Standardization of Sonographic Prognostication of Prenatally Diagnosed Congenital Diaphragmatic Hernia (CDH) Across the North American Fetal Therapy Network (NAFTNet)

by

Nimrah Abbasi

A thesis submitted in conformity with the requirements for the degree of Master of Science

Institute of Health Policy Management and Evaluation
University of Toronto

© Copyright by Nimrah Abbasi (2018)
Standardization of Sonographic Prognostication of Prenatally Diagnosed Congenital Diaphragmatic Hernia (CDH) Across the North American Fetal Therapy Network (NAFTNet)

Nimrah Abbasi

Master of Science

Institute of Health Policy, Management and Evaluation
University of Toronto
2018

Abstract

Congenital diaphragmatic hernia (CDH) is a potentially lethal condition, with antenatal prognostication relying heavily on sonographic lung area determination. Although critical for prenatal counselling, inter-observer variability has been demonstrated in lung area measurement. To understand this variability, current practice patterns were evaluated through surveys, interviews and human factors analysis with usability testing and inter-rater agreement was determined for the three published lung area measurement methods (i.e.: antero-posterior, longest and trace) across the North American Fetal Therapy Network (NAFTNet). When comparing lung area methods, the trace method demonstrated the highest inter-rater agreement. Human factors analysis identified variability in sonographic image selection criteria and lung area measurement technique. This data informed the development of a standardized protocol for antenatal sonographic assessment of CDH, which underwent iterative redesigns through consecutive Plan-Do-Study-Act cycles among physicians and sonographers to meet the needs of users with various expertise and ultimately standardize prenatal CDH prognostication across NAFTNet.
Acknowledgments

The Department of Obstetrics and Gynaecology at Mount Sinai Hospital and University Health Network at the University of Toronto and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) are acknowledged for their generous support towards this study.

The following individuals are acknowledged (in alphabetical order) for their support, participation, time and contribution towards the conception, development and completion of this thesis:

- E Asztalos, Division of Neonatology, Department of Paediatrics, Sunnybrook Health Sciences Center and University of Toronto; Toronto, Canada
- R Baker, Quality Improvement and Patient Safety, Institute of Health Policy Management and Evaluation, University of Toronto; Toronto, Canada
- A Benachi, Service de Gynécologie-Obstétrique, AP-HP; Centre Maladie Rare: Hernie de Coupole Diaphragmatique Hôpital Antoine Béclère, Université Paris-Sud; Clamart, France
- A Johnson, The Fetal Center, Children's Memorial Hermann Hospital, University of Texas Health Science Center; Houston, USA
- S McDonald, Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynaecology, McMaster University Medical Center and McMaster University; Hamilton, Canada
- K Murphy, Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynaecology, Mount Sinai Hospital and University of Toronto; Toronto, Canada
- R Ruano, Division of Maternal-Fetal Medicine, Department of Obstetric and Gynaecology, Mayo Clinic, Mayo College of Medicine; Rochester, USA
- G Ryan, Fetal Medicine Unit, Mount Sinai Hospital, University of Toronto; Toronto, Canada
- J Saada, Service de Gynécologie-Obstétrique, AP-HP; Centre Maladie Rare: Hernie de Coupole Diaphragmatique Hôpital Antoine Béclère, Université Paris-Sud; Clamart, France
- H Sangi-Haghpeykar, Department of Obstetrics and Gynaecology, Baylor College of Medicine; Houston, USA
- MS Cortes, Texas Children's Fetal Center, Baylor College of Medicine, Department of Obstetrics and Gynaecology; Houston, USA
- G Seaward, Fetal Medicine Unit, Mount Sinai Hospital, University of Toronto; Toronto, Canada
- PS Shah, Division of Neonatology, Department of Paediatrics, Mount Sinai Hospital and University of Toronto; Maternal-Infant Care (MiCare) Center, Mount Sinai Hospital; Toronto, Canada
The following members of the North American Fetal Therapy Network are acknowledged (in alphabetical order) for their participation and contribution towards this project:

- Aguilera M, Salaman DL, Division of Maternal-Fetal Medicine, Midwest Fetal Care center; Minneapolis, MN.
- Bahtiyar M, Division of Maternal- Fetal Medicine, Yale Fetal Care center, Yale School of Medicine; New Haven, CT
- Baschat AA, Miller JL, Center for Fetal Therapy, Department of Gynaecology and Obstetrics, Johns Hopkins University School of Medicine; Baltimore, MD
- Bebbington M, Division of Maternal-Fetal Medicine, Fetal Care Center, Washington University School of Medicine; St. Louis, MO
- Bennett KA, Newton M, Division of Maternal-Fetal Medicine, Vanderbilt Center for Women's Health, Vanderbilt University Medical Center; Nashville, TN
- Benson C, Department of Radiology, Brigham and Women’s Hospital and Harvard Medical School; Boston, MA
- Berman S, Treadwell M, Division of Maternal-Fetal Medicine, C.S. Mott Children’s Hospital, University of Michigan School of Medicine; Ann Arbor, MI
- Blumenfeld Y, Division of Maternal-Fetal Medicine, Stanford University School of Medicine; Palo Alto, CA
- Brown RN, Morency AM, Division of Maternal-Fetal Medicine, Royal Victoria Hospital, McGill University Health Center; Montreal, QC
- Coleman BG, Oliver ER, Center for Fetal Diagnosis and Treatment at the Children's Hospital of Philadelphia and the Perelman School of Medicine at University of Pennsylvania; Philadelphia, PA
- Carr S, Davis S, Division of Maternal-Fetal Medicine, Women's and Infants Hospital, Brown University; Providence, RI
- Center J, Markham K, Division of Maternal-Fetal Medicine, Wexner Medical Center, Ohio State University; Columbus, OH
- Chescheir N, Goodnight W, Division of Maternal-Fetal Medicine, North Carolina Women’s Hospital, University of North Carolina; Chapel Hill, NC
- Dashe J, Santiago-Muñoz P, Division of Maternal-Fetal Medicine, UT Southwestern’s William P. Clements Jr. University Hospital, University of Texas Southwestern Medical Center; Dallas, TX
- Donepudi R, Papanna R, The Fetal Center, Children's Memorial Hermann Hospital, Department of Obstetrics and Gynecology, McGovern Medical School, University of Texas Health Science Center; Houston, TX
- Drennan K, Division of Maternal-Fetal Medicine, Lattimore Medical Center, University of Rochester Medical Center; Rochester, NY
- Emery SP, Makaroun S, Division of Maternal-Fetal Medicine, Magee-Women’s Hospital, University of Pittsburgh School of Medicine; Pittsburgh, PA
- Estroff J, Department of Radiology, Boston Children’s Hospital and Harvard Medical School; Boston, MA
- Gagnon A, Tessier F, Division of Maternal-Fetal Medicine, British Columbia Women’s Hospital and Children’s Hospital of British Columbia, University of British Columbia; Vancouver, BC
- Goldstein R, Morgan T, The Fetal Treatment Center, Department of Radiology, University of California; San Francisco, CA
- Halabi S, Department of Radiology, Stanford Children’s Health, Stanford University; Palo Alto, CA
- Keunen J, Van Mieghem T, Fetal Medicine Unit, Mount Sinai Hospital, University of Toronto; Toronto, ON
- Lim FY, Polzin W, Fetal Care Center, Cincinnati Children's Hospital Medical Center; Cincinnati, OH
- Mehollin-Ray AR, Department of Radiology, Baylor College of Medicine; Houston, TX
- Meyers ML, Department of Paediatric Radiology, University of Colorado Denver, Anschutz Medical Centre; Denver, CA
- Miller R, Simpson L, Division of Maternal-Fetal Medicine, New York-Presbyterian & Morgan Stanley Children’s Hospital of New York, Columbia University Medical Center; New York, NY
- Turan, O, Division of Maternal-Fetal Medicine, University of Maryland Medical System; Baltimore, MD
- Zaretsky MV, Division of Maternal Fetal Medicine Children's Hospital of Colorado, University of Colorado School of Medicine; Aurora, CO
Table of Contents

Acknowledgments........................................................................................................... iii
Table of Contents.................................................................................................................... vi
List of Tables............................................................................................................................... viii
List of Figures............................................................................................................................. ix
List of Appendices...................................................................................................................... xi

Introduction............................................................................................................................... 1
Objectives & rationale for standardization of CDH prognostication within the
North American Fetal Therapy Network (NAFTNet)......................................................... 4

Materials & Methods................................................................................................................ 7
Establishment of a core working group................................................................. 7
Setting, participants and sample size................................................................. 7
Feasibility................................................................................................................................. 8
Funding support......................................................................................................................... 8
Ethics........................................................................................................................................ 9
Research design........................................................................................................................... 9
  • PDSA cycles in the learning, development and implementation of a
    standardized protocol for sonographic prognostication of prenatally
diagnosed CDH: Summary of research design and protocol....................... 9
Identification of potential sources of variability in lung area measurement: Root
cause analysis...................................................................................................................... 11

Research procedure & protocol for participants......................................................... 13
I. Qualitative research methods: Understanding sources of variability in
sonographic evaluation of CDH across NAFTNet.............................................. 13
  a. Surveys............................................................................................................................. 14
  b. Human factors analysis: Usability testing and heuristic analysis 15
     i) Usability testing............................................................................................................. 15
        o Usability testing among NAFTNet participants....16
        o Usability testing among representative end-users:17
     ii) Heuristic analysis.................................................................................................... 18
        o Heuristic analysis of the NAFTNet FETO
          consortium protocol by the core working group and
          sonographers................................................................. 19

II. Quantitative research methods: Evaluating inter-rater agreement in
image quality assessments and lung area and head circumference
measurements across NAFTNet................................................................. 19
  a. Determination of agreement in image quality ratings and lung area
     and head circumference measurements among NAFTNet participants
     from FETO and non-FETO centers......................................................... 19
  b. Determination of bias in lung area measurement methods among
     NAFTNet participants from FETO and non-FETO centers in
     comparison to the ER ..................................................................................... 21
Outcome measures................................................................................................................ 23
Results ...........................................................................................................................................23

I. Qualitative data ...................................................................................................................................23
   a. Survey data collected from NAFTNet FETO and non-FETO centers ........................................23
   b. Results of usability testing among NAFTNet participants from FETO and non-FETO centers .........23

II. Quantitative data: .................................................................................................................................28
   a. Determination of agreement in image quality ratings and lung area and head circumference measurements among NAFTNet participants from FETO and non-FETO centers .................................................................28
   b. Determination of bias in lung area measurement methods among NAFTNet participants from FETO and non-FETO centers in comparison to the ER .................................................................29

III. Development of a NAFTNet standardized protocol for o/e LHR measurement in prenatally diagnosed CDH ...............................................................................................................................29
   a. Results from usability testing among NAFTNet participants from non-FETO centers and NAFTNet o/e LHR protocol redesign ........................................................................................................30
   b. Results of heuristic analysis for NAFTNet o/e LHR protocol by core working group and sonographers .................................................................31
   c. Results of usability testing on NAFTNet o/e LHR protocol among representative end-users ....33

Discussion ..............................................................................................................................................34

Strengths ................................................................................................................................................36

Limitations .............................................................................................................................................37

Future steps ............................................................................................................................................39

Conclusion ...........................................................................................................................................39

References ............................................................................................................................................41
Tables

Table 1. Standardization of antenatal sonographic prognostication of congenital diaphragmatic hernia (CDH) across the North American Fetal Therapy Network (NAFTNet) Project team

Table 2. Heuristic analysis for NAFTNet protocol design

Table 3. Plan-do-study-act (PDSA) cycles for developing a standardized sonographic protocol for prognostication of prenatally diagnosed CDH across NAFTNet

Table 4. Image selection criteria for lung area assessment in CDH across NAFTNet based on survey data

Table 5. Advantages and usability issues with lung area measurement methods according to NAFTNet participants

Table 6. Agreement among NAFTNet participants for image quality rating of lung and head sonographic clips in fetal left CDH

Table 7. Agreement among NAFTNet participants for head circumference and lung area measurements using anteroposterior, longest and trace method in fetal left CDH

Table 8. Agreement between different lung area measurement methods among NAFTNet participants in fetal left CDH

Table 9. Agreement between each NAFTNet participant from the FETO consortium and the external reviewer by lung area measurement method in fetal left CDH

Table 10. Agreement between each NAFTNet participant from centers outside of the FETO consortium and the external reviewer by lung area measurement method in fetal left CDH

Table 11. Results of heuristic analysis for NAFTNet o/e LHR protocol by core working group and sonographers

Table 12. Results of usability testing on NAFTNet o/e LHR protocol with representative end-users
Figures

Figure 1. Sonographic lung area measurement methods in prenatally diagnosed CDH.

Figure 2. Number of CDH cases seen annually prenatally in NAFTNet centers.

Figure 3. Stakeholder mapping & team building.

Figure 4. Ishikawa diagram: sources of variability in lung area estimation.

Figure 5. Two-dimensional sonographic axial images of varying image quality obtained from fetuses with left CDH at the level of the four-chamber view of the heart.

Figure 6. Process map for lung area and head circumference measurement in CDH.

Figure 7. Sonographic measurement of fetal head circumference.

Figure 8. Calculators used for o/e LHR estimation in CDH prognostication across NAFTNet. Results are grouped according to whether participants are from centers within and outside the FETO consortium.

Figure 9. Lung area measurement method preferences among NAFTNet participants. Results are grouped according to whether participants are from centers within and outside the FETO consortium.

Figure 10: Mean lung area per CDH study using anteroposterior, longest and trace methods among participants from NAFTNet FETO centers.

Figure 11: Mean lung area per CDH study using anteroposterior, longest and trace methods among participants from NAFTNet centers outside the FETO consortium.

Figure 12: Agreement between NAFTNet reviewers and external reviewer using trace method per CDH study (NAFTNet FETO centers).

Figure 13: Agreement between NAFTNet reviewers and external reviewer using longest method per CDH study (NAFTNet FETO centers).

Figure 14: Agreement between NAFTNet reviewers and external reviewer using anteroposterior method per CDH study (NAFTNet FETO centers).

Figure 15: Agreement between NAFTNet reviewers and external reviewer using trace method per CDH study (NAFTNet centers outside the FETO consortium).

Figure 16: Agreement between NAFTNet reviewers and external reviewer using longest method per CDH study (NAFTNet centers outside the FETO consortium).
Figure 17: Agreement between NAFTNet reviewers and external reviewer using anteroposterior method per CDH study (NAFTNet centers outside the FETO consortium)
Appendices

A. Letter to reviewers
B. Participant consent form
C. Questionnaire
D. Data collection sheet
E. Usability testing script
F. Usability data collection sheet for FETO group
G. Usability data collection sheet for non-FETO group
H. Usability data collection sheet for sonographers and clinical fellows in training
I. NAFTNet o/e LHR protocol PDSA 4
J. NAFTNet o/e LHR protocol PDSA 6 & 7
K. NAFTNet o/e LHR protocol PDSA 8
L. NAFTNet protocol for sonographic measurement of o/e LHR for antenatal prognostication of CDH
Introduction

Congenital diaphragmatic hernia (CDH) arises from an embryologic defect in the diaphragm, resulting in intrathoracic herniation of abdominal viscera and secondary mass effect on the developing lungs. Approximately 1 in 2500 to 5000 of newborns are affected by CDH (1), with nearly 60% of lesions diagnosed by prenatal ultrasound (2). As CDH is potentially a lethal condition, with neonatal morbidity and mortality predominantly related to underdevelopment of fetal lungs, or pulmonary hypoplasia, antenatal prognostication relies mainly on fetal lung volume estimation. Lung-to-head ratio (LHR) or observed-to-expected lung-to-head ratio (o/e LHR) are sonographic tools used to estimated lung volume on prenatal ultrasound, and o/e LHR has been shown to be among the main predictors of neonatal mortality locally (3) and internationally (4). Severity is classified as extreme, severe, moderate and mild with o/e LHR <15%, 15-25%, 25-35% and >35% respectively (5). Survival rates from Mount Sinai Hospital and Hospital for Sick Children in infants with prenatally diagnosed CDH were approximately 20%, 50% and 70% for o/e LHR <25%, 25-35% and 35-45% (3), which are consistent with data from a large European antenatal CDH registry (6). Poor prenatal prognostic indicators by ultrasound include an o/e LHR < 25% (7, 8), LHR < 1 (6, 9) and presence of the liver within the thoracic cavity, which generally indicate a substantial diaphragmatic defect with significant mass effect on the lungs (8, 10-12).

With severe CDH, neonatal survival is overall poor and estimated at 20% (6) even with standard postnatal care in specialized units, often due to severe irreversible pulmonary hypoplasia as well as pulmonary hypertension (7). Pregnancy options with CDH of this severity have traditionally included expectant management with attempt at standard postnatal repair, postnatal palliation, or, alternatively, termination of pregnancy. Improved understanding of fetal lung development has lead to innovative prenatal interventions to help improve outcomes in this poor prognostic group. Over the course of gestation, fetal lung development depends on cyclical pressure changes generated by continual fetal lung secretion and excretion, promoting lung growth and differentiation (13). Applying these principles to fetal lamb CDH models, experimental therapy with tracheal occlusion was performed to prevent expulsion of lung fluid, leading to expanded, fluid-filled lungs, which was found to promote lung growth, improve pulmonary function
and reduce herniated viscera into the abdominal cavity (14). Subsequently, a small uncontrolled case series in human fetuses demonstrated improved survival with fetal endoscopic tracheal occlusion (FETO) for severe CDH, defined as LHR < 1.4 with liver herniation (15). A National Institutes of Health (NIH)-funded randomized control trial (RCT) was then conducted evaluating the efficacy of FETO in severe CDH as defined above, which demonstrated no advantage over standard postnatal care (16). The study has since been criticized for its high LHR cut-off, resulting in inclusion of very few fetuses with LHR < 1, which is typically used to define severe disease. Subsequently, the European fetal medicine centers evaluated the efficacy of FETO in severe CDH characterized by intrathoracic liver herniation with an LHR < 1 or an equivalent o/e LHR in a prospective cohort study of 210 fetuses. When compared to expectantly managed controls, a statistically significant increase in survival was noted from 24 to 49 % and 0 to 35% in fetuses treated with FETO with left and right sided CDH respectively (17). To further evaluate the efficacy of FETO, the TOTAL (Tracheal Occlusion To Accelerate Lung growth, http://www.totaltrial.eu/) trial, a European-based international randomized trial, is currently underway comparing FETO to routine postnatal surgical repair in the treatment of moderate and severe CDH [TOTAL Trial 1 for moderate pulmonary hypoplasia (NCT00763737 and TOTAL Trial 2 for severe pulmonary hypoplasia (NCT01240057)].

