2}$ $A^{m\acute{a}x} = \frac{60.000 * 0,866}{2} = 25.980$
4 – O valor de IMA é a divisão da soma das áreas de cada triângulo (passo 1) pela área máxima possível (passo 3)	$\frac{\left(\left(\frac{Se * E}{2} \text{Sen}60\right) + \left(\frac{E * RVP}{2} \text{Sen}60\right) + \left(\frac{RVP * (1 - RVN)}{2} \text{Sen}60\right) + \left(\frac{(1 - RVN) * DOR}{2} \text{Sen}60\right) + \left(\frac{DOR * J}{2} \text{Sen}60\right) + \left(\frac{J * Se}{2} \text{Sen}60\right)\right)}{25.980}$
5 – Manipulação aritmética	$\frac{(Se * (E + J)) + (RVN * (RVP + DOR)) + (DOR * J) + (E * RVP)}{60.000}$
6 – Considerando que são necessárias conversões em alguns dos parâmetros antes da substituição na fórmula, outro modo, que já engloba a conversão dos valores, pode ser apresentado	$\frac{\left((100Se * 100E) + (100E * RVP) + (RVP * (1 - 100RVN)) + ((1 - 100RVN) * DOR) + (DOR * 100J) + (100J * 100Se)\right) * \frac{\text{Sen}60}{2}}{25.980}$
7 – Manipulação aritmética	$\frac{(100 * (DOR * (J - RVN + 1) + 100J * Se - RVP(RVN - E - 1) + 100E * Se) * \frac{\text{Sen}60}{2}}{\frac{60.000 * 0,866}{2}}$
Fórmula final	$IMA = \frac{100 * Se * (E + J) + RVP^1 * (1 + E - RVN) + DOR^1 * (1 - RVN + J)}{600}$

Considerações Finais

Na primeira parte desta dissertação foi desenvolvida a expertise na avaliação da capacidade preditiva de classificadores por meio de uma abrangente análise da capacidade diagnóstica dos classificadores incluídos no partograma. Os resultados da análise indicam que o trabalho de parto é um fenômeno extremamente variável, com a possibilidade da ocorrência de desfechos adversos graves nas mais variadas situações. Praticamente metade de todas as mulheres incluídas neste trabalho – e não só as que tiveram desfechos adversos – cruzaram a linha de alerta do partograma. Com isso, a análise do partograma por meio dos principais indicadores de desempenho sugere que sua linha de alerta possui baixa capacidade de predição de desfechos adversos graves. Os resultados deste trabalho sugerem que os parâmetros incluídos no partograma possuem baixa capacidade de predição de desfechos adversos e, isoladamente, possuem reduzida capacidade de guiar a tomada de decisão clínica.

Na segunda parte desta dissertação, o gráfico-radar foi utilizado para facilitar a comunicação dos vários parâmetros utilizados na avaliação de desempenho. Foi também desenvolvido e testado um novo indicador de acurácia diagnóstica – o Índice de Máxima Acurácia. A utilização do gráfico-radar e do Índice de Máxima Acurácia parecem ser facilitadores da comunicação dos resultados de avaliação de acurácia diagnóstica. Sua utilização pode facilitar o trabalho dos desenvolvedores de diretrizes clínicas e de políticas de saúde.

Em 15 de fevereiro de 2018, em Genebra, Suíça, com base nos presentes resultados, a OMS alterou sua recomendação de uso para o partograma.

Referências

1. Moser RH. Diseases of Medical Progress. New England Journal of Medicine. 27 de setembro de 1956;255(13):606–14.
2. Barr DP. Hazards of modern diagnosis and therapy: the price we pay. J Am Med Assoc. 10 de dezembro de 1955;159(15):1452–6.
3. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet. 10 de janeiro de 2015;385(9963):117–71.
4. Šimundić A-M. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC. 20 de janeiro de 2009;19(4):203–11.
5. Wang N, Zeng NN, Zhu W. Sensitivity, Specificity, Accuracy, Associated Confidence Interval And ROC Analysis With Practical SAS Implementations. 2010;9.
6. Van den Bruel A, Cleemput I, Aertgeerts B, Ramaekers D, Buntinx F. The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. Journal of Clinical Epidemiology. 1º de novembro de 2007;60(11):1116–22.
7. Reid MC, Lane DA, Feinstein AR. Academic Calculations versus Clinical Judgments: Practicing Physicians' Use of Quantitative Measures of Test Accuracy 1. The American Journal of Medicine. 1º de abril de 1998;104(4):374–80.
8. Pearson K. On the theory of contingency and its relation to association and normal correlation. Londres. 1904. [internet]. [citado em 01 de março de 2019]. Disponível em: <https://archive.org/details/cu31924003064833/page/n2>
9. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. BMJ. 11 de junho de 1994;308(6943):1552.
10. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. novembro de 2003;56(11):1129–35.
11. Youden WJ. Index for rating diagnostic tests. Cancer. Wiley Online Library. 1950. [Internet]. [citado 22 de outubro de 2018]. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142%281950%293%3A1%3C32%3A%3AAID-CNCR2820030106%3E3.0.CO%3B2-3>

Anexos

Anexo 1. Partograma

Name	Gravida	Para.	Hospital no.
Date of admission	Time of admission	Ruptured membranes	hours
180			
170			
160			
150			
140			
130			
120			
110			
100			
Fetal heart rate			
Liquor Moulding			
10			
9			
8			
7			
6			
5			
4			
3			
2			
1			
0			
Hours			
Time			
5			
4			
3			
2			
1			
Contractions per 10 mins			
Oxytocin U/L drops/min			
Drugs given and IV fluids			
180			
170			
160			
150			
140			
130			
120			
110			
100			
90			
80			
70			
60			
Pulse			
and BP			
Temp °C			
Urine { protein			
acetone			
volume			

Source: WHO, used by permission

Anexo 2. Formulário de coleta de dados



World Health
Organization



Better Outcomes in Labour Difficulty Project

INDIVIDUAL DATA COLLECTION FORM

Instructions

1. This form is composed of sections.
2. Section A **MUST** be completed for all women admitted for vaginal delivery.
3. If the woman is not in first stage of labour or induction of labour (Q12a=0) or if any of the other conditions is present (Q12b to Q12l equals 1), the woman is **NOT** eligible to participate. Informed consent is mandatory.
4. All Sections **MUST** be completed as applicable for **ALL** women included in the study
5. In case of inadequate data in the woman's case record, obtain information from attending health provider.
6. In case of erroneous entry, cross out neatly, write the correct answer outside the box and append your initials.
7. In case the data is not known, missing, or not applicable, fill in with "9", "99", or "999" (except where skip instructions are provided).
8. The use of "9", "99", or "999" is not allowed in Section F and Q74.

Data obtained through this form will help the development of a Simplified, Effective, Labour Monitoring-to-Action tool (SELMA)

 World Health Organization	Better Outcomes in Labour Difficulty Project SELMA Development	INDIVIDUAL FORM Page 1/10
--	---	-------------------------------------

A	Hospital admission - All women	
----------	---------------------------------------	--

1) Identification:

a) Country code (N=Nigeria U=Uganda)

b) Facility code

c) Individual form number

d) Hospital ID#

2) Date of hospital admission

d	d	m	m	y	y

3) Time of hospital admission (24h format)

h	h	m	m

4) Age (years)

5) Marital status

0= Married / Cohabiting

1= Single / Separated / Divorced / Widowed

6) Gravidity

(Number of pregnancies including current)

7) Parity

(Number of previous births)

8) Number of previous abortions

(includes induced and spontaneous abortions)

9) Number of previous C-Sections

10) Ethnic group

Nigeria: 1= Ibo 2= Yoruba 3= Hausa 4= Fulani 5= TIV 6= Kanuri 7= Other Nig. 8= Non-Nigerian	Uganda: 9= Muganda/Musoga/Mugisu 10= Munyakore/Mukiga/ Munyoro/Mutoro 11=Acholi/Langi/Alur 12=Iteso/Karamojong 13=Lugbara/Madi 14=Other Ugandan 15= Non-Ugandan
---	---

11) Education level. Specify:

0= No education

1= Pre-primary education

2= Incomplete primary education

3= Complete primary education

4= Incomplete secondary education

5= Complete secondary education

6= Incomplete post-secondary/tertiary education

7= Complete post-secondary/tertiary or higher education

8= Other (e.g. Quranic / Nomadic education only)

12) Indicate whether the conditions below are present or not (0=No 1=Yes)

a) First stage of labour (spontaneous / induced)

b) Fetal death

c) Advanced first stage of labour (cervical dilation ≥7cm)

d) Multiple pregnancy

e) Gestational age less than 34 weeks

f) Elective C-section

g) Pre-labour C-section

h) Indication for emergency C-Section or laparotomy on admission

i) Attempted induction of labour, but no labour achieved

j) False labour

k) Non-emancipated minors w/o a guardian

l) Woman is not capable of giving consent due to labour distress or any health problem

If the woman is not in labour (Q12a=0) or if any of the other conditions is present (Q12b to Q12l equals 1), the woman is **NOT** eligible to participate.