Although, antenatal prognostication and selection for fetal intervention in prenatally diagnosed CDH relies largely on sonographic lung area estimation, predictive value of LHR for neonatal outcomes has been variable (4, 10, 12, 18-20), with some series questioning the prognostic value of LHR entirely (12, 18, 19). Possible explanations include varying sample size of studies, use of LHR, which varies more with gestational age than o/e LHR (4), use of different sonographic lung area estimation methods with differing inter-observer variability (21, 22) and inconsistent implementation of standardized protocols for lung area measurement (22, 23).

To understand potential sources of variability in antenatal sonographic assessment of CDH and its prognostic value, ultrasound parameters used in clinical practice to estimate lung area and how they are obtained are first reviewed. The sonographic LHR has been widely used to estimate pulmonary hypoplasia (11), and is calculated by
obtaining the ratio between the contralateral lung area and head circumference to correct for the physiologic increase in lung area that occurs with advancing gestational age (GA). However, since fetal lung volume grows exponentially relative to head circumference in both normal and fetuses affected with CDH (4, 21), a higher LHR may be derived at advanced GA irrespective of disease severity, potentially leading to incorrect prognostication. Thus, the reported variability in the prognostic value of sonographic lung area measurements (4, 10, 12, 18-20) may partly be attributed to gestational age at assessment. Consequently, Jani et al. introduced the concept of o/e LHR to further correct for increases in lung area with gestation, which is obtained by expressing the observed LHR in the contralateral lung in affected fetuses as a percentage of the expected mean LHR for GA in normal fetuses (4). The o/e LHR is an important prognostic indicator in both left- and right- sided CDH, and an independent predictor of mortality, irrespective of liver herniation and GA (4). In prenatally diagnosed CDH, o/e LHR has now become the mainstay of antenatal prognostication. Several references for expected LHR values exist (21, 22, 24, 25), and commonly used online o/e LHR calculators are available from www.perinatology.com and www.totaltrial.eu based on mathematical equations provided by Jani et al. (22) and Claus et al. (24) respectively, to obtain expected lung area values. Thus, an additional contributor to variability may be related to use of different expected reference values across and within centers when determining o/e LHR.

The presence of several different measurement methods for lung area estimation may also contribute to variation in sonographic prognostication of CDH. Currently, 3 different sonographic measurement techniques are used in practice for estimating lung area on an axial plane of the fetal chest (22): (1) multiplication of the anteroposterior diameter of the lung at the midclavicular line by the perpendicular (transverse) diameter at the midpoint of the anteroposterior diameter (AP method); (2) multiplication of the longest diameter of the lung by its longest perpendicular diameter (longest diameter method); (3) manual tracing of the limits of the lungs (trace method) (Figure 1). When Peralta et al. (21) compared the three different techniques for intra- and inter- observer agreement on images from 60 fetuses among 2 observers, the trace method was the most reproducible, whereas the longest diameter method was the least reproducible and overestimated lung volumes by 45%. When receiver operating characteristic (ROC)
curves were calculated for the statistical prediction of neonatal survival for the different lung area measurement techniques (26), the trace method was found to be the strongest predictor of postnatal survival with an area under the ROC curves (AUC) of 0.84 (26), which was also confirmed by a subsequent systematic review evaluating antenatal predictors of CDH (8). Despite this data, there is no single lung area method measurement in use universally and choice ultimately depends on sonographer and center preference.

Predictive value of o/e LHR for neonatal survival may also depend on level of expertise and center volume. In a study from 31 centers within the French National Center for Rare Diseases, the predictive value of o/e LHR for 28-day mortality was compared between large and small centers defined as an annual caseload of $\geq 14$ cases or $< 14$ cases per year respectively. During the period of study, 82 cases were managed by 2 large centers and 223 cases by 29 small centers. Although, overall o/e LHR was predictive of 28-day mortality, the ability to predict survival was statistically better in large centers; for a specificity of 0.30, the sensitivity was 0.71 in high volume centers compared to 0.55 in smaller centers (27). This data suggests that performance of this prognostic indicator varies with center experience and also highlights the potential need for training among health care providers in smaller centers to ensure accurate measurement.

**Objectives & rationale for standardization of CDH prognostication within the North American Fetal Therapy Network (NAFTNet)**

After discordant neonatal outcomes were reported in infants with prenatally diagnosed CDH of similar severity based on antenatal sonographic lung area estimation within the North American Fetal Therapy Network (NAFTNet), a voluntary association of 33 medical centers from Canada and United States of America with an expertise in prenatal diagnosis and fetal therapy (https://www.naftnet.org/), many physicians began to question the validity and reproducibility of these prenatal prognostic indicators. Furthermore, as several North American centers were evaluating the role of FETO therapy and preparing to offer this treatment and/or participate within the TOTAL trial, it became particularly prudent to standardize antenatal prognostication of CDH. Through
standardization of sonographic prognostication, accurate parental counseling and appropriate selection of candidates for fetal intervention can be ensured, and neonatal outcome data may also be generalized across centers. Thus, it became one of the mandates of the NAFTNet FETO consortium, which consists of 9 high volume fetal treatment centers within NAFTNet managing the majority of prenatally diagnosed CDH cases across the continent, to standardize prenatal prognostication of CDH across NAFTNet. However, before implementation of standardized practice, it was critical to first quantitatively and qualitatively assess variation in sonographic prognostication of CDH across NAFTNet currently.

An attempt to standardize CDH prognostication has occurred within the fetal medicine community, initiated by Peralta et al. (21) by first identifying the problem of inter-observer variability in lung area estimation using the different lung area measurement methods, which was followed by the publication of a protocol for lung area measurement by recognized experts in the field (22). However, what remains unclear is whether this protocol adequately addresses variability in clinical practice and if it has been interpreted and successfully implemented in prenatal ultrasound as intended by the authors. The disparate reports of neonatal outcomes across NAFTNet may suggest otherwise. This attempt at standardization of CDH prognostication resembles the typical “trial and error” approach to problems identified in healthcare settings, where a change is attempted without sufficient study of the underlying issue in a complex social system (28). In contrast, the science of improvement uses a multidisciplinary approach drawing on clinical science, statistical analyses, systems theory and the psychology of leading change, to develop rapid cycle tests of change to be used in the field to enable a deeper understanding of the problem and identify potential changes that can lead to improvement within a given context (29). The Model for Improvement provides a framework for developing, testing and implementing changes that lead to improvement, and seeks to find the balance between sufficient study and analysis of the problem with meaningful well-informed action by asking three critical questions: (1) what are we trying to accomplish; (2) how will we know that a change is an improvement and (3) what change can we make that will result in an improvement (30)? Rapid cycle tests of change in the form of Plan-Do-Study Act (PDSA) cycles are often employed directly in the field to
answer these questions and develop and refine interventions to address the identified problem (28). The four steps of the PDSA cycle consist of determining the objectives, prediction of findings, and detailing how to carry out the study in the “plan” stage, followed by the “do” phase, which consists of carrying out the planned study and collecting data. The “study” phase consists of analyzing and summarizing data with comparison to predictions, which can guide iterative improvements and inform the subsequent cycles in the “act” phase (28).

The objective of this project was to evaluate practice patterns in sonographic prognostication of prenatally diagnosed CDH and determine the lung area measurement method with the highest inter-rater agreement among fetal medicine specialists within NAFTNet. The data collected ultimately informed the development of a standardized protocol for o/e LHR determination in antenatal sonographic prognostication of CDH for implementation across NAFTNet. Applying the Model for Improvement, several PDSA cycles were employed to define the areas lacking standardization in CDH sonographic assessment and develop a protocol for CDH prognostication, which sought to reduce these sources of variability recognized in practice. In contrast to the traditional development of guidelines and protocols by a panel of experts, an emphasis was placed on involvement of multidisciplinary experts and frontline workers of various clinical backgrounds within cyclical tests of change, allowing for development and testing of the intervention on a smaller scale before publication and implementation of this protocol across a broader population.

As medical errors can arise from various individual and systemic factors, even with safe practices in place (31), the expanding field of Human Factors Engineering (HFE) emphasizes a user-centered design of devices and systems that is mindful of the needs and limitations of users throughout the design process, which may help reduce errors through informed development of systems and products that can ultimately facilitate successful implementation (32). This is in contrast to the traditional means of designing a system or a device, with little understanding of user needs or user input and an expectation of users to adapt after the design is complete (32). Although, HFE is becoming increasingly incorporated in the development of medical technology and systems, its use in medical protocols and policies remains infrequent. The principles of
HFE were applied in the development of the NAFTNet protocol for sonographic prognostication of CDH, specifically HFE techniques were used to better understand the critical components of lung area measurement and identify sources of variability in clinical practice. This novel approach enabled development of a user-centered protocol with the central aim of preventing errors and inconsistencies in antenatal CDH prognostication and ultimately to promote patient safety in the field of prenatal diagnosis and therapy.

Materials & Methods

Establishment of a core working group

This is a multicenter study with principal investigators from 4 NAFTNet FETO consortium centers, including Mount Sinai Hospital, University of Toronto (N.A., G.R.), Children’s Memorial Hermann Hospital, University of Texas Health Science Center McGovern Medical School (A.J.), Mayo Clinic College of Medicine (R.R.) and Texas Children's Fetal Center, Baylor College of Medicine (M.S.). Additional members of the core working group included biostatistical support from University of Toronto (P.S., X.Y.) and Baylor College of Medicine (H.S.), information technology for provision of standardized viewing software for ultrasound image review (A.S.) and maternal-fetal medicine specialists from an experienced European fetal medicine center to advise on study design and participate as an external reviewer (ER) in the study (A.B., J.S.) (Table 1). The ER are from an institution currently participating in the TOTAL trial and were selected based on academic contribution in the field of prenatal diagnosis and management of CDH as well as experience with standardization of CDH prognostication across fetal medicine centers within France.

Setting, participants and sample size

NAFTNet centers were recruited for participation at the biannual NAFTNet scientific meetings in November 2016 and April 2017, where the study aims, objectives and protocol were presented. Participating NAFTNet centers were categorized into “FETO” and “non-FETO” groups, with the former including centers within the FETO
consortium. The FETO consortium includes large volume fetal medicine centers with an annual CDH caseload of ≥ 15 cases (Figure 2), currently evaluating the role of FETO and includes centers with established programs or in the early stages of program development for FETO therapy. These centers are also currently participating in the TOTAL trial or represent sites that may potentially be recruited in upcoming years. Thus, centers within the FETO consortium may be considered a more experienced (i.e.: “expert”) subgroup within the network.

NAFTNet centers internally selected 1-2 fetal medicine or medical imaging specialists involved in reviewing most of the ultrasound images for fetal CDH locally for participation in the study. A total of 58 participants were included, with 19 members from all 9 centers within the FETO consortium and 30 participants from 17 non-FETO centers, an external “expert” reviewer (A.B.) and 5 sonographers and 3 clinical fellows training in fetal medicine. Participants from NAFTNet centers were of various clinical backgrounds and included 39 maternal-fetal medicine specialists (12 participants from FETO centers, 27 participants from non-FETO centers), 9 radiologists (6 participants from FETO centers, 3 participants from non-FETO centers) and 1 paediatric surgeon from a FETO center.

**Feasibility**

Stakeholder mapping was performed with the core working group to enable identification of key members of NAFTNet and the NAFTNet FETO consortium whose endorsement and input would be critical in developing a comprehensive protocol and influencing change in practice, due to their longstanding academic contributions to the fetal medicine community and specifically prenatal diagnosis of CDH (Figure 3). These prominent members within the FETO consortium and members of the NAFTNet board of directors along with all FETO centers were recruited individually at the NAFTNet FETO consortium meetings to ensure participation. Findings were presented at each FETO consortium gatherings to maintain engagement and input from key stakeholders and ultimately facilitate implementation of the protocol across NAFTNet.

**Funding support**
This North American multicenter collaboration was made possible with the generous support from grants provided by the Eunice Kennedy Shriver National Institute of Child Health and Development (NICHD 5R13HD059293-05) and Mount Sinai Hospital, University of Toronto Department of Obstetrics and Gynecology and University Health Network Research Fund.

**Ethics**

This study was approved by the NAFTNet steering committee with institutional review board approval from Baylor College of Medicine (H-37017) and Mayo Clinic College of Medicine, Rochester (16-008333) and research ethics board approval from Mount Sinai Hospital, Toronto (17-0168-C).

**Research design**

The study design was predominantly descriptive and observational, aimed at understanding and quantifying variability in practice patterns in prenatal sonographic assessment of CDH among fetal medicine specialists within NAFTNet. A standardized protocol for antenatal sonographic prognostication of CDH was developed through consecutive Plan-Do-Study-Act (PDSA) cycles, which refined the intervention via a continuous process of developing and conducting small tests of change.

- **PDSA cycles in the learning, development and implementation of a standardized protocol for sonographic prognostication of prenatally diagnosed CDH: Summary of study design and protocol**

Several PDSA cycles formed the basis of achieving a standardized approach for prenatal sonographic lung area assessment in CDH. In the recently published primer on PDSA by Leis *et al.* (33), it was suggested that the traditional ‘plan’ in PDSA can be replaced by ‘predict’, with each cycle including a clear hypothesis, a test of change to answer this hypothesis, followed by a focused analysis and conclusion describing the best step forward for the next PDSA cycle. Furthermore, Leis *et al.* emphasize that initial PDSA cycles should focus on learning about the local problem and development of the intervention rather than rushing towards implementation of an intervention (33).
Ignoring the initial learning and developmental PDSAs risks the possibility of incorrectly understanding and framing the problem, leading to uninformed action, ineffective interventions and disengagement, in part related to inadequate investment of resources into understanding cultural and structural barriers to change (33, 34). Drawing from these study designs, much emphasis was placed on conducting PDSAs in the investigation and problem framing stages in this project as defined by Reed et al. (34), which were essential in understanding and defining the problem within the NAFTNet context, identifying possible contributing factors, recognizing and engaging key stakeholders, and ultimately informing the design of the final intervention. PDSA cycles were categorized into learning, development and implementation PDSA cycles as described by Leis et al.(33). As in typical quality improvement projects, where progress rarely follows a simple linear trajectory, the PDSA cycles in this project alternate back and forth between learning and development prior to the final implementation PDSA, reflecting the iterative design process that occurred based on newly acquired knowledge and understanding through serial PDSA cycles.

Initial PDSA cycles were in the learning phase focusing on evaluating variability in practice patterns through surveys and human factor analytical tools and quantifying agreement in lung area estimation among the NAFTNet FETO consortium. This evaluation outlined the key steps involved in lung area measurement and problems with current protocols and measurement tools when applied in “expert” North American Fetal therapy centers. Subsequent PDSAs were involved in developing a standardized protocol for o/e LHR in antenatal sonographic prognostication of CDH based on initial data collected from the FETO consortium centers. This change idea modelled the detailed protocol developed by the Fetal Medicine Foundation that was highly successful in standardizing sonographic nuchal translucency measurement used in prenatal screening for trisomy 21(35). Subsequent PDSA cycles were performed in the learning phase to evaluate agreement and variability in practice for antenatal CDH prognostication among the remaining NAFTNet centers. These PDSA cycles were designed to identify usability issues unique to centers of varying expertise and volume, which informed further iterations of the protocol. Final PDSA cycles were in the development phase with use of various human factors analytical tools, which enabled evaluation of the protocol by the
core working group and educational sonographers and testing of the protocol among local sonographers and clinical fellows training in fetal medicine to identify additional usability issues among trainees and frontline workers and further refine the NAFTNet protocol for o/e LHR determination.

**Identification of potential sources of variability in lung area measurement: Root cause analysis**

To identify potential sources of variability in lung area measurement and guide development of interventions in standardization of CDH prognostication, a root cause analysis was conducted via a videoconference call with the core-working group (N.A., A.J., G.R., R.R.). An Ishikawa diagram was created to review policies (i.e.: related to lung area measurement), procedural aspects (i.e.: lung area measurement technique), staff (i.e.: physicians and sonographers acquiring images and institutions within which they practice), patients (i.e. maternal and fetal factors that may impact image acquisition and quality), equipment (i.e.: ultrasound technology available to obtain images) and information factors (i.e.: physician and sonographer knowledge and educational resources) that may be contributing to variability in sonographic fetal lung area assessment (Figure 4). Major sources of variability identified in the published literature were presented, which were primarily related to the presence of several lung area measurement methods used in practice with variable inter-observer agreement (21). Another possible source of variation in measurement technique identified in the literature was related to the unclear original description of sonographic lung area estimation (11, 22). Specifically, the text within the original article by Metkus *et al.* (11) describing sonographic lung area measurement is consistent with the longest diameter method, and the corresponding image within the article is more consistent with the AP technique. The use of the gestational-age dependent LHR rather than o/e LHR was also identified as an important contributor to heterogeneity in antenatal prognostication (22). The absence of a standardized protocol was also recognized as a major source of variation in practice, prompting the development of a protocol for measurement by Jani *et al.* (22), which remains the principal reference in use for o/e LHR measurement within the fetal medicine community. Inadequate training as a potential contributor to inaccurate lung area
assessments, and need for standards to maintain competency were also suggested in the literature (22).