13) Does the woman consent to participate in the study? (0=No 1=Yes 9=Not eligible)

Data Collector Name:	Date:
-----------------------------	--------------

 World Health Organization	Better Outcomes in Labour Difficulty Project SELMA Development	INDIVIDUAL FORM Page 2/10
--	---	-------------------------------------

B	Hospital admission - only eligible women	
14) Gainful occupation (0=No 1=Yes) <input style="width: 30px;" type="checkbox"/> 15) In case of any previous abortion, specify: (In case of no previous abortion, use 99) a) Number of induced abortions b) Number of spontaneous abortions 16) Any previous uterine surgery <input style="width: 30px;" type="checkbox"/> (0=No 1=Yes) (e.g. myomectomy) Note: Previous CS not included 17) Antecedent history of prolonged labour / obstructed labour <input style="width: 30px;" type="checkbox"/> (0=No 1=Yes) 18) Number of previous stillbirths 19) Outcome of last childbirth 0= no previous childbirth 1= live birth, still alive 2= live birth, deceased 3= stillbirth 20) In case of previous childbirths, what is the date of last delivery? ddmmyy 21) Pre-pregnancy health conditions (0=No 1=Yes) a) Chronic Hypertension b) Diabetes Mellitus c) HIV + d) AIDS / HIV wasting syndrome e) Chronic Anaemia (e.g. sickle cell anaemia) f) Obesity g) Heart disease h) Lung disease i) Renal disease j) Other chronic disease <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/>		22) Complications in current pregnancy (0=No 1=Yes) a) Placenta praevia b) Accreta/increta/percreta placenta c) Abruptio placentae d) Other obstetric haemorrhage e) Pre-eclampsia (excludes eclampsia) f) Eclampsia g) Pyelonephritis h) Malaria i) Preterm rupture of membranes j) Anaemia (Hct ≤ 26% or Hb ≤ 9g/dL) k) Gestational diabetes l) Other pregnancy complications <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> <input style="width: 20px;" type="checkbox"/> 23) Number of antenatal care visits 24) Women referred in labour from another health facility (0= No 1= Yes) 25) Best estimate of gestational age (in complete weeks) 26) Method of gestational age estimation 0= Symphysis-fundal height (cm) 1= Birth weight assessment 2= Last menstrual period 3= Neonatal physical findings (e.g. Ballard) 4= Ultra-sound scan 27) Mode of onset of labour 0= spontaneous 1= induced 28) If spontaneous onset of labour, specify duration of labour before admission in hours 29) Fetal movements in the last 2 hours, as self reported. (0=no changes/increased 1=reduced 2=absent)

 World Health Organization	Better Outcomes in Labour Difficulty Project SELMA Development	INDIVIDUAL FORM Page 3/10
B Hospital admission - only eligible women (first assessment)		
30) Time of first assessment (24h format) h h m m 31) Maternal anthropometric and physical exam. findings: a) Current weight (kg) b) Height (cm) c) Foot length (cm) d) Maternal heart rate e) Systolic blood pressure (mmHg) f) Diastolic blood pressure (mmHg) g) Axillary temperature (°C) h) Symphysis-fundal height (cm) i) Abdominal circumference (cm) j) Number of uterine contractions (10 min) k) Duration of uterine contractions l) Fetal heart rate (bpm) m) Fetal movements (0=No 1=Yes) n) Cervical dilatation (cm) (0=No 1=Yes) o) Cervix effacement 0=Thick (<30%) 2=Thin (up to 80%) 1=Medium (up to 50%) 3=Very thin (>80%) p) Cervix position (0=anterior 1=central 2=posterior)	q) Cervix consistency (0=soft 1=medium 2=firm) r) Fetal presentation (0= Cephalic 1=Breech 2=Transverse/other) s) Fetal station 0= above ischial spine 1= at ischial spine 2= below ischial spine t) Position of fetal head 0= Occiput Anterior (includes right and left) 1= Occiput transverse 2= Occiput posterior 3= Other u) Caput succedaneum 0=None 1=Mild 2=Moderate 3=Severe v) Moulding 0= None 2= Second degree 1= First degree 3= Third degree w) Amniotic membranes status 0= Intact 1= Ruptured without meconium 2= Rupture with stale meconium 3= Ruptured with fresh meconium x) Clinical pelvic assessment (Specify: 0=No 1=Yes 9=Not assessed) 1. Sacral promontory reached 2. Ischial spines prominent 3. Pubic angle admits less than 2 fingers 32) To what extent has the woman been bothered by: a) Emotional problems? (e.g. fear, anxiety, irritability, depression, sadness) b) Labour pain? (USE THE CHART) 33) Labour companionship? (0=No 1=Yes) 	

C Intrapartum data		Page 4/10	C1	C2	C3	C4	C5	C6
		34) Time of assessment:						
		35) Full Assessment? (0=No 1=Yes)						
36) Number of uterine contractions in 10 minutes								
37) Duration of uterine contractions (seconds)								
38) Fetal heart rate (b.p.m.)								
39) Fetal movements observed/felt (0=No 1=Yes)								
40) Cervical dilatation (cm)								
41) Fetal presentation (0=cephalic 1=breech 2=transverse lie / compound / other)								
42) Fetal station (0=above ischial spine 1= at ischial spine 2=below ischial spine)								
43) Moulding (0=None 1=First degree 2=Second degree 3=Third degree)								
44) Caput succedaneum 0=None 1=Mild 2=Moderate 3=Severe								
45) Position of fetal head 0=Occiput Anterior (includes right and left) 1=Occiput transverse 2=Occiput posterior 3=Other								
46) Amniotic membranes status 0= intact 1= ruptured without meconium 2= ruptured with stale meconium 3= ruptured with fresh meconium								
47) Systolic Blood Pressure (mmHg)								
48) Diastolic Blood Pressure (mmHg)								
49) Maternal heart rate (b.p.m.)								
50) Axillary temperature (°C, with decimals)								
51) Since the last assessment, to what extent has the woman been bothered by labour pain? (USE CHART)								
52) Since the last assessment, to what extent has the woman been bothered by emotional problems such as fear, anxiety, depression, irritability, or sadness? (USE CHART)								
53) Oxytocin infusion (0=No 1=Yes)								
54) Oxytocin dilution (0= 0.25 amp/500ml 1= 0.5 amp/500ml 2= 1 amp/500ml 3=2amp/500ml 4=Other)								
55) Oxytocin rate (drops/min)								
56) Analgesia (0=None 1=IV/IM Opioid 2=Epidural 3=Spinal 4=Other 5=Combined)								
57) Labour companionship (0=No 1=Yes)								
58) Predominant maternal position between assessments 0= upright, sitting, standing, walking, kneeling, squatting, all-4 1= recumbent, semi-recumbent, lateral, supine								
59) Oral fluid intake between assessments (0=No 1=Yes)								
60) Oral food intake between assessments (0=No 1=Yes)								
61) IV Fluids (0=No 1=Yes)								
Instructions 1. Record all full assessments (completely) 2. Between full assessments, record only abnormal values (e.g. FHR less than 120 or more than 160). This apply to FHR and maternal vital signs. 3. In full assessments, if information could not be retrieved, or was not assessed, or was missing, use 999. This only applies to full assessments. 4. Data collectors need to ensure that all information is obtained. If information is not in medical records, approach the assisting health provider or observe the pregnant woman.								

C Intrapartum data		Page 5/10	C7	C8	C9	C10	C11	C12
		34) Time of assessment:						
		35) Full Assessment? (0=No 1=Yes)						
36) Number of uterine contractions in 10 minutes								
37) Duration of uterine contractions (seconds)								
38) Fetal heart rate (b.p.m.)								
39) Fetal movements observed/felt (0=No 1=Yes)								
40) Cervical dilatation (cm)								
41) Fetal presentation (0=cephalic 1=breech 2=transverse lie / compound / other)								
42) Fetal station (0=above ischial spine 1= at ischial spine 2=below ischial spine)								
43) Moulding (0=None 1=First degree 2=Second degree 3=Third degree)								
44) Caput succedaneum 0=None 1=Mild 2=Moderate 3=Severe								
45) Position of fetal head 0=Occiput Anterior (includes right and left) 1=Occiput transverse 2=Occiput posterior 3=Other								
46) Amniotic membranes status 0= intact 1= ruptured without meconium 2= ruptured with stale meconium 3= ruptured with fresh meconium								
47) Systolic Blood Pressure (mmHg)								
48) Diastolic Blood Pressure (mmHg)								
49) Maternal heart rate (b.p.m.)								
50) Axillary temperature (°C, with decimals)								
51) Since the last assessment, to what extent has the woman been bothered by labour pain? (USE CHART)								
52) Since the last assessment, to what extent has the woman been bothered by emotional problems such as fear, anxiety, depression, irritability, or sadness? (USE CHART)								
53) Oxytocin infusion (0=No 1=Yes)								
54) Oxytocin dilution (0= 0.25 amp/500ml 1= 0.5 amp/500ml 2= 1 amp/500ml 3=2amp/500ml 4=Other)								
55) Oxytocin rate (drops/min)								
56) Analgesia (0=None 1=IV/IM Opioid 2=Epidural 3=Spinal 4=Other 5=Combined)								
57) Labour companionship (0=No 1=Yes)								
58) Predominant maternal position between assessments 0= upright, sitting, standing, walking, kneeling, squatting, all-4 1= recumbent, semi-recumbent, lateral, supine								
59) Oral fluid intake between assessments (0=No 1=Yes)								
60) Oral food intake between assessments (0=No 1=Yes)								
61) IV Fluids (0=No 1=Yes)								
Instructions 1. Record all full assessments 2. Between full assessments, record only abnormal values (e.g. FHR less than 120 or more than 160). This apply to FHR and maternal vital signs. 3. In full assessments, if information could not be retrieved, or was not assessed, or was missing, use 999. This only applies to full assessments. 4. Data collectors need to ensure that all information is obtained. If information is not in medical records, approach the assisting health provider or observe the pregnant woman								

C Intrapartum data		Page 6/10	C13	C14	C15	C16	C17	C18
		34) Time of assessment:						
		35) Full Assessment? (0=No 1=Yes)						
36) Number of uterine contractions in 10 minutes								
37) Duration of uterine contractions (seconds)								
38) Fetal heart rate (b.p.m.)								
39) Fetal movements observed/felt (0=No 1=Yes)								
40) Cervical dilatation (cm)								
41) Fetal presentation (0=cephalic 1=breech 2=transverse lie / compound / other)								
42) Fetal station (0=above ischial spine 1= at ischial spine 2=below ischial spine)								
43) Moulding (0=None 1=First degree 2=Second degree 3=Third degree)								
44) Caput succedaneum 0=None 1=Mild 2=Moderate 3=Severe								
45) Position of fetal head 0=Occiput Anterior (includes right and left) 1=Occiput transverse 2=Occiput posterior 3=Other								
46) Amniotic membranes status 0= intact 1= ruptured without meconium 2= ruptured with stale meconium 3= ruptured with fresh meconium								
47) Systolic Blood Pressure (mmHg)								
48) Diastolic Blood Pressure (mmHg)								
49) Maternal heart rate (b.p.m.)								
50) Axillary temperature (°C, with decimals)								
51) Since the last assessment, to what extent has the woman been bothered by labour pain? (USE CHART)								
52) Since the last assessment, to what extent has the woman been bothered by emotional problems such as fear, anxiety, depression, irritability, or sadness? (USE CHART)								
53) Oxytocin infusion (0=No 1=Yes)								
54) Oxytocin dilution (0= 0.25 amp/500ml 1= 0.5 amp/500ml 2= 1 amp/500ml 3=2amp/500ml 4=Other)								
55) Oxytocin rate (drops/min)								
56) Analgesia (0=None 1=IV/IM Opioid 2=Epidural 3=Spinal 4=Other 5=Combined)								
57) Labour companionship (0=No 1=Yes)								
58) Predominant maternal position between assessments 0= upright, sitting, standing, walking, kneeling, squatting, all-4 1= recumbent, semi-recumbent, lateral, supine								
59) Oral fluid intake between assessments (0=No 1=Yes)								
60) Oral food intake between assessments (0=No 1=Yes)								
61) IV Fluids (0=No 1=Yes)								
Instructions 1. Record all full assessments 2. Between full assessments, record only abnormal values (e.g. FHR less than 120 or more than 160). This apply to FHR and maternal vital signs. 3. In full assessments, if information could not be retrieved, or was not assessed, or was missing, use 999. This only applies to full assessments. 4. Data collectors need to ensure that all information is obtained. If information is not in medical records, approach the assisting health provider or observe the pregnant woman.								

 World Health Organization	Better Outcomes in Labour Difficulty Project SELMA Development	INDIVIDUAL FORM Page 7/10
CHART		D
		Intrapartum interventions
0 - Not at all		62) Cervical ripening (CR) (0=No 1=Yes) ☐ (if no cervical ripening, jump to Q63)
		a) Date CR starts d d m m y y
		b) Time CR starts h h m m
		c) Specify method of CR (0= No 1= Yes): 1. Oxytocin 2. Misoprostol 3. Other prostaglandin 4. Sweep membranes 5. Amniotomy 6. Mechanical (e.g. Foley Catheter)
1 - Slightly		63) Induction of labour (IOL) (0=No 1=Yes) ☐ (If no induction of labour, jump to Q64)
		a) Date IOL starts d d m m y y
		b) Time IOL starts h h m m
		c) IOL indication (0=No 1=Yes) 1. Suspected fetal growth impairment 2. Prelabour rupture of membranes 3. Chorioamnionitis 4. Pre-eclampsia/eclampsia 5. Gestational ≥ 41 weeks 6. Any other obstetric complication 7. Any other medical complication 8. Maternal request 9. IOL without medical reason
2 - Moderately		
3 - Quite a bit		
4 - Extremely		

D	Intrapartum interventions (Cont'd)													
	d) Specify method of induction (0= No 1= Yes): <table style="width: 100%;"> <tr><td>1. Oxytocin</td><td style="text-align: center;">☐</td></tr> <tr><td>2. Misoprostol</td><td style="text-align: center;">☐</td></tr> <tr><td>3. Other prostaglandin</td><td style="text-align: center;">☐</td></tr> <tr><td>4. Sweep membranes</td><td style="text-align: center;">☐</td></tr> <tr><td>5. Amniotomy</td><td style="text-align: center;">☐</td></tr> <tr><td>6. Mechanical (e.g. Foley Catheter)</td><td style="text-align: center;">☐</td></tr> </table>	1. Oxytocin	☐	2. Misoprostol	☐	3. Other prostaglandin	☐	4. Sweep membranes	☐	5. Amniotomy	☐	6. Mechanical (e.g. Foley Catheter)	☐	d) Complication ☐ 0= No 1= Cord prolapse 2= Fetal distress 3= Other
1. Oxytocin	☐													
2. Misoprostol	☐													
3. Other prostaglandin	☐													
4. Sweep membranes	☐													
5. Amniotomy	☐													
6. Mechanical (e.g. Foley Catheter)	☐													
	64) Augmentation of labour (0=No 1=Yes) ☐ (If no augmentation of labour, jump to Q65) a) Time of decision h h m m b) Time infusion starts h h m m c) Indication ☐ 0= weak/irregular contractions with slow progress 1= weak/irregular contractions w/o slow progress 2= other	66) Caesarean section (CS) ☐ 0= not performed and not decided on 1= not performed but decided on 2= performed (If no Caesarean section, jump to Q67) a) Time of decision h h m m b) Time of incision h h m m Conditions preceeding the CS: 0= Fetal distress 1= Prolonged labour 2= Cephalopelvic disproportion 3= Obstruction 4= Suspected/imminent uterine rupture 5= Failed operative vaginal delivery 6= Pre-eclampsia/eclampsia 7= Gestational $\geq$ 41 weeks 8= Intrapartum vaginal bleeding 9= Breech or other malpresentation 10= Previous CS 11= Failed induction 12= Maternal request 13= HIV 14= Genital Herpes/extensive condyloma 15= Previous uterine surgery 16= Other complication												
	65) Amniotomy (0=No 1=Yes) ☐ (If no amniotomy, jump to Q66) a) Date of amniotomy d d m m y y b) Time of amniotomy h h m m c) Indication ☐ 0= accidental rupture 1= routine rupture 2= slow progress	c) Primary indication of C-Section d) Secondary indication of C-Section 												

 World Health Organization	Better Outcomes in Labour Difficulty Project SELMA Development	INDIVIDUAL FORM Page 9/10
--	---	-------------------------------------

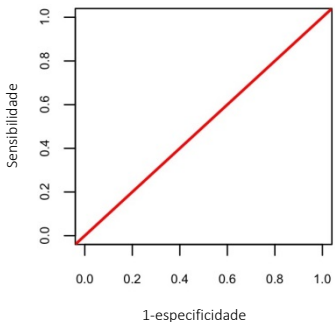
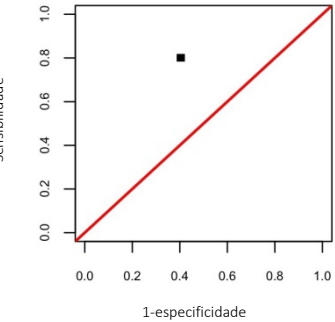
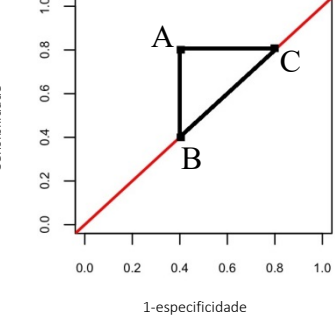
E	Childbirth and Neonatal Outcomes																													
67) Date of delivery <table style="display: inline-table; border-collapse: collapse;"> <tr> <td>d</td><td>d</td><td>m</td><td>m</td><td>y</td><td>y</td> </tr> <tr> <td> </td><td> </td><td> </td><td> </td><td> </td><td> </td> </tr> </table>		d	d	m	m	y	y							75) Newborn status at hospital discharge ☐ (0= Alive 1=Dead)																
d	d	m	m	y	y																									
68) Time of delivery <table style="display: inline-table; border-collapse: collapse;"> <tr> <td>h</td><td>h</td><td>m</td><td>m</td> </tr> <tr> <td> </td><td> </td><td> </td><td> </td> </tr> </table>		h	h	m	m					76) Newborn discharge, transfer or death a) Date <table style="display: inline-table; border-collapse: collapse;"> <tr> <td>d</td><td>d</td><td>m</td><td>m</td><td>y</td><td>y</td> </tr> <tr> <td> </td><td> </td><td> </td><td> </td><td> </td><td> </td> </tr> </table> b) In case of newborn death (Q75=1) what is the time of newborn death? <table style="display: inline-table; border-collapse: collapse;"> <tr> <td>h</td><td>h</td><td>m</td><td>m</td> </tr> <tr> <td> </td><td> </td><td> </td><td> </td> </tr> </table>	d	d	m	m	y	y							h	h	m	m				
h	h	m	m																											
d	d	m	m	y	y																									
h	h	m	m																											
69) Final mode of delivery. Specify: ☐ 0= Spontaneous vaginal delivery without episiotomy 1= Spontaneous vaginal delivery with episiotomy 2= Operative vaginal delivery (forceps or vacuum) 3= Caesarean section 4= Laparotomy		F Maternal Outcomes																												
70) Fetal presentation at delivery ☐ 0=Cephalic 1=Breech 2=Transverse Lie / Other		77) Was any of the following conditions identified during hospital stay? (0=No 1=Yes)																												
71) Infant sex (0=Male 1=Female) ☐		<i>Dystocia</i> a) Labour obstruction ☐ b) Prolonged first stage of labour ☐ c) Prolonged second stage ☐																												
72) Birth weight (g) <table style="display: inline-table; border-collapse: collapse;"><tr><td> </td><td> </td><td> </td><td> </td></tr></table>						<i>Haemorrhage</i> d) Placenta praevia ☐ e) Accreta/increta/percreta placenta ☐ f) Abruptio placenta ☐ g) Ruptured uterus ☐ h) Intrapartum haemorrhage ☐ i) Postpartum haemorrhage ☐																								
73) Neonatal conditions at birth a) Vital status (0= Alive 1= Stillbirth) ☐ b) Apgar score at 5 min <table style="display: inline-table; border-collapse: collapse;"><tr><td> </td><td> </td></tr></table>				<i>Infection</i> j) Puerperal endometritis ☐ k) Puerperal sepsis ☐ l) Wound infection ☐ m) Systemic infection / septicaemia ☐																										
74) Neonatal severe morbidity. Specify (0=No 1=Yes) a) Admission to neonatal ICU/special care unit ☐ b) Any intubation (at birth or during the 1st wk) ☐ c) Nasal CPAP ☐ d) Surfactant administration ☐ e) Cardio-pulmonary resuscitation ☐ f) Use of any vasoactive drug ☐ g) Use of anticonvulsants ☐ h) Use of phototherapy in the first 24 hours ☐ i) Use of any blood products ☐ j) Use of steroids to treat hypoglycaemia ☐ k) Use of therapeutic intravenous antibiotics ☐ l) Any surgery ☐ m) Any severe malformation ☐		<i>Hypertension</i> n) Severe hypertension ☐ o) Pre-eclampsia (excludes eclampsia) ☐ p) Eclampsia ☐																												
		<i>Other maternal complications</i> q) Embolic disease ☐ r) Other potentially life-threatening condition ☐																												