After presentation of these findings from the literature, the variability in lung area measurement and predictive value of this prognostic indicator was further scrutinized by the working group to identify additional contributing factors within NAFTNet. Related to policies and procedure, despite the presence of a published protocol for lung area measurement by Jani et al. (22), it remained uncertain whether this protocol was universally implemented across NAFTNet centers, if image selection criteria used in practice was consistent with the protocol and whether this protocol adequately addressed variability in practice patterns. The proportion of centers continuing to use LHR rather than o/e LHR was unknown and the references used for expected lung area values to determine o/e LHR across NAFTNet centers was also uncertain. A centralized auditing system to review images from sonographers to ensure competency was also noted to be lacking. When looking at staff-related causes, heterogeneity in lung area measurement methods used across NAFTNet centers was identified. Furthermore, potential political barriers to adoption of standard practice were recognized, as this would possibly involve moving towards implementation of lung area measurement methods different than those originally described by several members within NAFTNet. Insufficient sonographer training and variability in center experience and volume were also thought to potentially impact accuracy of lung area assessment. Patient factors such as maternal obesity, suboptimal fetal position, presence of additional fetal abnormalities, alterations in amniotic fluid volume, laterality of diaphragmatic defect (i.e.: right- or left- sided or bilateral CDH), size and severity of the defect and presence or absence of liver herniation, were all identified as potential contributors to suboptimal visualization of fetal lungs, thus impairing lung area estimation. Equipment factors such as older ultrasound machines and absence of visual aids and schematics for lung area measurement within ultrasound labs were also identified as possible contributors to poor image quality and measurement inaccuracies. Regarding information, potential knowledge gaps among sonographers and physicians regarding prognostic significance of o/e LHR for CDH prognostication and details of measurement technique were identified. Furthermore, the absence of a true gold standard for fetal lung volume determination prenatally was
recognized as an important limitation in identifying the most valid sonographic lung area estimation method for use in clinical practice.

Based on this root cause analysis, the working group selected the most critical contributors to variability in sonographic prognostication of CDH based on supporting evidence in the literature as well as perceived feasibility in addressing the various suspected causes. Maternal and fetal factors contributing to poor image quality were felt to represent technical challenges that could not be circumvented. Centralized auditing of images and improving available ultrasound technology were also felt to be out of the scope of the funding and resources available for this initiative. After exclusion of these causes, it was determined that the most important contributor to variability in sonographic lung area measurement was related to different lung area measurement methods used in practice across NAFTNet, as demonstrated in the literature (21). Thus, efforts were put primarily towards establishing and promoting the use of the lung area method with highest inter-rater agreement across NAFTNet. Another major factor promoting heterogeneity in clinical practice was the absence of a standardized protocol for use across NAFTNet. Furthermore, the available published protocol (22) was not felt to adequately standardize the measurement process, specifically by failing to recommend a specific lung area measurement method, or the use of o/e LHR with a particular reference for expected lung area value determination. Insufficient educational modules and visual aids were also felt to be important sources of knowledge gaps among sonographers and physicians, preventing adequate assessment of fetal CDH. Therefore, the next major underlying cause to be addressed was the absence of a comprehensive standardized protocol addressing lung area measurement for use across NAFTNet that could also bridge these knowledge gaps. Finally, key stakeholder individuals and institutions that contributed significantly to the field of prenatal management of CDH were identified within NAFTNet, and from whom buy-in, endorsement and input would be actively sought throughout protocol development.

Research procedure & protocol for participants

All participants received an introductory letter highlighting the purpose of the study and steps involved as well as a consent for participation with a questionnaire and
data collection sheet to be completed (see Appendix A ‘Letter to reviewers’, Appendix B ‘Participant consent form’, Appendix C ‘Questionnaire’ and Appendix D ‘Data collection sheet’). Initially consent was not obtained from FETO center participants during the early phases of the project, as they were also considered collaborators of this initiative. However, after further discussion with the Research Ethics Board at Mount Sinai Hospital, consent was recommended for all participants and therefore a retroactive consent for previously collected data was obtained from all FETO center participants. For non-FETO centers, consent was obtained prospectively. Both qualitative and quantitative methods were used to understand and estimate variability in practice for sonographic lung measurement in prenatally diagnosed CDH. The methodology is detailed below.

**I. Qualitative research methods: Understanding sources of variability in sonographic evaluation of CDH across NAFTNet**

The initial focus of this project was to understand the core task for sonographic CDH prognostication: lung area measurement. The original literature describing the technique as well as published protocols were reviewed to determine current available standards and recommendations. Questionnaires and videoconference calls were also conducted with participants to understand individual and institutional practice patterns within NAFTNet.

**a. Surveys**

All participants received a questionnaire consisting of 16 questions (see Appendix C ‘Questionnaire’) addressing the root causes identified among the core working group as potential sources of variability in lung area measurement including factors related to staffing and institution (e.g.: center volume, self-reported confidence with measurements), equipment (e.g.: image reporting system, type of ultrasound machine), information and knowledge (i.e.: relevance of o/e LHR for antenatal CDH prognostication, perception of inter-observer variability) and procedures and policies (e.g.: preferred methods of lung area measurement, image selection criteria, references used for o/e LHR determination).
All 49 NAFTNet participants and the ER completed surveys. Respondents were not required to answer all questions. All data was compiled by a single principal investigator (N.A.). Data collected from the surveys provided a broad overview of variability in clinical practice specific to NAFTNet and also helped quantify the contribution of the various root causes within NAFTNet (see Results below “Survey data collected from NAFTNet FETO and non-FETO centers”). As the project expanded rapidly to several non-FETO NAFTNet centers, questionnaires and consents were completed on an online platform rather than via email in the second half of the project, to assist with compilation of data and maintenance of confidentiality.

b. Human factors analysis: Usability testing and heuristic analysis

Videoconference calls were conducted with participants to further discuss and identify differences in practice patterns while sonographically assessing CDH. During interviews, participants were provided with de-identified sonographic axial clips of the fetal chest with isolated left CDH and were asked to perform sonographic measurements of lung area using the 3 different methods, which provided an opportunity to directly observe participants during assessments and compare image selection criteria and measurement techniques between participants. Human factors analytical tools such as usability testing and heuristic analysis were employed to better appreciate differences in practices, challenges with current methods for lung area estimation and provide the data to guide the development of an o/e LHR protocol for use within NAFTNet.

i) Usability testing

Usability testing can help identify user issues when a new device is in the users’ hands performing typical tasks, which may be overlooked during the design phase, when the tool is evaluated outside the working environment with experts and manufacturers rather than typical end-users (36). This form of assessment is performed with representative end-users while interacting with a new technology or system change usually in a simulated setting to reflect the true working environment (36). By providing participants with sonographic images from fetuses with CDH, a simulated environment was created in which usability testing on the available lung area measurement methods
could be performed, allowing identification of challenges and usability issues with the various methods. Measurements were obtained on 3 sonographic cases of varying image quality (Figure 5) to contrast usability issues with excellent image quality (i.e. optimal fetal position with appropriate two dimensional [2D] image optimization and adequate visualization of all necessary landmarks and lung borders [case 1]) and poor image quality (i.e. poor visualization of fetal lungs due to suboptimal fetal position with excessive rib shadowing over lung area [case 2], or in cases of maternal obesity [case 3]).

Participants were observed while obtaining measurements by a single principal investigator (N.A.) via videoconference calls. Participants were encouraged to “think aloud”, describe their measurement technique, identify important landmarks and discuss areas of ambiguity in current methods and protocols. These assessments enabled in-depth task analysis of lung area measurement and identify key steps involved (see below ‘Results of usability testing among NAFTNet participants from FETO and non-FETO centers’ and Figure 6).

The non-evaluative nature of the sessions was emphasized with participants. The objectives of these assessments were also reviewed, including the goals of gathering knowledge and expertise from fetal imaging specialists and to understand challenges with current measurement protocols for lung area assessment in CDH (see Appendix E ‘Usability testing script’). It was also discussed, that this data would ultimately guide future standardization efforts in CDH antenatal prognostication. All usability testing was performed by a single principal investigator (N.A.) and conducted over 20-30 minutes. Data collected was also anonymized and compiled by a single principal investigator (N.A.).

**Usability testing among NAFTNet participants**

Usability testing initially performed within the FETO group was aimed at gathering the key elements involved in lung area measurement among fetal medicine specialists with established expertise in prenatal diagnosis and also to understand potential pitfalls with the different methods used in practice (see Appendix F ‘Usability data collection sheet for FETO group’). Usability testing was performed on 18/19 FETO participants from all 9 FETO centers and the ER. From this initial data collection, the
necessary steps and variations in practice in lung area measurement among experts were identified (see below ‘Results of usability testing among NAFTNet participants from FETO and non-FETO centers’ and Figure 6), which directed the initial development of the NAFTNet protocol for lung area measurement (see Appendix I ‘NAFTNet o/e LHR protocol PDSA 4’).

Usability testing was then performed among non-FETO center participants using the same CDH sonographic images. This enabled identification of additional usability issues and areas of ambiguity in current protocols and measurement methods among centers with varying experience and center volume. Furthermore, it also captured whether the critical steps identified by the FETO group and the protocol by Jani J et al. (22) for image selection and lung area measurement using the different methods were being employed across NAFTNet centers (see Appendix G ‘Usability data collection sheet for non-FETO group’ and see below ‘Results of usability testing among NAFTNet participants from FETO and non-FETO centers’).

Due to the large number of participating NAFTNet centers outside the FETO consortium, usability testing was conducted only among selected centers based on center volume determined by self-reported annual CDH caseload from the survey data. Centers were classified based on annual volume (i.e. ≤5, >5-10, >10-15, >15-20, >20 cases per year) (Figure 2) and usability testing was performed on 13 participants from 11 non-FETO centers, including 2 centers with ≤5 cases per year, 4 centers with > 5-10 cases annually, 3 centers with >10-15 cases per year, 1 center with >15-20 cases and 1 center with > 20 cases annually.

• **Usability testing among representative end-users**

In the final stages of the NAFTNet o/e LHR protocol development for lung area measurement, the protocol underwent usability testing among representative end-users, which included 4 sonographers and 3 physicians in training in fetal medicine, 5 of whom had limited experience obtaining o/e LHR (< 5 measurements performed over the last year) and 2 users had no experience obtaining o/e LHR and none of the participants had formal training in measuring o/e LHR. The objective of this final usability testing was to
evaluate the efficacy of the protocol in training individuals to perform adequate fetal lung area assessments.

Participants were provided with sonographic images from case 1 (Figure 5) used in the usability testing for NAFTNet participants, which was a study of excellent image quality with optimal visualization of lung tissue and critical landmarks. Participants were asked to select a static plane for lung area measurement, list their image selection criteria and perform a lung area measurement to the best of their ability using the trace method with no guidance. Participants were encouraged to “think aloud” and describe their steps in detail. Subsequently, participants were provided with the NAFTNet o/e LHR protocol and asked to perform the same assessments on the same images. Image frame selection and quantitative lung area measurement was compared pre and post introduction of the protocol. Image selection criteria stated by participants was also compared before and after the protocol was provided, specifically evaluating whether participants discussed the following essential image selection criteria: 4-chamber view of the heart, axial section of the chest, presence of clear lung borders, optimal fetal position and 2D image optimization (i.e.: appropriate magnification, depth, focus, gain). Competency in trace method technique was evaluated by reviewing the lung perimeter tracing on the sonographic still image before and after introduction of the protocol and comparing whether all lung tissue was included without extrapulmonary structures (see Appendix H ‘Usability data collection sheet for sonographers and clinical fellows in training’). Participants were also asked to provide any general feedback for the protocol, particularly focusing on clarity of instructions and sequence of steps. Assessments before and after provision of the protocol were performed by a single principal investigator (N.A.).

ii) Heuristic analysis

With medical devices, heuristic analysis can be employed by experts in the field to identify design issues or heuristic violations that may lead to patient errors and inefficiencies when used in the typical work environment. Violations recognized can guide iterative improvements in design during development and early implementation phases (32). Although typically used to identify software usability issues, Zhang et al. adapted the traditional heuristic analysis tools and compiled a list of 14 heuristics for
medical devices entitled the “Nielsen–Shneiderman Heuristics” to ensure a good user-interface design in the healthcare setting (32). Many of the “Nielsen–Shneiderman Heuristics” can also be applied to the development of a new medical protocol or policy including: maintenance of consistency with current standards and convention; minimalist design; minimization of memory load; prevention of error; clear closure, with every task having a clear beginning and end and use of the user’s language (32). These heuristics were used to guide the critique and redesign of subsequent versions of the NAFTNet protocol for lung area measurement in CDH.

- **Heuristic analysis of the NAFTNet FETO consortium protocol by the core working group and sonographers**

  Drawing on the principles of heuristic analysis, the core-working group (N.A., A.J., R.R., G.R.) was provided with a list of the above-stated heuristics (see Table 2) and the protocol designed by the NAFTNet FETO consortium was evaluated for any heuristic violations as a group over a videoconference call. Following this discussion, all reported violations were subjectively rated as major or minor by a principal investigator (N.A.) depending on whether the violation could lead to misinterpretation of steps in lung area measurement and potential inaccurate sono graphic lung area estimation. All major violations were prioritized and addressed in redesigns of subsequent iterations. A similar heuristic analysis was performed on the subsequent iteration of the protocol with the lead educational sonographer and an experienced sonographer within the Fetal Medicine Unit at Mount Sinai Hospital individually. A list of the same heuristic violations were provided to the sonographers and all stated violations were compiled and rated as major or minor by N.A. using the same scale utilized within the working group and the subsequent protocol designs addressed identified violations (see ‘Results of heuristic analysis for NAFTNet o/e LHR protocol by core working group and sonographers’).

**II. Quantitative research methods: Evaluating inter-rater agreement in image quality assessments and lung area and head circumference measurements across NAFTNet**

To evaluate inter-rater agreement in lung area estimation and head circumference, participants were asked to assess image quality and perform measurements on de-
identified axial sonographic clips of the fetal chest and head from anatomical studies of fetuses with an isolated left CDH (see Appendix D ‘Data collection sheet’). Studies for use in this project were selected by three fetal medicine specialists within the core-working group (G.R., R.R., A.J.). All cases were obtained on a GE Voluson E8 machine between 22 to $33^{+6}$ weeks gestation and were of varying image quality to reflect heterogeneity of study quality encountered in clinical practice. The working group decided to use 10-15 studies. As there was no data in the literature to guide determination of an adequate sample size for number of CDH cases to include in order to assess inter-rater agreement with the number of participants included in this study, the sample size was selected based on practicality and feasibility to promote maximal study participation and ensure measurements could be performed within a reasonable time frame of 2-3 hours. Thirteen studies with sufficient visualization of the contralateral lung for measurement were selected by the core working and approved for use by the ER. All US studies were de-identified and encrypted by Trice Imaging Inc. Images were stored on a secure Trice Imaging Inc. cloud and reviewed by participants using standardized viewing software provided by Trice Imaging Inc. Measurements were completed by 19 FETO center participants and 29/30 non-FETO center participants and the ER.

Participants were provided with the commentary and published protocol for o/e LHR measurement by Jani et al. (22) with the introductory letter electronically (see Appendix A ‘Letter to reviewers’). Based on this criteria and their own subjective impression of the images, participants were asked to grade the quality of the sonographic clips based on the following scale of 1-4 (1= excellent/ optimal image quality that allows accurate determination of lung area or head circumference measurements; 2= good / good image quality, sufficient to determine lung area or head circumference measurements; 3= fair/ suboptimal image quality, just sufficient to determine lung area or head circumference measurements; 4= poor/ unacceptable image quality to obtain lung area or head circumference measurements). Participants were asked to select a static plane for measurement of the right lung area and head circumference and to indicate the frame number selected. Participants obtained a head circumference by measuring the biparietal diameter (BPD) by placing the callipers from the “outer” leading edge of the skull to the “inner” skull border and the occipito-frontal diameter (OFD), by placing the callipers
perpendicular to the BPD in the middle of the frontal and occipital skull bones and the following formula was used to calculate the head circumference: \([\text{OFD (mm)} + \text{BPD (mm)}] \times 1.62\) (Figure 7a). Alternatively, head circumference could be obtained by using an elliptical tracer tool around the outer border of the skull (Figure 7b). Either techniques are acceptable methods for head circumference measurement according to the Internal Society of Ultrasound in Obstetrics and Gynaecology practice guidelines for the routine midtrimester fetal ultrasound scan (37). Participants were asked to obtain lung area using the 3 published techniques (22), including AP, longest and trace method (Figure 1). Participants were referred to the protocol by Jani et al. (22) as a reference for the different lung area measurement methods. Of note, lung area methods were presented in an altered sequence on data collection sheets for each participant to counterbalance any potential bias that may have occurred if all reviewers were asked to obtain lung area measurements using methods in the same sequence (e.g.: if all reviewers were asked to obtain lung area by AP followed by longest and then trace method, a learning effect with the trace method may be seen given that it would have always been the last method used).

a. Determination of agreement in image quality ratings and lung area and head circumference measurements among NAFTNet participants from FETO and non-FETO centers

The challenge with establishing a preferred sonographic measurement method for lung area estimation arises from an inability to determine true fetal lung volume antenatally. Consequently, measurement methods cannot be compared to any gold standard to establish validity. Although the most valid lung area method could not be determined, agreement in lung area measurement using the 3 different methods was determined by obtaining the intraclass correlation (ICC 3, 1), which evaluates agreement among a group of operators, in this case the NAFTNet participants. This enabled identification of the method with the highest inter-rater agreement. A bootstrapping method was used to provide the 95% confidence intervals (CI). Intraclass correlation was also determined for head circumference measurements, primarily as a reference for an ideal sonographic parameter with anticipated high inter-rater agreement. Since fetal head circumference represents a component of basic fetal biometry with strict
measurement criteria, the anticipated variability in measurements between operators would be expected to be low. The usual interpretation of ICC values for agreement are as follows: <0.5 indicating poor agreement, 0.51-0.75 as moderate, 0.76-0.89 as good and 0.90-1 indicating excellent agreement (38). Agreement for image quality grading of head and chest images was assessed by determining Fleiss’ kappa, as is typically performed for assessment of agreement between more than 2 raters for categorical data. Suggested interpretation of kappa values for agreement in healthcare settings are as follows: 0-0.20 indicating no agreement, 0.21-0.39 as minimal, 0.40-0.59 as weak, 0.60-0.79 as moderate, 0.80-0.90 as strong and >0.90 indicating near perfect agreement (39). Overall, any kappa value below 0.60 may be interpreted as inadequate agreement (39).

b. Determination of bias in lung area measurement methods among NAFTNet participants from FETO and non-FETO centers in comparison to the ER

Average lung area measurement for each CDH study among NAFTNet reviewers as a group was compared to the ER to determine the mean difference (bias) for each lung area method by creating Bland and Altman (B & A) plots. Again, as there is no gold standard for lung area measurement, agreement with an “external expert” was used as a surrogate for assessment of external validity of lung area measurement methods. Bland & Altman analysis can be used to describe agreement between two quantitative measurements or operators by plotting the difference of the 2 measurements (i.e. sonographic lung area) on the Y-axis against the mean of the two measurements (i.e. sonographic lung area) on the X-axis. Agreement is determined by calculating the bias (i.e.: mean difference in lung area measurements) and the standard deviation of differences between two measurement methods or two operators (40). The B & A plot depict the mean difference (bias) as well as the limits of agreement, which are represented by 1.96 standard deviations above and below the mean difference (40). If the differences are normally distributed, then 95% of differences will be within the limits of agreement interval. Although B & A plots define the limits of agreement, they do not indicate whether these limits are clinically acceptable. Maximally acceptable differences will depend on the clinical setting.
Participants from FETO and non-FETO centers were analyzed separately and findings were compared. Additionally, mean difference (bias) between each participant and the ER was also determined for each lung area method. Statistical analyses were performed using SAS 9.4, and SAS macro packages of intracc.sas and Magree.sas (SAS Institute, Inc., Cary, NC).