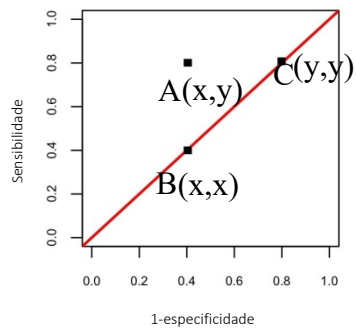
 World Health Organization	Better Outcomes in Labour Difficulty Project SELMA Development	INDIVIDUAL FORM Page 10/10
--	---	--------------------------------------

F	Maternal Outcomes	79) Please specify whether the women used any of the following interventions (0=No 1=Yes) <i>Haemorrhage</i> <table style="width: 100%;"> <tr><td>a) Oxytocin for treatment of PPH</td><td style="text-align: center;">☐</td></tr> <tr><td>b) Misoprostol for treatment of PPH</td><td style="text-align: center;">☐</td></tr> <tr><td>c) Ergotamine for treatment of PPH</td><td style="text-align: center;">☐</td></tr> <tr><td>d) Other uterotonic for treatment of PPH</td><td style="text-align: center;">☐</td></tr> <tr><td>e) Artery ligation or embolization</td><td style="text-align: center;">☐</td></tr> <tr><td>f) Balloon or condom tamponade</td><td style="text-align: center;">☐</td></tr> <tr><td>g) Repair of cervical laceration</td><td style="text-align: center;">☐</td></tr> <tr><td>h) Repair of uterine rupture</td><td style="text-align: center;">☐</td></tr> <tr><td>i) B-lynch suture</td><td style="text-align: center;">☐</td></tr> <tr><td>j) Hysterectomy</td><td style="text-align: center;">☐</td></tr> </table> <i>Infection</i> <table style="width: 100%;"> <tr><td>k) Therapeutic antibiotics (excludes prophylaxis)</td><td style="text-align: center;">☐</td></tr> </table> <i>Hypertension</i> <table style="width: 100%;"> <tr><td>l) Magnesium sulphate as anticonvulsant</td><td style="text-align: center;">☐</td></tr> <tr><td>m) Other anticonvulsant for eclampsia</td><td style="text-align: center;">☐</td></tr> </table> <i>Other interventions</i> <table style="width: 100%;"> <tr><td>n) Removal of retained products</td><td style="text-align: center;">☐</td></tr> <tr><td>o) Manual removal of placenta</td><td style="text-align: center;">☐</td></tr> <tr><td>p) Blood transfusion</td><td style="text-align: center;">☐</td></tr> <tr><td>q) Laparotomy</td><td style="text-align: center;">☐</td></tr> <tr><td>r) Admission to Intensive Care Unit</td><td style="text-align: center;">☐</td></tr> </table> 80) Specify whether the woman was referred to any higher complexity hospital (0=No 1=Yes) ☐ 81) Maternal status at hospital discharge (0= Alive 1=Dead) ☐ 82) Date of maternal discharge, transfer or death <table style="width: 100%; text-align: center;"> <tr> <td>d</td><td>d</td><td>m</td><td>m</td><td>y</td><td>y</td> </tr> <tr> <td> </td><td> </td><td> </td><td> </td><td> </td><td> </td> </tr> </table> The use of "9" is not allowed in section F and Q74	a) Oxytocin for treatment of PPH	☐	b) Misoprostol for treatment of PPH	☐	c) Ergotamine for treatment of PPH	☐	d) Other uterotonic for treatment of PPH	☐	e) Artery ligation or embolization	☐	f) Balloon or condom tamponade	☐	g) Repair of cervical laceration	☐	h) Repair of uterine rupture	☐	i) B-lynch suture	☐	j) Hysterectomy	☐	k) Therapeutic antibiotics (excludes prophylaxis)	☐	l) Magnesium sulphate as anticonvulsant	☐	m) Other anticonvulsant for eclampsia	☐	n) Removal of retained products	☐	o) Manual removal of placenta	☐	p) Blood transfusion	☐	q) Laparotomy	☐	r) Admission to Intensive Care Unit	☐	d	d	m	m	y	y						
a) Oxytocin for treatment of PPH	☐																																																	
b) Misoprostol for treatment of PPH	☐																																																	
c) Ergotamine for treatment of PPH	☐																																																	
d) Other uterotonic for treatment of PPH	☐																																																	
e) Artery ligation or embolization	☐																																																	
f) Balloon or condom tamponade	☐																																																	
g) Repair of cervical laceration	☐																																																	
h) Repair of uterine rupture	☐																																																	
i) B-lynch suture	☐																																																	
j) Hysterectomy	☐																																																	
k) Therapeutic antibiotics (excludes prophylaxis)	☐																																																	
l) Magnesium sulphate as anticonvulsant	☐																																																	
m) Other anticonvulsant for eclampsia	☐																																																	
n) Removal of retained products	☐																																																	
o) Manual removal of placenta	☐																																																	
p) Blood transfusion	☐																																																	
q) Laparotomy	☐																																																	
r) Admission to Intensive Care Unit	☐																																																	
d	d	m	m	y	y																																													

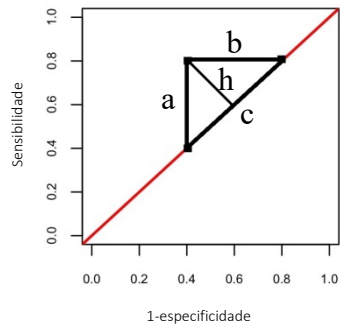
78) Was any of the following conditions identified? 0= No 1= Yes, within 24h of hospital stay 2= Yes, after 24h of hospital stay a) Cardiovascular dysfunction ☐ <i>Shock</i> <i>Cardiac Arrest</i> <i>Severe hypoperfusion (lactate >5 mmol/L or >45mg/dL)</i> <i>Severe acidosis (pH<7.1)</i> <i>Use of continuous vasoactive drugs</i> <i>Cardio-pulmonary resuscitation</i> b) Respiratory dysfunction ☐ <i>Acute cyanosis</i> <i>Gasping</i> <i>Severe tachypnea (respiratory rate>40 breaths per minute)</i> <i>Severe bradypnea (respiratory rate<6 breaths per minute)</i> <i>Severe hypoxemia</i> (O2 saturation <90% for ≥60min or PAO2/FiO2<200) <i>Intubation and ventilation not related to anaesthesia</i> c) Renal dysfunction ☐ <i>Oliguria non responsive to fluids or diuretics</i> <i>Severe acute azotaemia</i> (creatinine >300umol/ml or >3.5mg/dL) <i>Dialysis for acute renal failure</i> d) Coagulation dysfunction ☐ <i>Failure to form clots</i> <i>Severe acute thrombocytopenia (<50,000 platelets/ml)</i> <i>Massive transfusion of blood or red cells (≥ 5 units)</i> e) Hepatic dysfunction ☐ <i>Jaundice in the presence of pre-eclampsia</i> <i>Severe acute hyperbilirubinemia</i> (bilirubin>100umol/L or >6.0mg/dL) f) Neurologic dysfunction ☐ <i>Prolonged unconsciousness or coma (lasting >12 hours)</i> <i>Stroke or Global paralysis</i> <i>Uncontrollable fit / status epilepticus</i> g) Uterine dysfunction ☐ <i>Hysterectomy due to uterine infection or haemorrhage</i>	Data Collector Name: Date:
--	--

Anexo 3. A estatística J e o espaço ROC

	No Espaço ROC, a inclinação da linha de não-discriminação é dada por: $m = \frac{Y - Y_0}{X - X_0}$ $m = \frac{Y - 0}{X - 0}$ $m = \frac{Y}{X}$ $m = 1$ $\frac{Y}{X} = 1$ Na linha de não-discriminação, $X = Y$
	O ponto A representa o resultado de desempenho de um dado teste A. As coordenadas desse ponto são: ((1-Especificidade) , Sensibilidade) ou simplesmente: (x , y) $X = 1 - \text{especificidade}$ $Y = \text{sensibilidade}$
	Considere os pontos A, B, C



E suas coordenadas



$$a^2 = h^2 + \left(\frac{c}{2}\right)^2$$

$$h^2 = a^2 - \left(\frac{c}{2}\right)^2$$

$$h^2 = a^2 - \left(\frac{c^2}{4}\right)$$

$$h^2 = a^2 - \left(\frac{2a^2}{4}\right)$$

$$h^2 = a^2 - \left(\frac{a^2}{2}\right)$$

$$h^2 = \frac{a^2}{2}$$

$$h = \sqrt{\frac{a^2}{2}}$$

$$h = \sqrt{\frac{(\text{sensibilidade} - (1 - \text{especificidade}))^2}{2}}$$

$$h = \sqrt{\frac{(\text{sensibilidade} + \text{especificidade} - 1)^2}{2}}$$

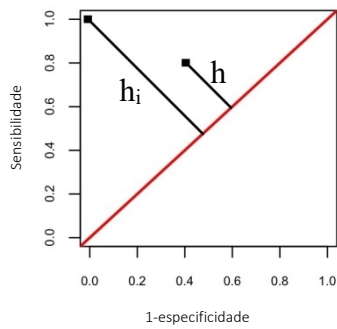
$$a = y - x$$

$$b = x - y$$

$$a^2 = b^2$$

$$c^2 = a^2 + b^2$$

$$c^2 = 2a^2$$



h_i é equivalente a metade da diagonal d um quadrado 1×1

$$d^2 = 1^2 + 1^2$$

$$d^2 = 2$$

$$d = \sqrt{2}$$

$$h_i = \frac{\sqrt{2}}{2}$$

Considerando que h é uma expressão de desempenho de um preditor e h_i é a expressão de desempenho de um teste perfeito, a razão entre h e h_i expressa a proporção ideal da desempenho diagnóstica alcançada por esse teste

$$\frac{h}{h_i} = \frac{\sqrt{\frac{(\text{sensibilidade} + \text{especificidade} - 1)^2}{2}}}{\frac{\sqrt{2}}{2}}$$

$$h/h_i = \text{sensibilidade} + \text{especificidade} - 1$$

A proporção h/h_i , quando escrita na forma:

$$(\text{sensibilidade} + \text{especificidade} - 1)$$

é equivalente a Estatística J.

Anexo 4. Definição dos preditores do partograma

Preditor	Regra para análise
Mecônio	Se a membrana está rompida com mecônio velho ou fresco, então o mecônio está presente
<i>Caput Succedaneum</i> significativo	Se o caput succedaneum é moderado ou severo, então é significativo
Temperatura materna anormal	Se a temperatura materna > 37,5 C ou < 35,5 C, então a temperatura é anormal
Amoldamento significativo	Se há amoldamento de segundo ou terceiro grau, então é significativo
Movimentos fetais não observados	Se nenhum movimento fetal foi observado ou sentido
Batimento cardíaco fetal anormal	Se o batimento cardíaco fetal é < 110 BPM ou > 160 BPM, então ele é anormal
Descida lenta	Se a descida é mais lenta que em 90% das mulheres presentes no banco de dados, é considerado que a descida é lenta Limite estabelecido a partir de um Modelo de Markov Multi-estado
Duração sub-ótima das contrações uterinas	Se a duração das contrações uterinas é > 60s ou Se o trabalho de parto está lento e a duração das contrações uterinas é < 40s, então a duração das contrações uterinas é sub-ótima
Pressão arterial sistólica anormal	Se a pressão arterial sistólica é > 159 mmHg ou < 85 mmHg, a pressão é anormal
Posição anormal da cabeça do feto	Se, no último exame vaginal, a posição da cabeça do feto é occipital-transversa, occipital-posterior ou uma posição diferente de occipital-anterior, então a posição da cabeça do feto é anormal
Batimento cardíaco materno anormal	Se o batimento cardíaco materno > 119 BPM ou < 60 BPM, então o batimento é anormal
Número sub-ótimo de contrações uterinas	Se o número de contrações uterinas em 10 minutos > 4 ou Se o trabalho de parto está lento e o número de contrações uterinas em 10 minutos < 3, então o número de contrações é sub-ótimo
Pressão arterial diastólica anormal	Se a pressão arterial diastólica > 105 mmHg, então é anormal

Anexo 5. Artigo publicado



DOI: 10.1111/1471-0528.15205
www.bjog.org

Cervical dilatation over time is a poor predictor of severe adverse birth outcomes: a diagnostic accuracy study

JP Souza,^a OT Oladapo,^a B Fawole,^b K Mugerwa,^c R Reis,^d F Barbosa-Junior,^d L Oliveira-Ciabati,^d D Alves,^d AM Gülmezoglu^a

^a UNDP/UNFPA/ UNICEF/WHO/World Bank Special Programme of Research Development and Research Training in Human Reproduction (IIRP), Department of Reproductive Health and Research, World Health Organization, Geneva, Switzerland ^b Department of Obstetrics and Gynaecology, College of Medicine, University of Ibadan, Ibadan, Nigeria ^c Department of Obstetrics and Gynaecology, Makerere University, Kampala, Uganda ^d Department of Social Medicine, Centre for Information and Informatics for Health (CIIS), Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil

Correspondence: JP Souza, UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Reproductive Health and Research, World Health Organization, Avenue Appia 20, 1202 Geneva, Switzerland. Email: jp.souza@usp.br

Accepted 13 February 2018. Published Online 17 April 2018.

Objective To assess the accuracy of the World Health Organization (WHO) partograph alert line and other candidate predictors in the identification of women at risk of developing severe adverse birth outcomes.

Design A facility-based, multicentre, prospective cohort study.

Setting Thirteen maternity hospitals located in Nigeria and Uganda.

Population A total of 9995 women with spontaneous onset of labour presenting at cervical dilatation of ≤ 6 cm or undergoing induction of labour.

Methods Research assistants collected data on sociodemographic, anthropometric, obstetric, and medical characteristics of study participants at hospital admission, multiple assessments during labour, and interventions during labour and childbirth. The alert line and action line, intrapartum monitoring parameters, and customised labour curves were assessed using sensitivity, specificity, positive and negative likelihood ratios, diagnostic odds ratio, and the f statistic.

Outcomes Severe adverse birth outcomes.

Results The rate of severe adverse birth outcomes was 2.2% (223 women with severe adverse birth outcomes), the rate of augmentation of labour was 35.1% (3506 women), and the

caesarean section rate was 13.2% (1323 women). Forty-nine percent of women in labour crossed the alert line (4163/8489). All reference labour curves had a diagnostic odds ratio ranging from 1.29 to 1.60. The f statistic was less than 10% for all reference curves.

Conclusions Our findings suggest that labour is an extremely variable phenomenon, and the assessment of cervical dilatation over time is a poor predictor of severe adverse birth outcomes. The validity of a partograph alert line based on the 'one-centimetre per hour' rule should be re-evaluated.

Funding Bill & Melinda Gates Foundation, United States Agency for International Development (USAID), UNDP/UNFPA/ UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (IIRP), and WHO (A65879).

Keywords alert line, childbirth, diagnostic accuracy, partograph, receiver operating characteristic space.

Tweetable abstract The alert line in check: results from a WHO study.

Linked article This article is commented on by F Okonofua. To view this article visit <https://doi.org/10.1111/1471-0528.15213>

Please cite this paper as: Souza JP, Oladapo OT, Fawole B, Mugerwa K, Reis R, Barbosa-Junior F, Oliveira-Ciabati L, Alves D, Gülmezoglu AM. Cervical dilatation over time is a poor predictor of severe adverse birth outcomes: a diagnostic accuracy study. BJOG 2018; <https://doi.org/10.1111/1471-0528.15205>.

Introduction

Labour and childbirth are natural processes with a relatively low frequency of complications among healthy pregnant women.^{1,2} Intrapartum maternal and fetal monitoring

is used to further minimise risks, and is expected to enable the early identification and prompt treatment of complications. The assessment of cervical dilatation is part of intrapartum monitoring, and is conducted by healthcare providers to determine the adequacy of labour progress.

© 2018 World Health Organization; licensed by John Wiley & Sons Ltd on behalf of Royal College of Obstetricians and Gynaecologists. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution and reproduction in any medium, provided that the original work is properly cited. In any reproduction of this article there should not be any suggestion that WHO or the article endorse any specific organization or products. The use of the WHO logo is not permitted. This notice should be preserved along with the article's URL.

The observed cervical dilatation is usually compared with reference labour curves to estimate the risk of labour complications and to guide the use of interventions.^{3–7}

The World Health Organization (WHO) partograph is a decision-making support tool designed to assist health providers in identifying women at risk of developing complications during labour, and to guide the use of interventions intended to mitigate any perceived risks.⁶ With the partograph, the identification of certain patterns of cervical dilatation or other risk factors may prompt the transfer of the woman to a higher-level health facility, the intensification of intrapartum monitoring, the augmentation of labour, or delivery by caesarean section.⁷ The WHO partograph is based on clinical principles, including the notion that 'normal' labour progress is defined by a cervical dilatation rate of not less than one centimetre per hour between 4 and 10 centimetres of cervical dilatation.⁶ This concept is the basis for the partograph 'alert line', which was derived from average labour curves developed during the 1950s and 1960s.^{8,9} Alternative and more recent labour curves have been developed to provide a reference for labour progress and to serve as the basis for new partograph designs.^{5,10}

Although some observational studies and other empirical evidence point towards the benefit of using the WHO partograph, experimental, head-to-head comparisons failed to demonstrate an effect of the partograph in improving health outcomes related to labour and childbirth.¹¹ Furthermore, different studies have pointed to limitations in the 'one centimetre per hour rule' as a valid benchmark for assessing the adequacy of labour progress.^{5,10,12} Our hypothesis is that if the partograph is unable to accurately identify women at risk of developing intrapartum complications, it will not be able to effectively guide labour management.

This article reports on findings of the WHO Better Outcomes in Labour Difficulty (BOLD) project. The present analysis assessed the diagnostic accuracy of the alert line, action line, and other parameters included in the WHO partograph as predictors of severe adverse birth outcomes. It also assessed the accuracy of customised labour curves to identify women at risk of developing severe adverse birth outcomes.

Methods

The BOLD project included quantitative, qualitative, and service-design research conducted in Nigeria and Uganda. The methodological details of the BOLD project have been described elsewhere.^{13,14} This analysis is based on the quantitative component, a facility-based, multicentre, prospective cohort study. In brief, this study included women admitted for vaginal birth with single live fetuses during the early first stage of labour across 13 hospitals in both

countries. Women with spontaneous onset of labour presenting at cervical dilatation of ≤ 6 cm and women undergoing induction of labour took part in the study. Women with multiple pregnancies, women with pregnancies with gestational ages of less than 34 weeks 0 days, women choosing elective caesarean section, and women who were incapable of giving consent because of labour distress or obstetric emergencies at arrival were excluded. Participating institutions had a minimum of 1000 deliveries per year, with stable access to caesarean section, augmentation of labour, and assisted vaginal birth. Midwives, obstetricians, or obstetric residents provided intrapartum health care to women in labour. Dopplers were used to assess fetal vital status at hospital admission and for intermittent monitoring through labour and childbirth. Labour management protocol, as well as the number and timing of pelvic examinations, were not standardised across participating institutions. None of the institutions subscribed to the active management of labour protocol during the study period. Although the partograph was a standard element of medical records in all participating health facilities, its prospective application to guide labour management during the study period varied widely across the hospitals.

Eligible women were recruited into the study between December 2014 and November 2015. From the medical records, trained research nurses prospectively extracted detailed information on the sociodemographic, anthropometric, obstetric, and medical characteristics of the study participants at hospital admission, multiple assessments during labour monitoring, interventions performed throughout the first and second stages of labour, and maternal and neonatal labour outcomes. Attending staff were approached to complement medical records data when needed. Data collection was limited to hospital stay of the mother and baby, and there was no post-hospital discharge follow-up.

The current analysis was based on information on maternal baseline and admission characteristics, repeated assessments of cervical dilatation versus time, and maternal and neonatal outcome data. Severe adverse birth outcomes were defined as the occurrence of any of the following: stillbirths, intra-hospital early neonatal deaths, neonatal use of anticonvulsants, neonatal cardiopulmonary resuscitation, Apgar score of ≤ 6 at 5 minutes, uterine rupture, and maternal death or organ dysfunction with dystocia. Details of the sample size calculation are provided in the supporting information (Box S1).