**Outcome measures**

The measure of interest in this project was inter-rater agreement within NAFTNet for quality rating of CDH sonographic images and lung area and head circumference measurements through determination of kappa statistics and ICC, respectively. This enabled assessment of internal validity of lung area measurement methods within NAFTNet. Bias or mean difference between NAFTNet reviewers and the ER was determined to evaluate external validity of current measurement methods. Furthermore, if inter-rater agreement was found to be poor for lung area measurements (ICC < 0.76) within NAFTNet centers outside of the FETO consortium, an assessment of inter-rater agreement and determination of bias would be repeated following introduction and implementation of the NAFTNet standardized protocol for o/e LHR to ensure competency in lung area measurement and to validate the protocol for training.

**Results**

Results from qualitative and quantitative data collection are presented below with comparison of findings between centers within and outside the NAFTNet FETO consortium. Findings from the PDSA cycles are summarized in Table 3, and are also further elaborated below.

**I. Qualitative data**

**a. Survey data collected from NAFTNet FETO and non-FETO centers**

Data from surveys revealed that all NAFTNet centers determined contralateral lung area for antenatal CDH prognostication, with one center using LHR and the remaining using o/e LHR. To obtain o/e LHR, the majority of participants were using the
online calculator provided by www.perinatology.com (n=27), followed by www.totaltrial.eu (n=7), and 1 participant used both calculators and compared values. Three participants used the expected lung area reference values provided by Peralta et al. (21) and 9 participants did not specify a reference (Figure 8).

Nearly 75% of participants from FETO centers described the o/e LHR as highly relevant for CDH prognostication. Among centers outside of the FETO consortium, only 30% of participants (n=9/30) described the o/e LHR as highly relevant, 50% of participants (n=14/30) reported the o/e LHR to be of moderate relevance, and 23% of participants (n=7/30) described it as somewhat relevant for antenatal CDH prognostication.

When asked to estimate the number of annual referrals for prenatally diagnosed CDH, 13 non-FETO centers reported evaluating 5-15 cases of CDH per year and all FETO centers reported seeing at least 15-20 cases annually (Figure 2). When questioned regarding self-perceived comfort with obtaining o/e LHR, participants from the FETO consortium were very confident (i.e. expert level) or fairly confident. Among non-FETO centers, 11/30 participants were very confident (i.e. expert level), 11/30 participants felt fairly confident and 8/30 were somewhat confident in obtaining o/e LHR.

Most participants were uncertain about the optimal number of o/e LHR assessments required prior to assigning an o/e LHR; however, most obtained 2-4 measurements. When assigning an o/e LHR, responses were similar across NAFTNet, with more than 50% of participants (n=26/48) using the average value obtained and one third of participants (n=14/48) using the highest value and 2 participants using the lowest value. Nearly 15% of participants (n= 6/48) used the qualitatively “best” measurement.

When participants were asked to rate inter-observer variability for o/e LHR value among experienced sonographers, trends were similar across NAFTNet. The majority of participants (n=34/49) felt there was some inter-observer variability but unlikely to be of clinical significance, 12 participants felt there was high inter-observer variability, 2 participants were uncertain and 1 participant felt it was insignificant. Regarding intra-observer variability for o/e LHR among experienced sonographers, the majority (n=33/49) felt there was some intra-observer variability that was unlikely to be of clinical significance.
significance, followed by high intra-observer variability (n=7/49), insignificant intra-observer variability (n=8/49) and 1 participant was uncertain.

Reporting systems used in practice varied, with most centers using Viewpoint, AS-OBGYN and Clickview. Other systems used include Powerscribe, Siemens Syngo, Observer, Digisonics, Astraia and locally developed software. Various ultrasound machines were available in centers, with more than half of fetal medicine centers using GE Voluson machines (E8, E9, E10). Additional ultrasound machines used in practice include Samsung WS80 elite; Phillips (IU22, Epiq 7, HD11X); Siemens (S2000 Helix, Acuson S2000) and Toshiba Aplio 500.

When participants were asked to list the most important imaging features for an optimal o/e LHR assessment, criteria were similar across NAFTNet centers (Table 4) and overall consistent with the protocol described by Jani et al. (22). Participants from FETO centers, however, were more precise in detailing necessary landmarks on the axial plane of the fetal chest to make accurate lung area measurements and also emphasized basic two-dimensional image optimization. Overall, FETO center participants were also more informed about pitfalls of each method for lung area estimation.

The majority of participants emphasized the importance of obtaining a 4-chamber view on an axial section of the fetal chest. However, FETO participants frequently detailed what comprised a “true” axial plane of the chest including visualization of a single rib on each side of the chest, to ensure the section was not oblique and falsifying lung area. Although challenging to achieve clear visualization of intracardiac anatomy including the atria, atrio-ventricular (AV) valves and interventricular septum in severe CDH, FETO participants insisted that attempts must be made to visualize these structures as best possible. Most participants also deemed clear lung borders in the absence of rib and extremity shadowing critical. Many participants from FETO centers, emphasized the importance of assessing an axial sweep of the chest rather than a still image before assigning severity, as a sweep may allow better appreciation of defect size and intrathoracic herniation of liver and/or other viscera. Furthermore, with the use of color Doppler, hepatic vasculature may be identified, which may further facilitate recognition of intrathoracic liver herniation. The majority of participants discussed the critical role of standard 2D image optimization. Many of the FETO center participants emphasized the
importance of obtaining an appropriate angle of insonation through the intercostal space to avoid rib shadowing and also specified ways to optimize the image with use of a high frequency probe (i.e.: 5-9 MHz) to better distinguish lung from surrounding tissue, and adjustment of focus, depth, gain, dynamic range and magnification, with the chest filling at least 75% of the screen. Preferred fetal position varied depending on the lung area method used in practice. The majority of participants using the longest or trace method preferred the contralateral lung close to the transducer with the fetal spine at 3 or 9 o’clock depending on fetal presentation. Those using the AP method preferred the fetus in supine position, with the spine at 6 o’clock and sternum at 12 o’clock. This position was felt to reduce rib shadowing, prevent overestimation of lung area, which may occur with decubitus or prone positioning, and allow placement of callipers in a true AP orientation relative to the fetal vertebrae and sternum.

Lung area measurement methods preferences varied across centers within and outside the FETO consortium (Figure 9). Among centers within the FETO consortium, the majority of participants preferred the trace method (n=10), followed by the longest (n=5) and AP method (n=4). However, when practice patterns were evaluated by institution, 4 centers used the longest method, 3 centers used the AP method and 2 centers used the trace method. Across the remaining NAFTNet centers, individual preferences were aligned with center preferences, with most participants preferring the longest method (n=13/29), followed by the trace method (n=11/29) and finally the AP method (n=5/29). Participants using the longest and AP method in practice often referenced an institutional preference since these methods were the originally described techniques. Most participants using the trace method recently adopted this technique after it was demonstrated to have greater inter-observer agreement (21).

b. Results of usability testing among NAFTNet participants from FETO and non-FETO centers

During usability testing of lung area measurement methods, most participants appeared hesitant with the AP method and uncertain of which landmarks to use. Many participants were uncertain of whether the AP diameter should be placed at the midclavicular line or where the measurement was longest, and many were uncertain of
where to obtain the transverse diameter. Many participants made reference to the discrepancy in the original paper describing sonographic lung area measurement (11) as previously discussed. In contrast, participants using this technique routinely in practice performed measurements with ease and typically placed the AP diameter along the midclavicular line from the atria to the posterior wall of the chest, parallel to the vertebral body and sternum and the transverse diameter from the descending aorta to the medial border of the rib.

Most participants found the longest diameter method easy to use as no specific landmarks for calliper placement were required. Many of the FETO center participants remarked that this technique could easily over-estimate lung area particularly with an oblique section of the chest, irregular shaped lungs or if callipers traversed the cardiac atria. During observations, some participants appeared to overestimate lung area, with callipers traversing structures other than lung, and this was more pronounced with irregularly shaped lungs.

When evaluating the trace method, perceived advantages included ability to better capture lung area, particularly with atypically shaped lungs. Many participants noted that lung borders needed to be particularly crisp. Several FETO center participants emphasized the need for high-resolution imaging to exclude mediastinal vessels, spleen, bowel and pericardium while manually tracing lung perimeters. Upon observation, some participants inadvertently included structures other than lung while using the trace method, most commonly mediastinal vessels. Advantages and usability issues with lung area measurement methods identified through usability testing among NAFTNet participants are summarized in Table 5.

When FETO participants were observed obtaining head circumference measurements, there was near complete uniformity in landmarks used for assessment, including visualization of the midline falx separating hemispheres symmetrically, cavum septum pellucidi, thalami, and posterior horns of lateral ventricles and absence of posterior fossa structures with clear visualization of skull borders. Non-FETO participants were not asked to perform head circumference measurements during videoconference calls, due to exceptional consistency noted in measurement technique
among FETO participants and thus this parameter was felt to be an unlikely source of variability in o/e LHR estimation in practice.

Summary of key measurement steps and variation in practices based on data collected from surveys and usability testing are summarized in a process map for lung area measurement in fetuses with CDH (Figure 6).

II. Quantitative data

a. Determination of agreement in image quality ratings and lung area and head circumference measurements among NAFTNet participants from FETO and non-FETO centers

Among NAFTNet reviewers, agreement for grading of image quality was overall poor, with an overall kappa of 0.05 (95% CI 0.03, 0.08) and 0.1 (95% CI 0.1, 0.2) for thorax and head clips respectively among FETO centers and a kappa of 0.07 (95%CI 0.05, 0.08) for thorax and 0.1 (95% CI 0.09, 0.1) for head images among non-FETO centers (Table 6). Agreement was excellent for head circumference measurement among NAFTNet FETO participants (ICC 0.99; 95% CI 0.992, 0.998) but notably lower for non-FETO centers (ICC 0.83; 95% CI 0.78, 0.99). Among the lung area measurement methods in FETO centers, agreement was highest for the trace method (ICC 0.94; 95% CI 0.83, 0.98), followed by the longest diameter method (ICC 0.89; 95% CI 0.75, 0.97) and lowest for the A-P method (ICC 0.83; 95% CI 0.67, 0.94). Although inter-rater agreement for all lung area measurement methods was lower for non-FETO centers compared to FETO centers, patterns were similar with an ICC of 0.86 (95% CI 0.79, 0.93) for trace method, 0.57 (95% CI 0.12, 0.90) for longest method and 0.54 (95% CI 0.1, 0.91) for AP method (Table 7). When average lung area estimates obtained by NAFTNet reviewers using the 3 different lung area measurement methods were compared, the highest agreement was demonstrated between the A-P and trace methods among the FETO group (ICC 0.99; 95% CI 0.91, 0.996) as well as the non-FETO group (ICC 0.95; 95% CI 0.70-0.99) (Table 8). Lung area estimates were highest using the longest method for each sonographic CDH case among FETO and non-FETO groups (Figure 10, 11).
b. Determination of bias in lung area measurement methods among NAFTNet participants from FETO and non-FETO centers in comparison to the ER

Using B & A analysis to evaluate agreement between NAFTNet participants and the ER, the trace method demonstrated the lowest bias. Among the FETO group, the mean difference (bias ± standard deviation [SD]) in lung area measurement was 14 ± 38 mm² for the trace method (Figure 12), 54 ± 60 mm² for the longest method (Figure 13) and 23 ± 50 mm² for the AP method (Figure 14). Similar trends were noted in the non-FETO centers, with a mean difference (bias) in lung area measurement of 19 ± 36 mm² for the trace method, 62 ± 57 mm² for the longest method and 42 ± 49 mm² for the AP method (Figures 15-17). When individual reviewers were compared to the ER, agreement was overall high for the trace method (Table 9, 10).

III. Development of a NAFTNet standardized protocol for o/e LHR measurement in prenatally diagnosed CDH

Quantitative and qualitative data collected from surveys, interviews and usability testing from NAFTNet FETO centers were presented at the biannual NAFTNet meeting in April 2017. Members of the FETO consortium present at the meeting were gathered following the meeting to form a working group for development of a standardized protocol for o/e LHR measurement. The previously published protocol for o/e LHR measurement created by Jani et al. (22) was reviewed and adapted to incorporate the preliminary data collected from the NAFTNet FETO consortium. Although, the overall practice for sonographic prognostication of CDH within the NAFTNet FETO consortium was largely consistent with the protocol by Jani et al. (22), certain additional specifications were recommended by the FETO group based on the data presented. Specifications related to landmarks of the axial view of the chest and 4-chamber view of the heart, optimal fetal position, and image optimization techniques were felt to be lacking. Based on the data available for inter-rater agreement for lung area measurement methods from the NAFTNet FETO consortium, it was decided that the trace method would be recommended for use in practice. Furthermore, since centers are preparing to join the TOTAL trial, the calculator provided by www.totaltrial.eu was recommended to determine o/e LHR, to ensure consistency across all NAFTNet sites. Finally, it was also
recommended to make a specification that if several estimates for lung area are obtained, then the qualitatively best image should be used to assign an o/e LHR, rather than use of the highest, average or lowest o/e LHR value, which could inaccurately estimate lesion severity if the image is of suboptimal quality. Competency assessment for o/e LHR measurement was not addressed, as it was felt that there was insufficient evidence to make any firm recommendations. The protocol by Jani et al. (22) was adapted accordingly based on the recommendations of the working group (see Appendix I ‘NAFTNet o/e LHR protocol PDSA 4’).

a. Results from usability testing among NAFTNet participants from non-FETO centers and NAFTNet o/e LHR protocol redesign

The adapted standardized protocol for o/e LHR by the FETO consortium was further revised following review of surveys, usability testing and results of inter-rater agreement for lung area and head circumference measurements from non-FETO NAFTNet centers, which identified several additional knowledge gaps (see Appendix J ‘NAFTNet o/e LHR protocol PDSA 6 & 7’). Prognostic value of o/e LHR in prenatally diagnosed CDH and advantages over LHR was included to address a knowledge gap identified, as many non-FETO center participants felt this parameter was only moderately relevant according to survey data. Sonographic images of a fetal chest and head with lung area and head circumference measurements respectively were included from a fetus with left CDH in optimal position for assessment. This visual aid was provided to further “minimize memory load”. Given lower inter-rater agreement in head circumference measurements among non-FETO centers, a step describing head circumference measurement with necessary landmarks was incorporated into the protocol. It was also specified that an elliptical tool should be used to measure head circumference rather than use of the following formula: head circumference = 1.62 x (biparietal diameter [BPD] and occipitofrontal [OFD] diameter, since this formula incorporates two measured parameters (i.e.: OFD and BPD) with more than one acceptable technique for measurement (i.e.: BPD may be measured from outer edge to inner edge of the fetal skull, which is the ‘leading edge’ technique or can be measured from outer edge to outer edge of the fetal skull (37)), which may contribute to inter-observer variability in
head circumference measurements. As many participants were noted to include mediastinal vessels and extrapulmonary tissue while tracing the perimeters of the lung during usability testing, the sonographic image of left CDH included in the protocol also demonstrated accurate tracing of the lung tissue. Further specifications of fetal position and optimal number of lung area measurements to obtain were also included, to address some of the uncertainty expressed by many participants regarding these issues during conference calls. As several participants were not using axial clips when assessing CDH sonographically, a recommendation to obtain a sonographic clip in addition to a still image of the axial fetal chest during prenatal assessment of CDH was included (see Appendix J ‘NAFTNet o/e LHR protocol PDSA 6 & 7’).

b. Results of heuristic analysis for NAFTNet o/e LHR protocol by core working group and sonographers

The revised protocol underwent a heuristic analysis by the core working group (N.A., A.J., G.R., R.R.) as well as the lead educational sonographer and an experienced sonographer in the Fetal Medicine Unit at Mount Sinai Hospital, based on an adaptation of the list of heuristic violations provided by Zhang et al. (32) for medical devices (Table 2). Following heuristic analysis with the core working group, the major violation recognized was related to tasks being unclear with several subtasks within each step, making the protocol difficult to follow. This was felt to represent a violation of clear closure, with every task having a clear beginning and an end. Criteria to ensure competency in o/e LHR measurement among sonographers was lacking, and recognized as a major violation in consistency with current standards and convention that could result in inadequately trained providers performing o/e LHR and potentially jeopardizing patient safety. Additionally, heuristic analysis with sonographers revealed that many practitioners were not aware of the severity categories for o/e LHR and corresponding survival rates were frequently forgotten. This represented a major heuristic violation related to “minimization of memory load” and “prevention of error”, which could potentially lead to incorrect prognostication. Additional violations recognized among both groups were related to “minimalist design”, with text described as “small and cramped” and the protocol was also described as “wordy”, “unclear” and “difficult to
read quickly”. User’s language was also not adequately utilized, specifically when addressing avoidance of rib shadowing and image optimization and within the explanation of the prognostic value of o/e LHR, which could lead to misinterpretation of critical steps within the protocol. The sonographers also noted that steps did not follow the typical sequence of image acquisition in practice, representing a minor violation of consistency with current standards and convention.

To address the major violations, steps were further simplified into smaller subtasks and reordered in a sequence to more accurately reflect image acquisition in practice. Severity classification of o/e LHR with associated mortality was also included in the protocol to assist with accurate counselling. Regarding competency assessment, it was recognized that ideally images should be submitted for review by a central body or organization with expertise in prenatal diagnosis; however, the associated requirements for funding and professional support to establish this type of organization were unlikely to be available in the near future. Furthermore, although one series had demonstrated that a minimum of 70 scans would be required to ensure competency in LHR measurements (41), this was also felt to be an excessive and unrealistic expectation for sonographers in training and inconsistent with competency requirements for other obstetrical ultrasound parameters (e.g.: nuchal translucency, cervical length measurement, uterine artery Doppler assessment). Thus, a recommendation was made for training to be performed under the supervision of an experienced sonographer or sonologist, with a requirement to obtain 5 images meeting the NAFTNet o/e LHR protocol criteria with image review by a local expert. These recommendations were based on an expert international consensus statement for obstetrics and gynaecological ultrasound competency assessment in residency training (42), which suggests establishing competency in a given ultrasound imaging skill on the basis of five image submissions that meet necessary criteria. However, the consensus statement also acknowledges that although this recommendation is supported by some data (43), the literature for competency assessment in ultrasound training remains sparse. Thus, it was recognized by the working group, that as competency assessment is better defined within the ultrasound literature, the criteria for competency in sonographic assessment of CDH may also be adjusted in the future.
To address the minor violations mostly related to design and layout of the protocol, wording was simplified, text was minimized, and font size was increased with spacing out of information to support a minimalist design and allow users to have an understanding of key measurement steps with a quick glance at the text and diagrams. (see Appendix J ‘NAFTNet o/e LHR protocol PDSA 6 & 7, Appendix K NAFTNet o/e LHR protocol PDSA 8 and Table 11).

c. Results of usability testing on NAFTNet o/e LHR protocol among representative end-users

Due to overall good inter-rater agreement for lung area measurements and demonstration of similar trends for the different lung area measurement methods among non-FETO NAFTNet centers, the core-working group decided against performing an assessment of inter-rater agreement among non-FETO centers for lung area measurement after implementation of the NAFTNet o/e LHR protocol. However, members of the core-working group emphasized the need for the protocol to be operationalized by less experienced users and frontline staff. Thus, usability testing was performed on the revised versions of the NAFTNet protocol following heuristic analysis among 4 less experienced sonographers and 3 physicians training in fetal medicine at Mount Sinai Hospital. This enabled identification of additional usability issues that may possibly impact correct interpretation and implementation of the protocol.

When asked to describe image selection criteria prior to introduction of the protocol, none of the participants listed all necessary elements for image selection criteria (i.e.: 4-chamber view of the heart, axial section of the chest, presence of clear lung borders, ideal fetal position and 2D image optimization). All participants mentioned the importance of a 4-chamber view of the heart and presence of clear lung borders and 3 participants mentioned need for an axial section of the chest. Only 2 participants made reference to fetal position and none mentioned the importance of or specified features of image optimization. When comparing assessments before and after introduction of the protocol (see Appendix K ‘NAFTNet o/e LHR protocol PDSA 8’), image frame selection for lung area measurement was similar, however, 5/7 participants mentioned all essential features for selection criteria and were quicker in selecting an appropriate image for
measurement after presentation of the protocol. Although, this rapidity in image selection may have been related in part to a learning effect by using the same fetal sonographic clip for the post-protocol assessment, the improved verbalization of selection criteria appeared to help focus the user’s attention on identifying the presence of critical image features when reviewing and selecting images. Furthermore, on image review, only 2/7 participants included tissue other than lung in their lung area measurement compared to 5/7 participants prior to introduction of the protocol. Most commonly, mediastinal vessels were incorrectly included in lung area estimation (see Table 12). When the protocol step describing lung perimeter tracing was reviewed with the 2 users making these errors after provision of the protocol, it was apparent that some users did not actually know or understand what composed the mediastinal vessels or where they were located. To address this usability issue, mediastinal structures were labelled in the sonographic image included in the protocol to clarify anatomical location and further prevent inclusion in lung area estimation (see Appendix L ‘NAFTNet protocol for sonographic measurement of o/e LHR for antenatal prognostication of CDH’).

**Discussion**

Compilation of data from surveys and usability testing conducted across nearly 75% of NAFTNet centers demonstrated overall consistency in sonographic assessment of CDH, with all centers using an o/e LHR or LHR as part of antenatal prognostication. The majority of participants felt confident in obtaining lung area measurements and most felt this parameter was at least moderately relevant for antenatal prognostication. Image selection criteria for lung area assessment were also largely consistent, which included an axial view of the chest that included a 4-chamber view of the heart and clear lung borders without shadowing and standard 2D optimization. However, participants from FETO centers appeared more familiar with the pitfalls of various lung area measurement methods and were more detailed in specifying selection criteria and steps involved in lung area measurement, likely attesting to their experience with antenatal assessment of CDH. Most members of the FETO consortium preferred the trace method for lung area assessments, in contrast to the non-FETO centers, which preferred the longest diameter
method. However, in practice, most FETO centers used the longest diameter for lung area estimation. Overall, participants demonstrated the most unfamiliarity with the AP method and inconsistency in landmarks used for measurements.

When assessing inter-rater agreement for the sonographically measured components of o/e LHR, an excellent agreement was noted among participants from the NAFTNet FETO consortium for head circumference and agreement was highest when the trace method was used for lung area estimation in CDH. Among non-FETO centers, the trace method also demonstrated the highest inter-rater agreement, however, overall, agreement was comparatively lower for all measurements, particularly head circumference. Notably, disparity between the lung area methods was greater among non-FETO centers, with AP and longest methods demonstrating significantly poorer inter-rater agreement. This further highlights the advantages of using the trace method across centers of varying expertise and volume. When lung area estimates were compared using the 3 methods, the longest method yielded the highest lung area estimates for each CDH case and the greatest agreement was demonstrated between the A-P and trace method across NAFTNet. The bias was lowest for the trace method and highest for the longest diameter method when comparing NAFTNet participants to the ER. These results are consistent with the series from Peralta et al. (21) evaluating inter-rater agreement in lung area measurement methods.

Despite providing the published protocol for o/e LHR by J. Jani et al. (22) prior to assessments, agreement was poor among FETO and non-FETO centers for image quality ratings. This finding could reflect a lack of consensus for image selection criteria for optimal sonographic lung area measurement among fetal medicine specialists. This poor agreement may also indicate inconsistent implementation or interpretation of the published protocol for o/e LHR (22), or alternatively, that this protocol does not sufficiently account for more subtle sources of variability in o/e LHR measurements. After evaluation of practice patterns across NAFTNet through usability testing, additional specifications for image selection criteria were recognized, which may in turn further improve inter-rater agreement in image quality assessments as well as measurements. Furthermore, to promote consistency in measurement and prognostication across
NAFTNet, the trace method and the calculator provided by www.totaltrial.eu was recommended for use in clinical practice.

**Strengths**

This is the largest multicenter collaboration evaluating practice patterns and inter-rater agreement in lung area measurement for CDH prognostication across the major fetal therapy centers in North America. Although previous series have evaluated agreement in lung area measurements, these studies have been conducted within single centers (21, 23) or in normal fetuses (21). Furthermore, this project is distinct from previous studies by focusing on understanding variability in CDH sonographic assessment among fetal medicine specialists from institutions of varying size and expertise through surveys, interviews and human factors analytical tools, which helped guide the iterative improvements of the NAFTNet o/e LHR protocol. Acknowledging that the initial o/e LHR protocol by Jani et al. (22) and the NAFTNet FETO consortium were designed by experts, which may not entirely address the user needs of less experienced sonographers, usability testing of the revised protocol was performed among frontline clinicians and sonographers of varying experience to identify additional usability issues. Ultimately, this allowed for the development of a highly “user-centered” standardized protocol, in contrast to most protocols in medicine, which are often immediately launched into the workforce prior to any form of usability testing or heuristic analysis and with little input from end users.

Bias was limited in this study by blinding all participants to the measurements of other participants and neonatal outcomes of the fetuses from which the ultrasound images were obtained. Additionally, a single investigator collected and compiled data from surveys, interviews and usability testing, thus facilitating comparison of practices across NAFTNet. Finally, many participants remarked that this study was an important quality improvement initiative in itself, by allowing specialists to reflect on their individual and institutional practices and identify areas that remain unclear. Furthermore, by performing repetitive lung area measurement exercises, many participants remarked feeling more confident in performing sonographic assessments of CDH.
Interaction with participants during interviews and usability testing was not only instrumental in providing meaningful insight into variability in practice across NAFTNet, but also engaged participants by allowing them to provide a more personal contribution to the project. As standardization of lung area measurements would involve moving towards a single method of measurement different from the originally described techniques by some of the major North American fetal therapy centers, a particular emphasis was placed on closely involving key stakeholders in “co-design” of the protocol, which has facilitated uptake of a standardized practice across NAFTNet. Additionally, since data collected from the FETO centers informed the development of the initial protocol, a strong sense of ownership developed within the FETO consortium for this initiative, which further promoted endorsement of this initiative within NAFTNet.

Limitations

Surveys and videoconference calls were not conducted anonymously, which represents an important limitation of this study. However, this approach was deliberate to ensure that practices from centers of all levels of experience could be assessed. By evaluating centers of differing sizes, unique challenges were identified, which helped adapt the protocol accordingly to facilitate universal implementation. Moreover, to ensure confidentiality, all data was compiled and anonymized by a single investigator. Finally, although overall participation was high, usability testing was performed on 60% of non-FETO centers, with 1-2 representatives per site. Thus, variations in practice could not be evaluated across all NAFTNet centers and within institutions. However, after completion of usability testing within this project, similar usability issues were recurrently identified among participants, making it less likely that new challenges would be detected with more participants.

Another major limitation of this study includes its design, whereby physicians were asked to obtain measurements on stored images, which may over-estimate inter-rater agreement (44). Ideally, participants would obtain their own images during real-time scanning and lung area estimates could then be compared between participants. However, this approach would be unfeasible due to the rarity of this condition and the associated impracticalities associated with live scanning of a single patient by multiple
international reviewers. Furthermore, using pre-selected clips provides an element of realism, as assessments in clinical practice are often performed on suboptimal images due to technical limitations.

Although the trace method was found to have the highest inter-rater agreement, the most valid method remains unknown, as there is no gold standard available for antenatal fetal lung volume estimation. Furthermore, clinically acceptable ICC cut-offs for inter-rater agreement in fetal sonographic measurements are difficult to define. Higher ICC thresholds may be required for fetal measurements, due to the larger true variance in fetal dimensions across gestation when compared to adult measurements (44). Martins et al. have suggested defining moderate reproducibility with an ICC cut off of 0.98-0.99, and for use in clinical practice with caution(44). However, none of the lung area methods reach this cut-off based on the data collected within NAFTNet. Furthermore, although, the trace method has the narrowest limits of agreement, it remains uncertain whether these limits are clinically acceptable. As the limits of agreement for the trace method include a wide range of measurement differences, this could potentially result in clinically significant inter-observer variability in o/e LHR and ultimately antenatal prognostication.

While detailed in its description of o/e LHR measurement, the NAFTNet o/e LHR protocol is still limited in that it focuses on left-sided CDH. However, the protocol can be extrapolated to right-sided CDH as general principles for image selection and acquisition likely remain similar. Furthermore, the protocol ultimately relies on sonographer skill to obtain the correct plane in the fetal chest, however, by precisely outlining image criteria, this now enables a more systematic approach in recognizing an image as ‘acceptable’ or ‘unacceptable’ for assessment. Although usability testing of the protocol was performed on less experienced sonographers and trainees within one FETO site (i.e. Mount Sinai Hospital) in the final PDSA, perhaps more widespread usability testing among sonographers at other FETO and non-FETO centers may have identified additional usability issues and barriers to appropriate implementation. Finally, although not deemed necessary by the core-working group, an assessment of inter-rater agreement among NAFTNet centers following the introduction of the o/e LHR protocol may have helped validate the use of this protocol in practice.
**Future steps**

Future initiatives to ensure successful implementation and sustainability of the protocol long-term may be conducted through periodic auditing of images locally or centrally at a center of excellence or through a committee within the NAFTNet FETO consortium. This type of quality assurance initiative may require a substantial commitment of financial and staffing resources, however, this may be reasonable particularly in cases of fetal intervention or in settings of randomized controlled trials evaluating efficacy of fetal intervention in CDH to ensure fetuses are appropriately being selected for treatment or randomization, respectively. If implementation of the protocol remains suboptimal for less experienced users, additional interventions such as eLearning modules or videos may also be considered. Future initiatives evaluating inter-rater agreement in sonographic prediction of intrathoracic liver herniation and quantification of liver herniation and total lung volume by magnetic resonance imaging would also be valuable, as liver herniation has been shown to be a predictor of neonatal morbidity and mortality in CDH (8, 10-12). Ultimately this data could be incorporated into a comprehensive standardized antenatal prognostication protocol for CDH combining the various prognostic indicators and imaging modalities used in practice.

**Conclusion**

Lung area estimation and determination of o/e LHR represent the cornerstone of antenatal prognostication for CDH. However, if attempts are not made to address the variability in practice patterns across fetal medicine centers, accurate prenatal counselling and appropriate selection for fetal intervention cannot be ensured, which may result in a significant quality gap in patient care. Furthermore, the use of unstandardized prognostic tools can confound interpretation of neonatal outcome data and efficacy of fetal therapy. Through the development and implementation of a protocol for lung area estimation and o/e LHR determination, it is anticipated that a standardized approach for prenatal prognostication of CDH can be achieved across NAFTNet centers and that the principles
of this study design can be applied in the development of standardized protocols for assessment of other prenatally diagnosed fetal disorders.
References


Tables

Table 1. Standardization of antenatal sonographic prognostication of congenital diaphragmatic hernia (CDH) across the North American Fetal Therapy Network (NAFTNet) Project team

Table 2. Heuristic analysis for NAFTNet protocol design

Table 3. Plan-do-study-act (PDSA) cycles for developing a standardized sonographic protocol for prognostication of prenatally diagnosed CDH across NAFTNet

Table 4. Image selection criteria for lung area assessment in CDH across NAFTNet based on survey data

Table 5. Advantages and usability issues with lung area measurement methods according to NAFTNet participants

Table 6. Agreement among NAFTNet participants for image quality rating of lung and head sonographic clips in fetal left CDH

Table 7. Agreement among NAFTNet participants for head circumference and lung area measurements using anteroposterior, longest and trace method in fetal left CDH

Table 8. Agreement between different lung area measurement methods among NAFTNet participants in fetal left CDH

Table 9. Agreement between each NAFTNet participant from the FETO consortium and the external reviewer by lung area measurement method in fetal left CDH

Table 10. Agreement between each NAFTNet participant from centers outside of the FETO consortium and the external reviewer by lung area measurement method in fetal left CDH

Table 11. Results of heuristic analysis for NAFTNet o/e LHR protocol by core working group and sonographers

Table 12. Results of usability testing on NAFTNet o/e LHR protocol with representative end-users
Table 1. Standardization of antenatal sonographic prognostication of congenital diaphragmatic hernia (CDH) across the North American Fetal Therapy Network (NAFTNet) Project team

### Core Working Group

- **Team Lead/Process Owner**: Nimrah Abbasi (Fetal Medicine Unit, Mount Sinai Hospital, University of Toronto; Toronto, Canada)
- **Executive Sponsor (local)**: Greg Ryan (Head, Fetal Medicine Unit, Mount Sinai Hospital, University of Toronto; Toronto, Canada)
- **Executive sponsors (NAFTNet FETO consortium)**
  - Anthony Johnson (The Fetal Center, Children's Memorial Hermann Hospital, University of Texas Health Science Center; Houston, TX, USA)
  - Rodrigo Ruano (Division of Maternal-Fetal Medicine, Department of Obstetric and Gynaecology, Mayo Clinic, Mayo College of Medicine; Rochester, MN, USA)
  - Magda Sanz Cortes (Texas Children's Fetal Center, Baylor College of Medicine, Department of Obstetrics and Gynaecology; Houston, TX, USA)
- **External reviewers**
  - Alexandra Benachi & Julien Saada (Service de Gynécologie-Obstétrique, AP-HP, & Centre Maladie Rare: Hernie de Coupole Diaphragmatique, Hôpital Antoine Béclère, Université Paris-Sud; Clamart, France)
- **Biostatistics**
  - Prakesh S Shah (Department of Paediatrics, Mount Sinai Hospital and University of Toronto and Maternal-Infant Care (MiCare) Center, Mount Sinai Hospital; Toronto, Canada)
  - Xiang Y. Ye (Maternal-Infant Care (MiCare) Center, Mount Sinai Hospital; Toronto, Canada)
  - Haleh Sangi-Haghpeykar (Department of Obstetrics and Gynaecology, Baylor College of Medicine, Houston; Houston, TX, USA)
- **Information technology support/Trice Imaging Inc. viewing software**
  - Aamir Siddiqui (Trice Imaging Inc.)