Data analysis

Simple frequencies and proportions were used to describe the characteristics of the study population. Sensitivity, specificity, positive and negative likelihood ratios, diagnostic odds ratios, and the *J* statistic (Youden's index),

with 95% confidence intervals, were used to estimate the diagnostic accuracy of the alert line and the action line in the identification of women who would develop a severe adverse birth outcome.^{15–18} We used the true-positive rate (i.e. sensitivity) and the false-positive rate (i.e. 1 – specificity) to graphically represent the diagnostic accuracy of the partograph parameters in the receiver operating characteristic (ROC) space.¹⁹ Each point estimate in the ROC space represents a classification result for binary parameters, and the interpretation of the ROC space is similar to the ROC curve: optimal results are associated with high true-positive rates combined with low false-positive rates. The *J* statistic summarises the performance of a binary classifier,¹⁶ and also expresses the proportion of ideal performance of a diagnostic test (Box S2). The supporting information provides additional details related to the calculation and interpretation of these statistics (Tables S1–S3).

The alert line and the action line are classifiers currently applied to all women, regardless of their obstetric characteristics (e.g. nulliparous, multiparous, spontaneous or induced labour, or previous caesarean section). We hypothesised that cervical dilatation curves customised according to the obstetric characteristics of the population could have a better accuracy than the generic alert and action lines. The study population was stratified into mutually exclusive, totally inclusive obstetric groups according to the 10-group Robson classification:²⁰ group 1 (nulliparous, single cephalic pregnancy, 37 weeks of gestation or more, with spontaneous onset of labour); group 2 (nulliparous women, single cephalic pregnancy, 37 weeks of gestation or more, with induced onset of labour); group 3 (multiparous women without previous caesarean section, with single cephalic pregnancy, 37 weeks of gestation or more, with spontaneous onset of labour); group 4 (multiparous women without previous caesarean section, with single cephalic pregnancy, 37 weeks of gestation or more, with induced onset of labour); group 5 (all multiparous women with at least one previous caesarean section, single cephalic pregnancy, at 37 weeks of gestation or more); and group 10 (all women with singleton cephalic preterm pregnancy at less than 37 weeks of gestation at childbirth). As a result of the eligibility criteria, this study has no women from group 8 (multiple pregnancies) or with caesarean section before labour. Women with non-cephalic presentations (groups 6, 7, and 9) were grouped together. Groups 1–5 and 10, were further divided according to the use of augmentation of labour (present or absent), totalling 12 subgroups. Using data from women who did not have any severe adverse birth outcome, customised labour curves were generated for each of these 12 subgroups. Data from women pertaining to groups 6, 7, and 9 were not used to generate customised curves because of the small numbers involved. The customised cervical dilatation curves were created using a multi-state Markov

model,^{21,22} which represented the cervical dilation pattern through intermediate states from 2 cm to 10 cm, and childbirth by selected percentiles and obstetric group (i.e. one labour curve for each obstetric group and selected percentile). In this model, each centimetre of cervical dilatation represented an intermediate state, and childbirth was the final ‘absorbing’ state. The model was generated as a progressive unidirectional labour-to-childbirth model, and the time of state change was determined by a set of transition intensities. The transition intensity represents the instantaneous likelihood of moving from one state to another, and is generated as part of the multi-state Markov model. For each one of the 12 obstetric subgroups, the multi-state Markov model generated labour curves representing the progress of labour in women that was either faster or at the 50th, 60th, 70th, 80th, 90th, and 95th percentiles.

Once the percentile curves were generated for each obstetric subgroup of women without severe adverse birth outcomes, women were classified as having crossed or not having crossed each of the percentile curves of their relevant obstetric subgroup. The study population was then consolidated and all women who crossed their relevant 50th percentile curves were grouped together (i.e. women in which labour progressed more slowly than the customised 50th percentile curve). Similarly, women were classified as having labour that progressed either slower or faster/equal to the relevant 60th, 70th, 80th, 90th, and 95th percentiles. We estimated the accuracies of the customised percentile curves in the identification of women who would develop a severe adverse birth outcome, by comparing women with labour progress that was slower than the specific percentile with women in which labour progressed faster or equal to that percentile. Sensitivity, specificity, positive and negative likelihood ratios, diagnostic odds ratios, with 95% confidence intervals, the *J* statistic, and ROC space plotting were used to estimate the accuracy of the percentile curves in the identification of women who would develop a severe adverse birth outcome.

Statistical analyses were carried in R and Microsoft EXCEL (2010).²³

Results

The analysis flow is shown in Figure 1. Thirteen hospitals (nine from Nigeria and four from Uganda) and a total of 9995 women (4964 from Nigeria and 5031 from Uganda) took part in this study. The average age of the participants was 27.9 years (± 5.0 years); 3.2% were younger than 20 years of age (320 women) and 11.0% were 35 years old or older (1100 women). The majority of the participants had a partner (97.6%, 9753 women); 5.2% of the participants had either incomplete primary education or no education (525 women), 5.6% had complete primary

education (564 women), 42.5% had either complete or incomplete secondary education (4245 women), and 45.4% had either complete or incomplete post-secondary/tertiary education (4537 women). A total of 4076 nulliparous women took part in the study (40.8%), and among women with at least one previous birth (59.2%, 5919 women), 535 (5.4%) had had a previous caesarean section. A total of 667 women (6.7%) had no antenatal care visit, 4229 (42.3%) had between one and three antenatal care visits, and 5007 (50.6%) had four visits or more; 1228 women (12.3%) developed pre-labour complications during the current pregnancy. The majority of women initiated labour spontaneously (8984 women, 89.9%) at between 37 and 41 weeks of gestation (91.2%, 9111 women), with only 594 (5.9%) being referred from another health facility during labour. All women participating in this study had singleton pregnancies, 98.6% (9845 women) of which were in cephalic presentation. The mean number of cervical assessments between 4 and 10 cm was 2.22 (± 1.02). Augmentation of labour was used in 3506 women (35.1%). Pharmacological analgesia was rarely used (2.0%, 196 women). Table 1 presents the distribution of the study population according to the 10-group Robson classification. The overall intrapartum caesarean section rate was 13.2% (1323 women), and the rate of severe adverse birth outcomes was 2.2% (223 women with severe adverse birth outcomes; Table S4).

Nearly half of women with at least two assessments of cervical dilatation between 4 cm and childbirth crossed the alert line (49%; 4163/8489). Figure 2 illustrates the progress of labour in the study population. In the upper panel, each grey line represents the progress of an individual women without severe adverse birth outcomes, and each red line represents the progress of an individual woman with adverse outcomes. In the lower panel, the labour curves for women in the 95th percentile, without augmentation of labour, is displayed by obstetric group. Video S1 displays an animation of labour progress of all women in labour that reached at least 4 cm of cervical dilatation.

The sensitivity, specificity, positive and negative likelihood ratios, diagnostic odds ratio, and *J* statistic for the WHO partograph alert and action lines, and the 50, 60, 70, 80, 90, and 95th percentiles are presented in Tables 1 and S5. For all reference curves, women who crossed the curves tended to show a mild increase in the odds of severe adverse birth outcomes, when compared with women who did not cross the reference lines. All reference curves had a diagnostic odds ratio ranging from 1.29 to 1.60. All reference curves had positive likelihood ratios smaller than 1.5 and negative likelihood ratios greater than 0.85. The *J* statistic was less than 10% for all reference curves. Figure 3 presents the ROC space analysis, with all the aforementioned predictors showing a poor diagnostic performance.

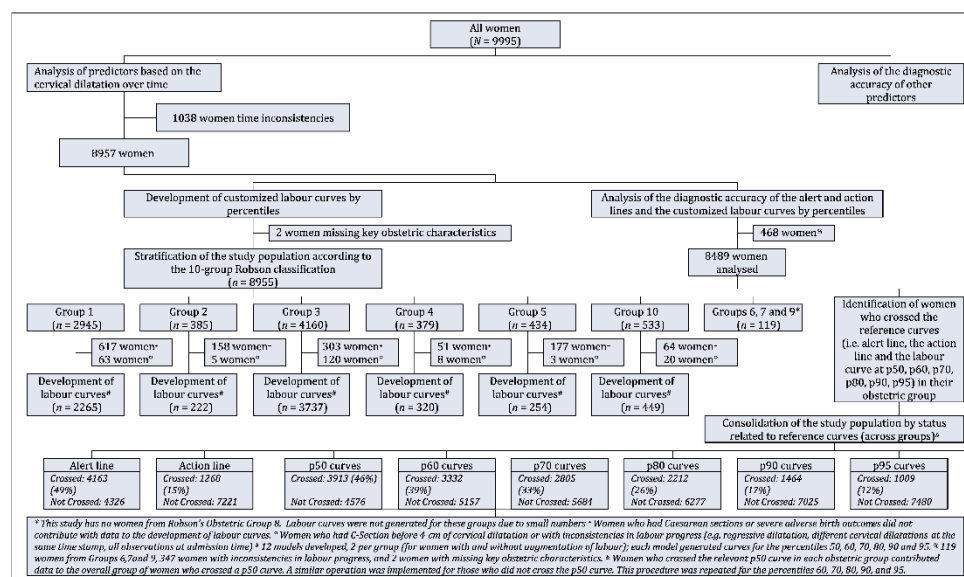


Figure 1. The analysis flowchart.

Table 1. Diagnostic accuracy of the alert line and the action line for severe adverse birth outcomes ($n = 8603$)

Clinical signs	Severe adverse birth outcomes		Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)	Diagnostic odds ratio (95% CI)	J statistic (Youden's index) (95% CI)
	Present	Absent						
Alert line	Crossed	110 4053	56.7% (49.7–63.5)	51.1% (50.1–52.2%)	1.16 (1.02–1.32)	0.85 (0.72–1.00)	1.37 (1.03–1.83)	7.8% (0.8–14.9)
	Not crossed	84 4242						
Action line	Crossed	38 1230	19.6% (14.6–25.7)	85.2% (84.4–85.9)	1.32 (0.99–1.77)	0.94 (0.88–1.01)	1.40 (0.98–2.01)	4.8% (0.9 to 10.4)
	Not crossed	156 7065						

Figure S1 and Table S6 present the diagnostic accuracy of various predictors included in the partograph (the definitions of these predictors are presented in Table S7). Abnormal fetal heart rate, absence of fetal movements, significant moulding, significant caput succedaneum, meconium, and maternal hyperthermia (fever) were associated with mild to moderate increased odds of severe adverse birth outcomes. Similarly to the labour curves, the examined predictors presented poor performance in the prediction of severe adverse birth outcomes.

Discussion

Main findings

Labour is an extremely variable phenomenon, and our findings suggest that the assessment of cervical dilatation over time is a poor predictor of severe adverse birth outcomes. Labour curves depicting the cervical dilatation over time (including the WHO partograph alert and action lines) showed poor diagnostic accuracy to identify women at risk of severe adverse birth outcomes during labour. We draw one main inference from these findings: the validity of a partograph alert line based on the 'one-centimetre per hour' rule should be re-evaluated.

Strengths and limitations

These findings are relevant to the care provided in health facilities, particularly in sub-Saharan Africa, and have potential implications for clinical practice. Despite the procedures adopted to ensure appropriate study implementation and high quality data, some limitations need to be considered, however. The primary data source in this study was routine hospital records, complemented by information obtained from clinical staff. We opted for this approach to minimise any interference with the standard practice in health facilities, but acknowledge that it could be associated with irregular and, at times, incomplete and intermittent assessment and recording of maternal and fetal status during labour. Although unlikely given the clinical workload, the availability of Doptonics provided by the study in the labour wards may have facilitated fetal monitoring and contributed to an increased identification of fetal distress, which could have affected the clinical management and outcomes. We were also able to determine the fetal vital status at arrival for all women, which resulted in an accurate assessment of intrapartum, intra-hospital fetal mortality. This assessment enabled the disentangling of pre-hospital fetal deaths from intra-hospital fetal deaths, and uncovered a low rate of intra-hospital fetal mortality, despite the constraints to optimal care in health facilities. None of the participating hospitals subscribed to a systematic implementation of the active management of labour; although this could contribute to a less standardised

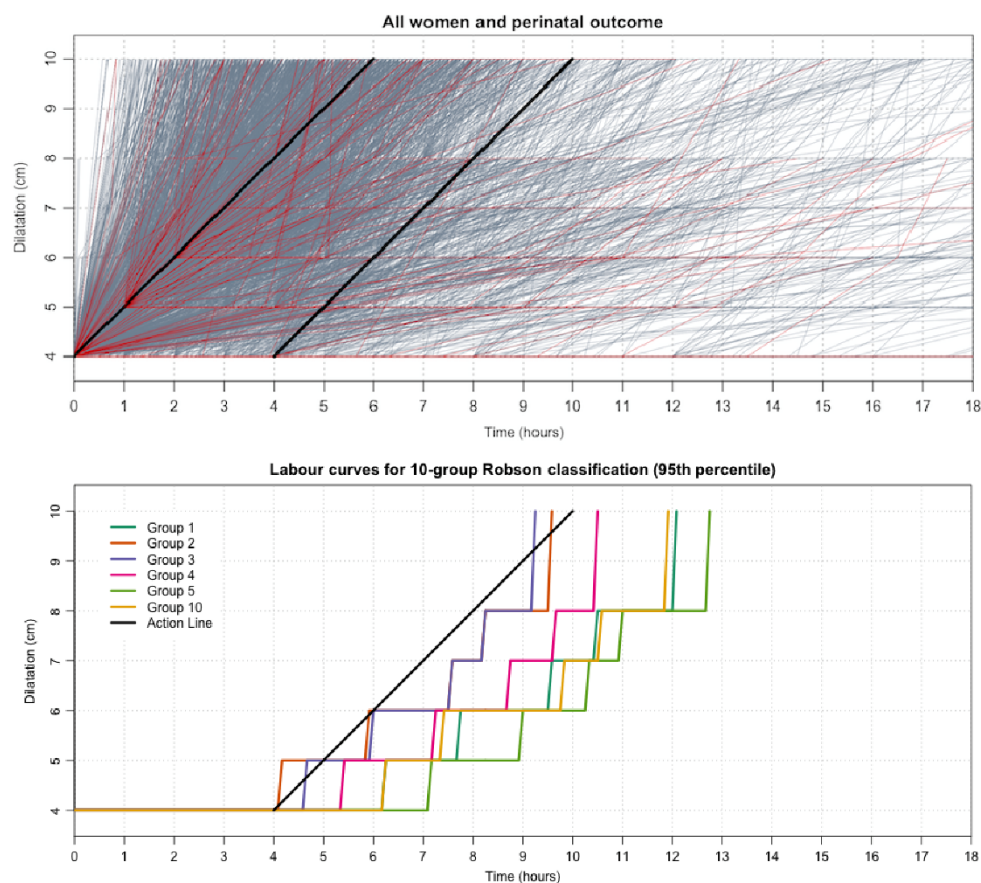


Figure 2. Upper panel: cervical dilatation over time (all women with at least two cervical dilatation assessments between 4 cm and childbirth). Grey lines denote labour progress of women without severe adverse birth outcomes; red lines denote labour progress of women with severe adverse birth outcomes. Lower panel: labour curves for selected groups of the 10-group Robson classification (95th percentile, women without augmentation of labour).

management of labour, it favoured a less interventionist approach and an intra-hospital labour progression that was more closely related to the natural progression in many women. Given the differences of workload and health-facility protocols, the standardisation of intrapartum maternal fetal monitoring and recording was a challenging task. Several mechanisms were used to minimise methodological heterogeneity and to increase the quality of the data as much as possible (such as research assistant training, the use of a visual check of the data collection forms before data entry, automated queries,

double-checking of selected medical records, and a thorough audit of unclear cases, especially those resulting in mortality). It should also be considered that crossing the alert or action lines could have prompted health providers to implement interventions in the current cohort population. These interventions could have modified the final outcome, either for good or bad.

Interpretation

Health facilities in low-resource settings often struggle with a shortage of human resources and life-saving commodities,

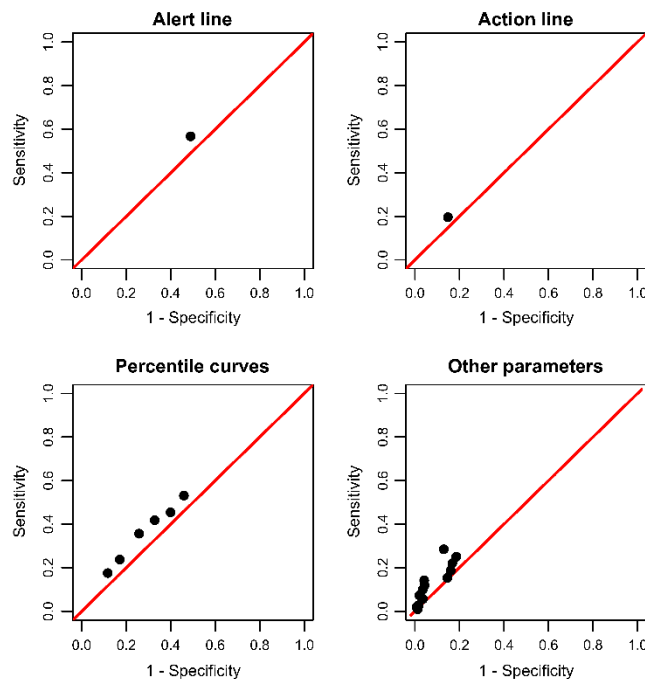


Figure 3. Analysis of the ROC space (alert and action line, customised percentile curves, and other parameters included in the partograph).

training resources, and health infrastructure, which limit the early identification and effective management of labour complications. Conversely, any overestimation of the risk of complications and over-medicalisation of care during labour and childbirth may lead to iatrogenic complications, avoidable suffering, and a waste of limited resources.²⁴ In an attempt to optimise intrapartum care, several organisations recommend the use of the WHO partograph to guide labour monitoring and management.²⁴ The ‘one-centimetre per hour rule’, as illustrated by the partograph alert line, has also (formally or informally) been used to prompt labour interventions in many settings around the world.^{4,25} Global efforts to promote the use of the partograph in the last three decades have been met with mixed results. Although most healthcare providers working in maternity settings know about the partograph, it is frequently used retrospectively for recording purposes instead of providing prospective support for clinical decision making. Possible reasons for these shortcomings include difficulties in its use and interpretation.^{26,27} Our findings suggest that the poor predictive performance of the partograph – and the consequent

effect in supporting effective decision making – could contribute to the lack of interest in using the tool prospectively.

As countries navigate through the obstetric transition,²⁸ a marked trend towards the medicalisation of labour and childbirth is observed. Several determinants of the medicalisation of labour and childbirth are at play, including models of care based on the notion that a normal labour abides by the ‘one-centimetre per hour’ rule. This notion has been embedded in generations of healthcare providers across the world, and the implicit or explicit influence of this notion in obstetric and midwifery culture cannot be over-emphasised; however, as suggested by our findings, a cervical dilatation rate of ‘one-centimetre per hour’ may be unrealistically fast for a substantial proportion of women in labour. The mismatch between the unrealistic expectations of healthcare providers and the physiology of labour may give rise to the constructed ‘need’ for an intervention in a natural process that could otherwise be slower than currently expected but end well and naturally. The poor accuracy of the tool means on one hand that a high proportion of women would receive an intervention without a valid justification, and on the other

hand women at risk would not be recognised in time to avoid the adverse outcome. The excessive use of interventions may also contribute to adverse outcomes. For example, the augmentation of labour is a well-established risk factor for fetal distress; unnecessary augmentation of labour, prompted by the 'one-centimetre per hour' rule, may be harmful, particularly in settings with limited capacity for providing appropriate, intermittent fetal monitoring. Another potential adverse effect of the above mismatch is increased tension, anxiety, and frustration among the staff, which could be a contributing factor to disrespect, abuse, and mistreatment of women during labour and childbirth. Allowing an increase in the average duration of labour in health facilities has a direct impact on the occupancy rate of labour-ward beds, however, which could further complicate the shortage of hospital beds and overcrowding of health facilities. Reducing the number of interventions during labour could reduce staff workload. Research to determine the short- and long-term consequences of a less invasive intrapartum care model at the individual and at the health-systems level is warranted.

The poor performance of the partograph and the customised labour curves may not be a surprising finding. The rationale for using the partograph for preventing labour problems goes back several decades, when it was introduced for the timely referral from peripheral health facilities to prevent the complications of obstructed labour.⁹ Fetal and early neonatal outcomes are much more likely to be impacted by events that are not related to the cervical dilatation rate, such as placental abruption, cord compression, cord prolapse, meconium aspiration, and intrauterine growth restriction, among many other reasons. In South Africa, for example, only 6% of fetal and early neonatal deaths were associated with prolonged labour;²⁹ however, we should not overlook the finding that slower labours compared with faster labours (in different percentiles) tended to be associated with a mild increase in the risk of adverse outcomes. Nevertheless, this association alone can hardly provide a basis for a reliable classification tool because of the excessive number of false positives. For example, nearly half of the study population crossed the alert line, making the policy of transferring women who crossed the alert line to referral hospitals impractical. In this context, and given the limitations of static, paper-based diagnostic tools, the development and testing of more sophisticated, dynamic, easy-to-use tools for improved risk classification during labour is a priority. A cluster-randomised trial, comparing a static paper-based partograph with a dynamic, multivariable prediction model would be ideal research to be carried out next.