### Collaborators from the North American Fetal Therapy Network (NAFTNet)

- Aguilera M, Salaman DL, Division of Maternal-Fetal Medicine, Midwest Fetal Care center; Minneapolis, MN.
- Bahtiyar M, Division of Maternal- Fetal Medicine, Yale Fetal Care center, Yale School of Medicine; New Haven, CT
- Baschat AA, Miller JL, Center for Fetal Therapy, Department of Gynaecology and Obstetrics, Johns Hopkins
• Bebbington M, Division of Maternal-Fetal Medicine, Fetal Care Center, Washington University School of Medicine; St. Louis, MO
• Bennett KA, Newton M, Division of Maternal-Fetal Medicine, Vanderbilt Center for Women's Health, Vanderbilt University Medical Center; Nashville, TN
• Benson C, Department of Radiology, Brigham and Women’s Hospital and Harvard Medical School; Boston, MA
• Berman S, Treadwell M, Division of Maternal-Fetal Medicine, C.S. Mott Children’s Hospital, University of Michigan School of Medicine; Ann Arbor, MI
• Blumenfeld Y, Division of Maternal-Fetal Medicine, Stanford University School of Medicine; Palo Alto, CA
• Brown RN, Morency AM, Division of Maternal-Fetal Medicine, Royal Victoria Hospital, McGill University Health Center; Montreal, QC
• Coleman BG, Oliver ER, Center for Fetal Diagnosis and Treatment at the Children's Hospital of Philadelphia and the Perelman School of Medicine at University of Pennsylvania; Philadelphia, PA
• Carr S, Davis S, Division of Maternal-Fetal Medicine, Women's and Infants Hospital, Brown University; Providence, RI
• Center J, Markham K, Division of Maternal-Fetal Medicine, Wexner Medical Center, Ohio State University; Columbus, OH
• Chescheir N, Goodnight W, Division of Maternal-Fetal Medicine, North Carolina Women’s Hospital, University of North Carolina; Chapel Hill, NC
• Dashe J, Santiago-Munoz P, Division of Maternal-Fetal Medicine, UT Southwestern’s William P. Clements Jr. University Hospital, University of Texas Southwestern Medical Center; Dallas, TX
• Donepudi R, Papanna R, The Fetal Center, Children's Memorial Hermann Hospital, Department of Obstetrics and Gynaecology, McGovern Medical School, University of Texas Health Science Center; Houston, TX
• Drennan K, Division of Maternal-Fetal Medicine, Lattimore Medical Center, University of Rochester Medical Center; Rochester, NY
• Emery SP, Makaroun S, Division of Maternal-Fetal Medicine, Magee-Women’s Hospital, University of Pittsburgh School of Medicine; Pittsburgh, PA
• Estroff J, Department of Radiology, Boston Children’s Hospital and Harvard Medical School; Boston, MA
• Gagnon A, Tessier F, Division of Maternal-Fetal Medicine, British Columbia Women’s Hospital and Children’s Hospital of British Columbia, University of British Columbia; Vancouver, BC
• Goldstein R, Morgan T, The Fetal Treatment Center, Department of Radiology, University of California; San...
Francisco, CA

- Halabi S, Department of Radiology, Stanford Children’s Health, Stanford University; Palo Alto, CA
- Keunen J, Van Mieghem T, Fetal Medicine Unit, Mount Sinai Hospital, University of Toronto; Toronto, ON
- Lim FY, Polzin W, Fetal Care Center, Cincinnati Children's Hospital Medical Center; Cincinnati, OH
- Mehollin-Ray AR, Department of Radiology, Baylor College of Medicine; Houston, TX
- Meyers ML, Department of Paediatric Radiology, University of Colorado Denver, Anschutz Medical Centre; Denver, CA
- Miller R, Simpson L, Division of Maternal-Fetal Medicine, New York- Presbyterian & Morgan Stanley Children's Hospital of New York, Columbia University Medical Center; New York, NY
- Turan, O, Division of Maternal-Fetal Medicine, University of Maryland Medical System; Baltimore, MD
- Zaretsky MV, Division of Maternal Fetal Medicine Children's Hospital of Colorado, University of Colorado School of Medicine; Aurora, CO

**Local collaborators (Fetal Medicine Unit, Mount Sinai Hospital):**

- YM Lee (Lead sonographer educator, Fetal Medicine Unit)
- Sonographers/clinical fellows training in fetal medicine
<table>
<thead>
<tr>
<th>Heuristic principles (1)</th>
<th>Heuristic violation (Yes/No)</th>
<th>Explain violation and specify where violation occurs in protocol</th>
<th>Rate violation (Major/Minor$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintenance of consistency with current standards and convention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimalist design (i.e.: simple, easy to follow)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimization of memory load (i.e.: provides all necessary information and not relying on individuals to remember excessive facts and data)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevention of error (i.e.: addresses common errors and pitfalls)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear closure, with every task having a clear beginning and end</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of the user’s language (i.e.: simple vocabulary, easy to interpret text)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^\text{Rating of violation based on potential to lead to misinterpretation of steps involved in lung area measurement resulting in inaccurate lung area estimation.}$
Table 3. Plan-do-study-act (PDSA) cycles for developing a standardized sonographic protocol for prognostication of prenatally diagnosed CDH across NAFTNet

<table>
<thead>
<tr>
<th>PDSA cycle</th>
<th>PDSA stage</th>
<th>Plan/Prediction</th>
<th>Do</th>
<th>Study</th>
<th>Act</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Learning</td>
<td>We predict that there is variability in practice patterns for lung area measurement among NAFTNet FETO group</td>
<td>Qualitative surveys, videoconference calls/interview conducted among participants with direct observation of measurement technique and usability testing of current lung area measurement methods (i.e.: Anteroposterior (A-P), longest (L) and trace (T) method) on sonographic clips of CDH among NAFTNet FETO group</td>
<td>-Similar image selection criteria and technique for head circumference - Variability in image selection criteria for lung area assessment - Trace method preferred among most participants, despite varying institutional preferences -Many usability issues identified with AP and L - Poor agreement for image quality rating - kappa statistic: Chest: 0.05 (95% CI 0.03, 0.08) Lung: 0.12 (95% CI 0.1, 0.15)</td>
<td>-Key steps for optimal image selection and areas lacking consensus identified -Usability issues identified with current lung area measurement methods -Institutional barriers to change and key stakeholders identified</td>
</tr>
</tbody>
</table>
| 2          | Learning   | We predict that there is significant inter-observer variability in lung area estimation among NAFTNet FETO group, with the highest inter-rater agreement using the trace method | Establish baseline agreement (Intraclass correlation [ICC]) for lung area methods (i.e.: AP, L and T) and head circumference (HC) measurements on 13 sonographic clips of CDH among NAFTNet FETO group | ICC for measurements:  
• HC: 0.99 (95% CI 0.992, 0.998)  
• A-P: 0.83 (95% CI 0.67, 0.94)  
• L: 0.89 (95% CI 0.75, 0.97)  
• T: 0.94 (95% CI 0.83, 0.98)  
• Highest agreement between A-P and T (ICC 0.99; 95% CI 0.91, 0.996)) | -High inter-rater agreement for head circumference and unlikely to contribute to variability in o/e LHR -Highest inter-rater agreement for trace method among NAFTNet FETO consortium |
| 3 | **Learning** | As no gold standard exists to determine fetal lung volume, validity of measurement techniques may be addressed by comparing lung area measurement methods and head circumference between NAFTNet participants and an external reviewer (ER) from a European center with established expertise in CDH. Compare agreement in lung area measurements using AP, L and T methods between 19 NAFTNet reviewers and ER on 13 sonographic clips of CDH using Bland & Altman (B & A) plots. When comparing NAFTNet reviewers to ER, the bias (mean difference) was lowest for T and highest for L method. Bias (mean difference): A-P: 23 ± 50 mm², L: 54 ± 60 mm², T: 14 ± 38 mm². Trace method has the highest agreement for lung area measurement between NAFTNet and ER. Trace is the preferred method for lung area estimation. |
| 4 | **Development** (See Appendix NAFTNet o/e LHR protocol PDSA 4) | Using data collected from interviews, direct observation, usability testing and statistical analysis evaluating inter-rater agreement for HC and lung area measurement methods, the protocol proposed by Jani et al. (2) was modified accordingly by NAFTNet FETO consortium to improve standardization of o/e LHR. Consensus and protocol development at biannual NAFTNet Meeting (April 2017) with multidisciplinary representatives (i.e.: obstetrics, radiology and paediatric surgery) from each participating NAFTNet FETO center. Consensus reached with decision to perform all lung area assessments using trace method. Protocol proposed by Jani et al. (2) adapted to address areas lacking consensus identified during usability testing. Concerns raised regarding implementation of consensus statement among less experienced centers. Protocol will need to be adapted based on usability testing performed in NAFTNet centers outside of the FETO consortium. |
| 5 | **Learning** | Less experienced (Non-FETO) NAFTNet centers will have lower agreement for lung area measurements compared to FETO NAFTNet centers. Qualitative surveys, videoconference calls/ interviews with direct observation of measurement technique and usability testing of current lung area measurement methods conducted among NAFTNet non-FETO centers. Establish baseline agreement for lung area methods (i.e.: AP, Kappa statistic for image quality rating: Chest: 0.07 (95% CI 0.05, 0.08), Lung: 0.1 (95% CI 0.09, 0.12) (ICC) for measurements: HC: 0.83 (95% CI 0.78, 0.99), A-P: 0.54 (95% CI 0.1, 0.91). Trends similar between FETO and non-FETO centers with trace method demonstrating highest inter-rater agreement. Agreement overall lower in non-FETO group for all lung area methods as well as head circumference measurements. |
L, T) and HC measurements among NAFTNet non-FETO group and compare agreement in lung area measurements using 3 methods (AP, L and T) between non-FETO group and ER

- L: 0.57 (95% CI 0.12, 0.90)
- T: 0.86 (95% CI 0.79, 0.93)
- Highest agreement between A-P and trace (ICC 0.95; 95% CI 0.70-0.99)
- Highest lung area estimates with L

When comparing NAFTNet reviewers to ER, the bias (mean difference) was lowest for T and highest for L method

Bias (mean difference):
- A-P: 42 ± 49 mm²
- L: 62 ± 57 mm²
- T: 19 ± 36 mm²

6 **Development**  
(See Appendix NAFTNet o/e LHR protocol PDSA 6 & 7)

Additional usability issues and knowledge/skills gaps will be identified among fetal medicine centers of varying size and expertise, requiring further specifications within NAFTNet protocol

- Usability testing of lung area measurement methods among non-FETO centers to further understand challenges with current protocols among NAFTNet centers of varying size and expertise
- Knowledge gaps identified in relevance of o/e LHR as a prognostic indicator, number of measurements to obtain and which o/e LHR value to assign
- Usability issues identified with current measurement methods including optimal fetal position, avoidance of extrapulmonary tissue in lung area measurement
- Assessments generally made on still images rather than sweeps
- Further specifications within protocol (i.e.: optimal fetal position, use of axial sweep with color Doppler to aid identification of intrathoracic liver, number of images to acquire and use of average o/e LHR) to address identified knowledge gaps and usability issues
- Technique for head circumference measurement included due to lower inter-rater agreement among non-FETO centers
- Visual aid/sonographic image included

7 **Development**  
Less experienced physicians

- Heuristics analysis of protocol
- Main heuristic violations
- Tasks clarified
|   | Development/Implementation (See Appendix NAFTNet o/e LHR protocol PDSA 8) | Implementation of protocol among end users (i.e.: local sonographers and clinical fellows training in fetal medicine) may identify additional usability issues among frontline workers | - Usability testing among sonographers and clinical fellows training in fetal medicine | Additional heuristic violations:  
- “Minimalist design”, “Prevention of error”: wording/tasks unclear, text heavy  
- Usability issues identified with protocol: users still incorporating mediastinal vessels in lung area measurement | - Protocol adapted to clarify wording, simplify text  
- Mediastinal vessels identified on sonographic image to prevent users from including these structures in lung area measurement |}

|   |   | and sonographers may have difficulty implementing a protocol for o/e LHR measurement developed by “experts” | with core working group (N.A., A.J., G.R., R.R.) and lead educational sonographer and experienced sonographer locally identified:  
- “Every task having a clear beginning and an end”: several steps having several subtasks without clear closure  
- “Minimization of memory load” and “prevention of error”: many sonographers unaware of prognostic value and corresponding mortality risk with o/e LHR; sonographers noted to incorporate extra-pulmonary tissue in lung area estimation  
- “Consistency with current standards and convention”: no standards for competency, steps inconsistent with typical sequence for image acquisition  
- “Minimalist design”: excessive text | - Relevance of o/e LHR in prognostication included  
- Sonographic image showing accurate lung area measurement should be included  
- Competency addressed by making a recommendation based on the consensus statement for obstetric and gynaecological ultrasound competency assessment in residency training (3), which recommends obtaining 5 images meeting image criteria or a 'pass score' for each competency assessment/ultrasound imaging task.  
- Steps placed in sequence consistent with typical image acquisition in practice  
- Text simplified  
- Protocol will need to be operationalized by frontline workers (sonographers) |   |
| (Not completed due to acceptable agreement among NAFTNet centers) | standardized protocol for o/e LHR assessment will improve agreement in measurements across NAFTNet centers | interview participants with direct observation of measurement technique and usability testing of NAFTNet protocol  
- Obtain agreement for lung area (T) and HC among NAFTNet non-FETO group and compare agreement in lung area measurement (T) and HC between non-FETO group and ER on de-identified sonographic clips of CDH with both old and new cases | with protocol  
- Compare agreement pre- and post-standardization among non-FETO NAFTNet centers | protocol based on usability issues identified  
- Usability testing may be required locally at NAFTNet site to ensure successful implementation  
- May consider additional interventions like eLearning modules if poor implementation of protocol  
- Future initiatives to evaluate intrathoracic liver herniation by US and MRI and incorporate into a comprehensive antenatal prognostication protocol for CDH |
Table 4. Image selection criteria for lung area assessment in CDH across NAFTNet based on survey data

<table>
<thead>
<tr>
<th>Image criteria</th>
<th>All NAFTNet participants (n=49)</th>
<th>FETO centers (n=19)</th>
<th>Non-FETO centers (n=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four-chamber view of heart</td>
<td>43</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td>Axial view of the chest</td>
<td>28</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>Clear lung borders</td>
<td>27</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>Absence of shadowing from ribs, spine and extremities</td>
<td>19</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Optimal fetal position (Total)</td>
<td>19</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>• Lung close to transducer (i.e.: spine at 3 o’clock or 9 o’clock)</td>
<td>14</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>• Supine position (i.e.: spine at 6 o’clock, sternum at 12 o’clock)</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>• Not specified</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Image optimization (E.g.: appropriate gain, depth, focal point, magnification, dynamic range)</td>
<td>11</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Visualization of descending aorta</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 5. Advantages and usability issues with lung area measurement methods according to NAFTNet participants

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Usability issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anteroposterior diameter</td>
<td>“Originally described.”</td>
<td>“Where do I put the calipers?”</td>
</tr>
<tr>
<td>method</td>
<td></td>
<td>“I don’t know what landmarks to use.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“How do I do this?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“I never use this method.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Underestimate?”</td>
</tr>
<tr>
<td>Longest diameter method</td>
<td>“Used in early publications.”</td>
<td>“Irregular shaped lungs?”</td>
</tr>
<tr>
<td></td>
<td>“Most comfortable.”</td>
<td>“Avoid traversing the atria.”</td>
</tr>
<tr>
<td></td>
<td>“Easy--just take the longest measurement.”</td>
<td>“Can overestimate.”</td>
</tr>
<tr>
<td></td>
<td>“No specific landmarks needed for calipers.”</td>
<td></td>
</tr>
<tr>
<td>Trace method</td>
<td>“Most reproducible.”</td>
<td>“Control needed.”</td>
</tr>
<tr>
<td></td>
<td>“Evidence-based.”</td>
<td>“Can overshoot.”</td>
</tr>
<tr>
<td></td>
<td>“Easy.”</td>
<td>“May include mediastinal vessels, bowel and liver.”</td>
</tr>
<tr>
<td></td>
<td>“Makes more sense.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Less subjective.”</td>
<td>“Lung borders must be clear.”</td>
</tr>
<tr>
<td></td>
<td>“Can compare images easily.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Take in more lung volume.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Useful with irregular lung shape.”</td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Agreement among NAFTNet participants for image quality rating of lung and head sonographic clips in fetal left CDH

<table>
<thead>
<tr>
<th></th>
<th>Image quality categories$^\S$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(95% CI 0.07, 0.2)</td>
<td>(95% CI -0.04, 0.04)</td>
<td>(95% CI 0.01, 0.1)</td>
<td>(95% CI 0.03, 0.1)</td>
<td>(95% CI 0.03, 0.08)</td>
</tr>
<tr>
<td>Lung sonographic</td>
<td>FETO centers†</td>
<td>0.1</td>
<td>0</td>
<td>0.06</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>clips (kappa)</td>
<td></td>
<td>(95% CI 0.1, 0.2)</td>
<td>(95% CI -0.04, 0.04)</td>
<td>(95% CI 0.01, 0.1)</td>
<td>(95% CI 0.03, 0.1)</td>
<td>(95% CI 0.03, 0.08)</td>
</tr>
<tr>
<td></td>
<td>Non-FETO centers¶</td>
<td>0.2</td>
<td>0.01</td>
<td>0.06</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(95% CI 0.2, 0.3)</td>
<td>(95% CI 0.02, 0.03)</td>
<td>(95% CI 0.04, 0.09)</td>
<td>(95% CI 0.01, 0.07)</td>
<td>(95% CI 0.05, 0.08)</td>
</tr>
<tr>
<td>Head sonographic</td>
<td>FETO centers†</td>
<td>0.2</td>
<td>0.03</td>
<td>0.1</td>
<td>0.07</td>
<td>0.1</td>
</tr>
<tr>
<td>clips (kappa)</td>
<td></td>
<td>(95% CI 0.2, 0.3)</td>
<td>(95% CI 0.01, 0.08)</td>
<td>(95% CI 0.1, 0.2)</td>
<td>(95% CI 0.03, 0.1)</td>
<td>(95% CI 0.1, 0.2)</td>
</tr>
<tr>
<td></td>
<td>Non-FETO centers¶</td>
<td>0.3</td>
<td>0.01</td>
<td>0.1</td>
<td>0.04</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(95% CI 0.2, 0.3)</td>
<td>(95% CI -0.02, 0.03)</td>
<td>(95% CI 0.09, 0.1)</td>
<td>(95% CI 0.01, 0.07)</td>
<td>(95% CI 0.09, 0.1)</td>
</tr>
</tbody>
</table>

$^\S$ Image quality categories: 1 = Excellent/ optimal image quality that allows accurate measurement; 2 = Good / good image quality, sufficient to obtain measurements; 3 = Fair/ suboptimal image quality, just sufficient to obtain measurements; 4 = Poor/ unacceptable image quality to obtain measurements.