Conclusion

Our findings suggest that the validity of a partograph alert line based on the 'one-centimetre per hour' rule should be

re-evaluated. Labour is an extremely variable phenomenon, and emphasis should be given to individualised, supportive, person-centered care during labour and childbirth.

Disclosure of interests

None declared. Completed disclosure of interests form available to view online as supporting information.

Contribution to authorship

JPS, OTO, and AMG developed the research protocol, with input from members of the project steering committee and advisory group, Nigeria and Uganda research teams, data science team, and the project coordination and support team. The analysis plan was developed by JPS with input from OTO. Data analysis was carried out by RR, FB, LOC, and JPS. JPS drafted this article, with substantial contributions from OTO, BF, KM, RR, FB, LOC, DA, and AMG. All authors reviewed the draft manuscript for intellectual content and approved the final version for publication.

Details of ethics approval

Scientific and technical approval was obtained from the Review Panel on Research Projects (RP2) of UNDP/UNFPA/UNICEF/WHO/World Bank Special Program of Research, Development and Research Training in Human Reproduction (HRP). Ethical approval was obtained from the World Health Organization Ethical Review Committee (protocol A65879, approval date 25 August 2014), the Makerere University School of Health Sciences Research and Ethics Committee, Uganda (protocol #SHSREC REF 2014-058), University of Ibadan/University College Hospital Ethics Committee (UI/EC/14/0223), Federal Capital Territory Health Research Ethics Committee, Nigeria (protocol FHREC/2014/01/42/27-08-14), and Ondo State Government Ministry of Health Research Ethics Review Committee, Nigeria (AD 4693/160).

Acknowledgements

This manuscript was developed as part of the World Health Organization's BOLD (Better Outcomes in Labour Difficulty) project. The BOLD project implementation was a collaborative effort of a large number of academic staff, hospital personnel and researchers from 13 hospitals across Uganda and Nigeria. We acknowledge the contributions of all midwives, doctors, facility administrators, and all maternity unit staff in the maternity units of all hospitals that participated in the project. The authors would like to thank the members of the following groups for their input during project implementation and analysis: the BOLD project steering committee and advisory group (Priya Agrawal, Fernando Althabe, Deborah Armbruster, France Donnay, G Justus Hofmeyr, Vanora Hundley, Jen Jardine, David Kyadondo, Tina Lavender, Simon Lewin, Jerker Liljestrand,

Matthews Mathai, Lanre Olutayo, Janna Patterson, Robert Pattinson, Kate Ramsey, Harshadkumar C Sanghvi, Mary Ellen Stanton, Petra Tenhoope-Bender, Mari Tikkanen, Melanie Wendland, Jim Zhang); the Nigeria research team (Amos Adebayo, Nike Adesina, Adesina Akintan, Olu-bunmi Alabi, Francis E Alu, Halima Bello, Felicia Fasipe, Bukola Fawole, Hadiza Idris, Oluwatosin Lawal, Ola Okike, Lawal Oyeneyin, and Sanni Wilfred); the Uganda research team (Salim Bataale, Michael Bukennya, Josaphat Byamugisha, Wasswa Damien, Kidza Mugerwa, Othiniel Musana, and Miriam Nakalembe); the data science team (Domingos Alves, Francisco Barbosa-Júnior, Gleise Kele Bernucci, Hayala Cristina Cavenague, Livia Ciabati, Alexandre Cristóvão Maiorano, Jéssica Marcelino, Leonardo Moscovici, Lariza Laura de Oliveira, Gleici Perdoná, Rodrigo Reis, Tiago Lara Michelin Sanches, and Joao Paulo Souza); and the project coordination and support team (Lily Zilla Atutornu, Rajiv Bahl, Meghan Bohren, Metin Gülmezoglu, Catherine Cécile Kiener, Olufemi T Oladapo, Joao Paulo Souza, Özge Tunçalp, and Joshua P Vogel). The manuscript presents the views of the named authors only.

Funding

Bill & Melinda Gates Foundation, United States Agency for International Development (USAID), UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), and WHO (A65879). The external funders played no role in conducting research and writing this article.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. The analysis flowchart.

Table S1. Calculation of diagnostic accuracy statistics.

Table S2. Equivalence of diagnostic accuracy statistics.

Table S3. Suggested interpretation of diagnostic accuracy statistics.

Table S4. Distribution of the study population according to obstetric groups and the frequency of severe adverse birth outcomes.

Table S5. Diagnostic accuracy of customized labour curves by percentiles.

Table S6. Diagnostic accuracy of other predictors included in the partograph.

Table S7. Definitions of the predictors included in the Table S6 and Figure S1.

Box S1. Sample Size calculation.

Box S2. The J statistic in the ROC space

Video S1. All women and perinatal outcome ■

References

- Snowden JM, Tilden EL, Snyder J, Quigley B, Caughey AB, Cheng YW. Planned Out-of-Hospital Birth and Birth Outcomes. *N Engl J Med* 2015;373:2642–53.
- de Jonge A, Geerts CC, van der Goes BY, Mol BW, Buitendijk SE, Nijhuis JG. Perinatal mortality and morbidity up to 28 days after birth among 743 070 low-risk planned home and hospital births: a cohort study based on three merged national perinatal databases. *BJOG* 2015;122:720–8.
- Cohen WR, Friedman EA. The assessment of labor: a brief history. *J Perinat Med* 2018;46:1–8. <https://doi.org/10.1515/jpm-2017-0018>
- O'Driscoll K, Stronge JM. Active management of labor. *BMJ* 1973;3:590.
- Zhang J, Landy HJ, Branch DW, Burkman R, Haberman S, Gregory KD, et al. Contemporary patterns of spontaneous labor with normal neonatal outcomes. *Obstet Gynecol* 2010;116:1281–7.
- World Health Organization. *Preventing Prolonged Labour: A Practical Guide. The Partograph Part I. Principle and Strategy*. Geneva, Switzerland: WHO, 1993a. WHO/FHE/MSM 93.8
- World Health Organization. *Managing Complications in Pregnancy and Childbirth: A Guide for Midwives and Doctors*. IMPAC. WHO/RHR.007. Geneva, Switzerland: World Health Organization, 2000.
- Friedman E. The graphic analysis of labor. *Am J Obstet Gynecol* 1954;68:1568–75.
- Philpott RH, Castle WM. Cervicographs in the management of labour in primigravidae. I. The alert line for detecting abnormal labour. *J Obstet Gynaecol Br Commonw* 1972;79:592–8.
- Neal JL, Lowe NK. Physiologic partograph to improve birth safety and outcomes among low-risk, nulliparous women with spontaneous labor onset. *Med Hypotheses* 2012;78:319–26.
- Lavender T, Hart A, Smyth RM. Effect of partogram use on outcomes for women in spontaneous labour at term. *Cochrane Database Syst Rev* 2013;7:CD005461.
- Hamilton EF, Warrick PA, Collins K, Smith S, Garite TJ. Assessing first-stage labor progression and its relationship to complications. *Am J Obstet Gynecol* 2016;214:358.e1–8.
- Souza JP, Oladapo OT, Bohren MA, Mugerwa K, Fawole B, Moscovici L, et al. The development of a Simplified, Effective, Labour Monitoring-to-Action (SELMA) tool for Better Outcomes in Labour Difficulty (BOLD): study protocol. *Reprod Health* 2015;12:49.
- Bohren MA, Oladapo OT, Tunçalp Ö, Wendland M, Vogel JP, Tikkanen M, et al. Formative research and development of innovative tools for “Better Outcomes in Labour Difficulty” (BOLD): study protocol. *Reprod Health* 2015;26:50.
- Powers DM. Evaluation: from precision, recall and f-measure to ROC, informedness, markedness & correlation. *J Mach Learn Technol* 2011;2:37–63.
- Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.
- Glas AS, Uijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003;56:1129–35.
- Cook C, Cleland J, Huijbregts P. Creation and critique of studies of diagnostic accuracy: use of the STARD and QUADAS methodological quality assessment tools. *J Man Manip Ther* 2007;15:93–102.
- Fawcett T. An Introduction to ROC Analysis. *Pattern Recogn Lett* 2006;27:861–74.
- Robson MS. Classification of caesarean section. *Fetal Maternal Med Rev* 2001;12:23–39.
- Jackson CH. Multi-state models for panel data: The msm package for R. *J Stat Softw* 2011;38:1–29.
- Reis R. Labour Curves and Markov models. MSc [Dissertation]. Ribeirão Preto, Brazil: University of São Paulo, 2017.

- 23 R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2014 (<http://www.R-project.org>)
- 24 Miller S, Abalos E, Chamillard M, Ciapponi A, Colaci D, Comandé D, et al. Beyond too little, too late and too much, too soon: a pathway towards evidence-based, respectful maternity care worldwide. *Lancet* 2016;388:2176–92.
- 25 Maffi I. The detour of an obstetric technology: active management of labor across cultures. *Med Anthropol* 2016;35:17–30.
- 26 Bedwell C, Levin K, Pett C, Lavender DT. A realist review of the partograph: when and how does it work for labour monitoring? *BMC Pregnancy Childbirth* 2017;17:31.
- 27 Ollerhead E, Osrin D. Barriers to and incentives for achieving partograph use in obstetric practice in low- and middle-income countries: a systematic review. *BMC Pregnancy Childbirth* 2014;16:281.
- 28 Souza JP, Tunçalp Ö, Vogel JP, Bohren M, Widmer M, Oladapo OT, et al. Obstetric transition: the pathway towards ending preventable maternal deaths. *BJOG* 2014;121(Suppl 1):1–4.
- 29 Saving babies 2012-2013: Ninth report on perinatal care in South Africa. Pretoria, South Africa: RC Pattinson, N Rhoda, Tshepesa Press, 2014