†Inter-rater kappa statistic was used to assess inter-rater agreement for image quality assessment of lung and head sonographic clips among 19 physicians from NAFTNet FETO consortium centers

¶Inter-rater kappa statistic was used to assess inter-rater agreement for image quality assessment of lung and head sonographic clips among 29 physicians from NAFTNet centers outside the FETO consortium
Table 7. Agreement among NAFTNet participants for head circumference and lung area measurements using anteroposterior, longest and trace method in fetal left CDH

<table>
<thead>
<tr>
<th>Agreement (ICC)</th>
<th>Head Circumference</th>
<th>Anteroposterior</th>
<th>Longest</th>
<th>Trace</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FETO centers†</strong></td>
<td>0.99 (95% CI 0.992, 0.998)</td>
<td>0.83 (95% CI 0.67, 0.94)</td>
<td>0.89 (95% CI 0.75, 0.97)</td>
<td>0.94 (95% CI 0.83, 0.98)</td>
</tr>
<tr>
<td><strong>Non-FETO centers¶</strong></td>
<td>0.83 (95% CI 0.78, 0.99)</td>
<td>0.54 (95% CI 0.1, 0.91)</td>
<td>0.57 (95% CI 0.12, 0.90)</td>
<td>0.86 (95% CI 0.79, 0.93)</td>
</tr>
</tbody>
</table>

†Intraclass correlation (ICC) was used to assess inter-rater agreement for measurement of lung area and head circumference among 19 physicians from NAFTNet FETO consortium centers
¶Intraclass correlation (ICC) was used to assess inter-rater agreement for measurement of lung area and head circumference among 29 physicians from NAFTNet centers outside the FETO consortium
Table 8. Agreement between different lung area measurement methods among NAFTNet participants in fetal left CDH

<table>
<thead>
<tr>
<th></th>
<th>Anteroposterior and longest</th>
<th>Anteroposterior and trace</th>
<th>Longest and trace</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FETO centers†</strong> (ICC)</td>
<td>0.85 (95% CI 0.54, 0.90)</td>
<td>0.99 (95% CI 0.91, 0.996)</td>
<td>0.83 (95% CI 0.26, 0.90)</td>
</tr>
<tr>
<td><strong>Non-FETO centers¶</strong> (ICC)</td>
<td>0.87 (95% CI 0.55, 0.92)</td>
<td>0.95 (95% CI 0.70, 0.99)</td>
<td>0.80 (95% CI 0.28, 0.89)</td>
</tr>
</tbody>
</table>

†Intraclass correlation (ICC) was used to assess inter-rater agreement between different lung area methods among 19 physicians from NAFTNet FETO consortium centers.
¶Intraclass correlation (ICC) was used to assess inter-rater agreement between different lung area methods among 29 physicians from NAFTNet centers outside the FETO consortium.
Table 9. Agreement between each NAFTNet participant from the FETO consortium and the external reviewer by lung area measurement method in fetal left CDH

<table>
<thead>
<tr>
<th>NAFTNet Reviewer</th>
<th>Anteroposterior (mm²)</th>
<th>Longest (mm²)</th>
<th>Trace (mm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias ± std</td>
<td>ICC (95% CI)</td>
<td>Bias ± std</td>
</tr>
<tr>
<td>1</td>
<td>10.9±51.6</td>
<td>0.87 (0.47, 0.97)</td>
<td>52.1±99.8</td>
</tr>
<tr>
<td>2</td>
<td>-52.4±84.4</td>
<td>0.61 (0.51, 0.93)</td>
<td>4.6±68</td>
</tr>
<tr>
<td>3</td>
<td>23.2±64.7</td>
<td>0.87 (0.31, 0.96)</td>
<td>8.9±65.7</td>
</tr>
<tr>
<td>4</td>
<td>7±52.5</td>
<td>0.91 (0.66, 0.98)</td>
<td>31.2±89.9</td>
</tr>
<tr>
<td>5</td>
<td>49.1±76.9</td>
<td>0.85 (0.44, 0.95)</td>
<td>50.7±85</td>
</tr>
<tr>
<td>6</td>
<td>42.4±84.9</td>
<td>0.83 (0.48, 0.93)</td>
<td>54.2±68</td>
</tr>
<tr>
<td>7</td>
<td>52.3±61.4</td>
<td>0.92 (0.51, 0.96)</td>
<td>69.4±65.8</td>
</tr>
<tr>
<td>8</td>
<td>35.4±51.7</td>
<td>0.92 (0.59, 0.97)</td>
<td>91.2±67.8</td>
</tr>
<tr>
<td>9</td>
<td>52.6±91.8</td>
<td>0.77 (0.38, 0.95)</td>
<td>70±133.5</td>
</tr>
<tr>
<td>10</td>
<td>40.3±62.6</td>
<td>0.85 (0.45, 0.95)</td>
<td>40.6±56.6</td>
</tr>
<tr>
<td>11</td>
<td>82.3±125.9</td>
<td>0.65 (0.39, 0.98)</td>
<td>138.6±143</td>
</tr>
<tr>
<td>12</td>
<td>51.6±52.8</td>
<td>0.89 (0.51, 0.98)</td>
<td>95.3±80.6</td>
</tr>
<tr>
<td>13</td>
<td>11.5±34.7</td>
<td>0.94 (0.70, 0.97)</td>
<td>39.2±51.8</td>
</tr>
<tr>
<td>14</td>
<td>-45.6±133.1</td>
<td>0.21 (0.0, 0.85)</td>
<td>9.4±75.7</td>
</tr>
<tr>
<td>15</td>
<td>17.7±54.8</td>
<td>0.90 (0.57, 0.99)</td>
<td>77.5±88.4</td>
</tr>
<tr>
<td>16</td>
<td>18.6±69.5</td>
<td>0.84 (0.37, 0.95)</td>
<td>55.8±80.9</td>
</tr>
<tr>
<td>17</td>
<td>-44.7±73.1</td>
<td>0.69 (0.60, 0.94)</td>
<td>11.1±39.8</td>
</tr>
<tr>
<td>18</td>
<td>46.4±64.9</td>
<td>0.87 (0.60, 0.97)</td>
<td>56.1±74.7</td>
</tr>
<tr>
<td>19</td>
<td>31.7±49.9</td>
<td>0.91 (0.38, 0.98)</td>
<td>73.2±55.6</td>
</tr>
</tbody>
</table>
Table 10. Agreement between each NAFTNet participant from centers outside of the FETO consortium and the external reviewer by lung area measurement method in fetal left CDH

<table>
<thead>
<tr>
<th>NAFTNet Reviewer</th>
<th>Anteroposterior (mm²)</th>
<th>Longest (mm²)</th>
<th>Trace (mm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias ± std</td>
<td>ICC (95% CI)</td>
<td>Bias ± std</td>
</tr>
<tr>
<td>1</td>
<td>3.8±53.1</td>
<td>0.87 (0.45,0.97)</td>
<td>33.5±49.1</td>
</tr>
<tr>
<td>2</td>
<td>19.5±64.9</td>
<td>0.83 (0.8,0.96)</td>
<td>44.7±133</td>
</tr>
<tr>
<td>3</td>
<td>132±132.9</td>
<td>0.66 (0.53,0.95)</td>
<td>120.8±121.9</td>
</tr>
<tr>
<td>4</td>
<td>41.4±65.9</td>
<td>0.86 (0.64,0.97)</td>
<td>49.5±60.7</td>
</tr>
<tr>
<td>5</td>
<td>54.5±71.9</td>
<td>0.8 (0.55,0.95)</td>
<td>59.6±98.9</td>
</tr>
<tr>
<td>6</td>
<td>74.8±81.4</td>
<td>0.79 (0.69,0.99)</td>
<td>114.5±128.3</td>
</tr>
<tr>
<td>7</td>
<td>101.5±72.2</td>
<td>0.85 (0.5,0.95)</td>
<td>121.1±86.4</td>
</tr>
<tr>
<td>8</td>
<td>32.5±71.2</td>
<td>0.8 (0.72,0.96)</td>
<td>79.8±65.1</td>
</tr>
<tr>
<td>9</td>
<td>54.7±66.1</td>
<td>0.87 (0.63,0.97)</td>
<td>68.6±68.7</td>
</tr>
<tr>
<td>10</td>
<td>49.4±65.3</td>
<td>0.87 (0.59,0.99)</td>
<td>95.3±70.3</td>
</tr>
<tr>
<td>11</td>
<td>106.9±95.6</td>
<td>0.81 (0.65,0.93)</td>
<td>127.7±85.6</td>
</tr>
<tr>
<td>12</td>
<td>4.5±53.7</td>
<td>0.91 (0.78,0.98)</td>
<td>-16.9±71</td>
</tr>
<tr>
<td>13</td>
<td>68.3±69.8</td>
<td>0.86 (0.59,0.97)</td>
<td>75.9±93.5</td>
</tr>
<tr>
<td>14</td>
<td>15.5±95.2</td>
<td>0.67 (0.52,0.89)</td>
<td>39.8±90.7</td>
</tr>
<tr>
<td>15</td>
<td>6.4±49.4</td>
<td>0.92 (0.78,0.98)</td>
<td>103.4±88.1</td>
</tr>
<tr>
<td>16</td>
<td>123.6±374.4</td>
<td>0.1 (0.09,0.98)</td>
<td>130.9±395</td>
</tr>
<tr>
<td>17</td>
<td>29.7±102.6</td>
<td>0.72 (0.65,0.99)</td>
<td>44±144.5</td>
</tr>
<tr>
<td>18</td>
<td>-16.8±96</td>
<td>0.57 (0.53,0.86)</td>
<td>-30.5±115.2</td>
</tr>
<tr>
<td>19</td>
<td>16.9±54</td>
<td>0.91 (0.74,0.99)</td>
<td>65.6±54.6</td>
</tr>
<tr>
<td>20</td>
<td>-10.2±48.6</td>
<td>0.88 (0.81,0.96)</td>
<td>-13±87</td>
</tr>
<tr>
<td>21</td>
<td>42.5±97</td>
<td>0.6 (0.58,0.89)</td>
<td>57.6±159.9</td>
</tr>
<tr>
<td>22</td>
<td>164.2±370.9</td>
<td>0.14 (0.1,0.95)</td>
<td>169.7±395.6</td>
</tr>
<tr>
<td>23</td>
<td>11.1±54.2</td>
<td>0.87 (0.76,0.97)</td>
<td>28.5±93.4</td>
</tr>
<tr>
<td>24</td>
<td>68.7±69.8</td>
<td>0.86 (0.5,0.97)</td>
<td>87.9±52.9</td>
</tr>
<tr>
<td>25</td>
<td>-6.8±36.7</td>
<td>0.94 (0.75,0.99)</td>
<td>5.5±75.7</td>
</tr>
<tr>
<td>26</td>
<td>39.2±56.9</td>
<td>0.89 (0.45,0.98)</td>
<td>76.6±103</td>
</tr>
<tr>
<td>27</td>
<td>-5.6±58.2</td>
<td>0.85 (0.79,0.95)</td>
<td>-12±56.8</td>
</tr>
<tr>
<td>28</td>
<td>3.5±46.9</td>
<td>0.92 (0.6,0.98)</td>
<td>20.3±101.1</td>
</tr>
<tr>
<td>29</td>
<td>-17.4±78</td>
<td>0.66 (0.39,0.91)</td>
<td>37.7±82.5</td>
</tr>
</tbody>
</table>
Table 11. Results of heuristic analysis for NAFTNet o/e LHR protocol by core working group and sonographers

<table>
<thead>
<tr>
<th>Heuristic principle (1)</th>
<th>Heuristic violation (Yes/No)</th>
<th>Explain violation and specify where violation occurs in protocol</th>
<th>Rate violation (Major/Minor)</th>
<th>Protocol Modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintenance of consistency with current standards and convention</td>
<td>Yes</td>
<td>-Competency not addressed, which is essential to ensure accurate lung area measurement. Protocol by Jani J. (2) recommends that sonographers measure 70 LHR to ensure competency based on a study looking at the learning curve for LHR measurement in trainees (4). However, this number was felt to be unrealistic and inconsistent with competency assessment for other ultrasound parameters. -Steps do not follow typical sequence of image acquisition in practice (e.g.: after obtaining an axial section of the chest, adequate fetal position should be ensured prior to image optimization; importance of obtaining axial sweeps of chest should follow lung area measurement description rather than following head measurement)</td>
<td>Major (may result in inaccurate lung area assessments)</td>
<td>Training recommended with an experienced sonographer, with 5 images meeting criteria in NAFTNet protocol required to ensure competency. Recommendations based on expert international consensus statement for obstetrics and gynaecological ultrasound competency assessment in residency training (3). Sequence of steps modified (see Appendix NAFTNet o/e LHR protocol PDSA 6, 7 &amp; 8 and final protocol)</td>
</tr>
<tr>
<td>Minimalist design</td>
<td>Yes</td>
<td>Excess text/ “wordiness” Section on prognostic value of o/e LHR unclear, “difficult to follow and read quickly” and “wordy”.</td>
<td>Minor</td>
<td>Wording simplified and text minimized, font size increased and text spaced out (see Appendix NAFTNet o/e LHR protocol PDSA 6, 7 &amp; 8 and final protocol)</td>
</tr>
<tr>
<td>Minimization of memory load</td>
<td>Yes</td>
<td>o/e LHR severity classification and associated mortality difficult to recall for many practitioners</td>
<td>Major (may result in incorrect severity)</td>
<td>Severity classification and associated survival included (see Appendix NAFTNet o/e LHR protocol PDSA 8)</td>
</tr>
<tr>
<td>Prevention of error</td>
<td>Yes</td>
<td>Incorrect interpretation and unclear wording of steps -Inclusion of extra-pulmonary tissue in lung area estimate. Optimal fetal position unclear</td>
<td>Major (may result in incorrect lung area measurement, optimal fetal position and</td>
<td>-Sonographic image with optimal lung area measurement, optimal fetal position and</td>
</tr>
<tr>
<td>Violation</td>
<td>Rating</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 4 incorrectly worded: The image should be frozen prior to final magnification to ensure landmarks are clearly visible.</td>
<td>Minor</td>
<td>Step 4 clarified as follows: “The image should be optimized prior to final magnification to ensure landmarks are clearly visible.”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tasks unclear and difficult to follow and composed of several subtasks within each step.</td>
<td>Major (may result in inaccurate lung area)</td>
<td>Tasks simplified, and broken down into smaller tasks (see Appendix NAFTNet o/e LHR protocol PDSA 6, 7 &amp; 8 and final protocol)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unclear wording</td>
<td>Minor</td>
<td>Wording clarified: “Avoid rib shadowing by scanning through an intercostal space.”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

§Rating of violation based on potential to lead to misinterpretation of steps involved in lung area measurement resulting in inaccurate lung area estimation.
Table 12. Results of usability testing on NAFTNet o/e LHR protocol with representative end-users

<table>
<thead>
<tr>
<th>Image selection criteria</th>
<th>Prior to introduction of protocol (n=7*)</th>
<th>Post introduction of protocol (n=7*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-chamber view</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Axial section of the chest</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Clear lung borders</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Fetal position</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Image optimization</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

**Lung area measurement using trace method**

| Calipers placed over contralateral lung area | 7                                      | 7                                  |
| Include tissue other than lung (inclusion of mediastinal vessels) | 5                                      | 2 (inclusion of mediastinal vessels) |
References

Figures

Figure 1. Sonographic lung area measurement methods in prenatally diagnosed CDH.

Figure 2. Number of CDH cases seen annually prenatally in NAFTNet centers

Figure 3. Stakeholder mapping & team building

Figure 4. Ishikawa diagram: sources of variability in lung area estimation

Figure 5. Two-dimensional sonographic axial images of varying image quality obtained from fetuses with left CDH at the level of the four-chamber view of the heart.

Figure 6. Process map for lung area and head circumference measurement in CDH

Figure 7. Sonographic measurement of fetal head circumference

Figure 8. Calculators used for o/e LHR estimation in CDH prognostication across NAFTNet. Results are grouped according to whether participants are from centers within and outside the FETO consortium

Figure 9. Lung area measurement method preferences among NAFTNet participants. Results are grouped according to whether participants are from centers within and outside the FETO consortium

Figure 10: Mean lung area per CDH study using anteroposterior, longest and trace methods among participants from NAFTNet FETO centers

Figure 11: Mean lung area per CDH study using anteroposterior, longest and trace methods among participants from NAFTNet centers outside the FETO consortium

Figure 12: Agreement between NAFTNet reviewers and external reviewer using trace method per CDH study (NAFTNet FETO centers)

Figure 13: Agreement between NAFTNet reviewers and external reviewer using longest method per CDH study (NAFTNet FETO centers)

Figure 14: Agreement between NAFTNet reviewers and external reviewer using anteroposterior method per CDH study (NAFTNet FETO centers)

Figure 15: Agreement between NAFTNet reviewers and external reviewer using trace method per CDH study (NAFTNet centers outside the FETO consortium)

Figure 16: Agreement between NAFTNet reviewers and external reviewer using longest method per CDH study (NAFTNet centers outside the FETO consortium)
Figure 17: Agreement between NAFTNet reviewers and external reviewer using anteroposterior method per CDH study (NAFTNet centers outside the FETO consortium)
Figure 1. Sonographic lung area measurement methods in prenatally diagnosed CDH. Two-dimensional sonographic axial view of the chest at the level of the four-chamber view of the heart in a fetus with left sided CDH. The three published methods for lung area measurement are illustrated: (a) Antero-posterior (AP) method: multiplying the AP diameter of the lung at the mid-clavicular line by the transverse, perpendicular diameter at the midpoint of the AP diameter method; (b) Longest diameter method: multiplying the longest diameter of the lung by its longest perpendicular diameter; (c) Trace method: manually tracing the outer margins of the lungs.
Figure 2. Number of CDH cases seen annually prenatally in NAFTNet centers

CDH annual caseload per participating NAFTNet centers

- ≤5 cases: Non-FETO (n=17), FETO (n=8)
- >5-10 cases: Non-FETO (n=17), FETO (n=8)
- >10-15 cases: Non-FETO (n=17), FETO (n=8)
- >15-20 cases: Non-FETO (n=17), FETO (n=8)
- >20 cases: Non-FETO (n=17), FETO (n=8)
Figure 3. Stakeholder mapping & team building
Figure 4. Ishikawa diagram: sources of variability in lung area estimation

Fishbone/Ishikawa diagram: Sources of variability in sonographic estimation of lung area in CDH

Policies:
- Standardized protocol?
- Audits/feedback?

Staff:
- Individual center preferences
- Differences in training/experience

Equipment:
- Visual aids while scanning for measurement criteria
- Ultrasound machine, technology

Procedure:
- Image selection criteria
- Lung area measurement method

Patients:
- LHR vs. o/e LHR
- Calculators/reference for o/e LHR

Information:
- Coexisting anomalies
- Left vs. Right CDH, severity, intrathoracic liver
- Maternal habitus
- Amniotic fluid abnormalities
- Fetal position
- Gestational age

Potential areas for interventions

Inter-observer variability for lung area measurement in CDH
Figure 5. Two-dimensional sonographic axial images of varying image quality obtained from fetuses with left CDH at the level of the four-chamber view of the heart. Case 1 (a) is a well-optimized image axial view of the chest with a 4-chamber view of the heart and contralateral lung close to the transducer in absence of shadowing. Case 2 and 3 are suboptimal images for fetal lung area measurement due to poor fetal position with contralateral lung far from the transducer (b) and maternal obesity with an oblique view of the chest (c).
Figure 6. Process map for lung area and head circumference measurement in CDH
Figure 7. Sonographic measurement of fetal head circumference.

A. Head circumference is obtained by using the elliptical tool. B. Head circumference is obtained by measuring the biparietal diameter (BPD) by placing the calipers from the “outer” leading edge of the skull to the “inner” skull border and the occipito-frontal diameter (OFD), by placing the calipers perpendicular to the BPD in the middle of frontal and occipital skull. Head circumference is derived from the following formula: \([\text{OFD (mm)} + \text{BPD (mm)}] \times 1.62\).
Figure 8. Calculators used for o/e LHR estimation in CDH prognostication across NAFTNet. Results are grouped according to whether participants are from centers within and outside the FETO consortium.

Calculators used for o/e LHR estimation by NAFTNet reviewers (n=46*)

* 2 participants were excluded, as they do not use o/e LHR in clinical practice, 1 non-responder
Figure 9. Lung area measurement method preferences among NAFTNet participants. Results are grouped according to whether participants are from centers within and outside the FETO consortium.
**Figure 10:** Mean lung area per CDH study using anteroposterior, longest and trace methods among participants from NAFTNet FETO centers
Figure 11: Mean lung area per CDH study using anteroposterior, longest and trace methods among participants from NAFTNet centers outside the FETO consortium.
Figure 12: Agreement between NAFTNet reviewers and external reviewer using trace method per CDH study (NAFTNet FETO centers)
**Figure 13:** Agreement between NAFTNet reviewers and external reviewer using longest method per CDH study (NAFTNet FETO centers)
Figure 14: Agreement between NAFTNet reviewers and external reviewer using anteroposterior method per CDH study (NAFTNet FETO centers)
**Figure 15:** Agreement between NAFTNet reviewers and external reviewer using trace method per CDH study (NAFTNet centers outside the FETO consortium)
Figure 15: Agreement between NAFTNet reviewers and external reviewer using longest method per CDH study (NAFTNet centers outside the FETO consortium)
**Figure 16:** Agreement between NAFTNet reviewers and external reviewer using anteroposterior method per CDH study (NAFTNet centers outside the FETO consortium)
Appendices

A. Letter to reviewers

B. Participant consent form

C. Questionnaire

D. Data collection sheet

E. Usability testing script

F. Usability data collection sheet for FETO group

G. Usability data collection sheet for non-FETO group

H. Usability data collection sheet for sonographers and clinical fellows in training

I. NAFTNet o/e LHR protocol PDSA 4

J. NAFTNet o/e LHR protocol PDSA 6 & 7

K. NAFTNet o/e LHR protocol PDSA 8

L. NAFTNet protocol for sonographic measurement of o/e LHR for antenatal prognostication of CDH
Appendix A. Letter to reviewers
September 25, 2017

Dear Colleagues,

You are receiving this email because you have agreed to participate in the “NAFTNet sonographic standardization of CDH prognostication project”.

As you may recall the primary objective of this study is to evaluate inter-observer variability in the measurement of the o/e LHR.

This is a very important project as centers prepare to join the ongoing TOTAL trial. It is also important, as some centers have noted that their survival outcome data in the most severe CDH group seems to have improved over the last 5-8 years. It is paramount that we compare “like” with “like”, so that data from different NAFTNet centers can be exchanged.

As a participant in this study, you will be asked to review 16 studies with 2 clips each, to assess lung area and head circumference. The clips and measurement software will be sent to you by Aamir Siddiqi from Trice Imaging.

We would like you to obtain the following measurements:

1. **Lung area (LA)** using the three different published methods (i.e.: Trace, longest and Anteroposterior (AP) diameter) for LA estimation as demonstrated below (Jani J. et. Ultrasound Obstet Gynecol. 2012;39(1):2-6):

2. **Biparietal diameter and an occipito-frontal diameter** on the fetal head images. Head circumference (HC) will be derived from these values.

You will also be asked to grade the quality of the images according to the following scale:

1. **Excellent** (Optimal image quality that allows accurate determination of LA and HC)
2. **Good** (Good image quality, sufficient to determine LA and HC)
3. **Fair** (Suboptimal image quality, just about sufficient to determine LA and HC)
4. **Poor** (Unacceptable image quality for LA and HC)

* For images that you have graded as “fair” or “poor”, we would like for you to explain why you have done so.

Finally, we would also like you to indicate the frame # of the image(s) which you selected to perform measurements. All measurements should be recorded on the attached excel spreadsheet. In our experience, it will take you ~ 2-2.5 hours to review and score all of the images.
Centers will also be contacted by Nimrah Abbasi, one of our team members, who is a recent graduate from the MFM Fellowship at the University of Toronto, and is currently completing an MSc degree in the field of quality improvement and patient safety, with a specific interest in improving quality and standardization in obstetric ultrasound.

As a part of an effort to formalize and standardize sonographic prognostication of prenatally diagnosed CDH, it would be ideal if Nimrah could interview as many reviewers as possible during the measurement process. This will allow us to better understand and deconstruct how sonographers select and assess images and obtain measurements, which can then be used to formulate a protocol for obtaining these measurements and for training other sonographers. These interviews can be organized via videoconference (Skype, Facetime or GoToMeeting) calls at a time that is convenient for you and should take ~20 minutes.

All documents should be returned to Nimrah via email at Nimrah.abbasi@sinahealthsystem.ca. All identifying information regarding site and reviewer will be removed before statistical analysis and any information collected during interviews will be anonymized.

We would also like you to complete a survey related to your experience with measuring o/e LHR and grading stomach position in CDH prognostication administered via Survey Monkey, which will be sent out in a separate email from Nimrah. Again, all information collected during the course of this study will be de-identified and kept confidential.

Aamir Siddiqi from Trice imaging will also be contacting you shortly to help review the process of how to download clips onto your center’s reporting system. If any further technical questions arise over the course of the study, please email him directly at aamir@triceimaging.com.

Our proposed timeline for this study is as follows:

- Complete questionnaire and spreadsheet by October 20, 2017.
- Videoconference calls to be completed with centers by November 3, 2017
- Data analysis will be completed by October 30, 2017

This study has been approved by the Mayo Clinic College of Medicine, Baylor College of Medicine IRB, Mount Sinai Hospital (Toronto, Canada) REB and the Steering Committee of the NAFTNet. The clips have been de-identified by the PI (R. Ruano & TRICE imaging). The study is supported by NAFTNet, as such the publication guidelines will follow network guidelines. [https://www.naftnet.org/Forms/tabid/163/Default.aspx](https://www.naftnet.org/Forms/tabid/163/Default.aspx)

The attachment outlines the guidelines for this project.

We would like to thank you for agreeing to participate in this project and look forward to working together. If there are any questions or concerns, please feel free to contact us.

Sincerely,
Nimrah Abbasi, Tony Johnson, Magda Sanz, Rodrigo Ruano & Greg Ryan
Appendix B. Participant consent form
CONSENT TO PARTICIPATE IN A RESEARCH STUDY

Title: Standardization of Prenatal Sonographic Prognostication of Fetuses with Congenital Diaphragmatic Hernia (CDH) Across the North America Fetal Therapy Network (NAFTNet)

Investigator: Dr. Nimrah Abbasi

Co-Investigators: Dr. Greg Ryan, Dr. Rodrigo Ruano, Dr. Anthony Johnson, Dr. Magda Sanz

Introduction
You are being asked to take part in a research study with the primary objective of standardizing antenatal sonographic prognostication of congenital diaphragmatic hernia (CDH) across fetal medicine centers. Please read this explanation about the study before deciding to participate. Please feel free to ask the study investigator to explain anything that you do not understand and make sure that all of your questions have been answered prior to signing this consent form.

Participation in this study is voluntary.

Background and Purpose
The primary objective is to evaluate inter-observer variability in the measurement of o/e LHR and stomach position grading (as a surrogate for liver herniation estimation) amongst NAFTNet centers. This is a very important project as centers prepare to join the ongoing TOTAL trial. It is also important, as some centers have noted that their survival outcome data in the most severe CDH group seems to have improved over the last 5-8 years. It is paramount that we compare “like” with “like”, so that data from different NAFTNet centers can be exchanged.

Study Design
Participation in this study will entail completion of a survey addressing how participants use antenatal sonographic prognostication in the management of fetal CDH, performing lung area and head circumference measurements and grading stomach position on de-identified sonographic video clips of fetal CDH, and for some, an interview via video-conference calls with the investigator (NA), during which practice patterns and measurement process will be discussed. The interviews should take approximately 20-30 minutes. The survey will be administered via Survey Monkey and may be accessed following agreement to participate in the study. This should take about 5 minutes to complete.

For the measurements, you will be asked to review 16 studies with 2 clips each, to assess lung area, stomach position and head circumference. Measurements will be recorded on the data collection sheet attached to the original email. The collection sheet will be returned to a single co-investigator (NA) via email, and all identifying information, regarding site and reviewer, will be removed and anonymized data will be compiled on a single data collection sheet prior to proceeding with statistical analysis. In our experience, it will take you ~ 2-2.5 hours to review and score all of the images.

Risks and Benefits to Being in the Study
There are no known risks to participating in the study. You may not receive any direct benefit from being in this study; however, information learned from this study will help us develop a standardized protocol for CDH antenatal prognostication by the NAFTNet centres.
Voluntary Participation
Your participation in this study is voluntary. You may decide not to be in this study, or to be in the study now and then change your mind later. You may leave the study at any time. You may refuse to answer any question you do not wish to answer, or not answer an interview question by saying “pass”. We will share any new information with you that is learned during the study that might affect your decision to stay in the study.

Confidentiality
The information that is collected for the study will be kept in a locked and secure area by the study investigator for 7 years. Only the study team and the Mount Sinai Hospital Research Ethics Board will be allowed to look at study records. The Mount Sinai Hospital Research Ethics Board has access to check that the information collected for this study is correct and to make sure the study followed proper laws and guidelines.

All information collected during this study will be kept confidential and will not be shared with anyone outside the study, unless required to do so by law. Any information about you that is sent out of the hospital will be coded and will not show your name, address or any information that identifies you. You will not be named in any reports, publications or presentations that may come from this study. If you decide to leave the study, the information about you that was collected before you left the study will still be used. No new information will be collected without your permission.

Questions About the Study
If you have any questions, concerns or would like to speak to the study team for any reason, please contact any of the principle investigators: Dr. Nimrah Abbasi at 514-605-6076, Dr. Greg Ryan at 416-586-8415, Dr. Rodrigo Ruano at 832-416-3310, Dr. Magda Sanz at 832-270-1471 or Dr. Anthony Johnson at 713-204-4949.

If you have any questions about your rights as a research participant or have concerns about this study, please contact Ronald Heslegrave, Ph.D., Chair of the Mount Sinai Hospital Research Ethics Board (REB) or the Research Ethics office number at 416-586-4875. The REB is a group of people who oversee the ethical conduct of research studies and are not part of the study team. Everything that you discuss will be kept confidential.

Consent
By signing below, I confirm that this study has been explained to me and my questions have been answered. I agree to participate in this study and I understand that I may leave the study at any time.

Printed name: ________________________________

Signature: ________________________________

Date: ___________________________
Appendix C. Questionnaire
As participants of the **NAFTNet sonographic standardization of CDH prognostication project**, we would like to ask you the following questions regarding your experience with sonographic assessment of CDH:

1. Please enter your name and institution of practice

2. What is your clinical subspecialty (e.g.: maternal-fetal medicine, radiology, pediatric surgery etc.)?

3. Which is your preferred method for estimating lung area?
   a. Trace method
   b. Longest diameter method
   c. Anterior-posterior (AP) diameter method

4. When assigning an o/e LHR, which value do you usually use?
   a. Average
   b. Highest
   c. Lowest

5. On average, how many lung area measurements do you obtain to calculate the o/e LHR?

6. How relevant is the o/e LHR value when prognosticating CDH severity in your clinical practice?
   a. Irrelevant
   b. Somewhat relevant
   c. Moderately relevant
   d. Highly relevant

7. How would you rate the **inter**-observer variability for the o/e LHR value amongst experienced sonographers (i.e.: reproducibility of o/e LHR measurement among different sonographers)?
   a. High inter-observer variability
   b. Some inter-observer variability, but unlikely to be significant
   c. Insignificant inter-observer variability
   d. Uncertain

8. How would you rate the **intra**-observer variability for the o/e LHR value among experienced sonographers (i.e. reproducibility of o/e LHR measurement for the same sonographer)?
   a. High intra-observer variability
b. Some intra-observer variability, but unlikely to be significant
c. Insignificant intra-observer variability
d. Uncertain

9. Please rate your confidence in obtaining the o/e LHR measurement.
a. Not confident
b. Somewhat confident
c. Fairly confident
d. Very confident (i.e.: expert level)

10. In your opinion, what are the most important imaging features that should be present for a LHR study to be considered to be of the highest quality? Please list these features in their order of priority:

   i) 
   ii) 
   iii) 
   iv) 
   v) 

11. Which ultrasound machine(s) do you routinely use in clinical practice?

12. Which calculator/nomogram do you use to obtain the o/e LHR value?

13. Which ultrasound reporting system do you use?

14. On average, how many cases of fetal diaphragmatic hernia do you see in your unit on an annual basis?

Thank you for taking the time to complete this survey.
Appendix D. Data collection sheet
### LHR Inter-rater agreement study

**Lung Area measurement using anterior-posterior (A-P) diameter method (mm x mm)**

<table>
<thead>
<tr>
<th>Case</th>
<th>Rate image quality for lung images (1=excellent, 2=good, 3=fair, 4=poor)</th>
<th>Please explain why you rated this image as you did*</th>
<th>Image (Frame) selected for lung area measurement</th>
<th>Lung Area measurement using anterior-posterior (mm x mm)</th>
<th>Long Area measurement using longest (L) diameter method (mm)</th>
<th>Long Area measurement using trace method (mm2)</th>
<th>Please explain why you rated this image as you did*</th>
<th>Image (Frame) selected for head measurements</th>
<th>Biparietal diameter (BPD) (mm)</th>
<th>Occipito-frontal diameter (OFD) (mm)</th>
<th>Head circumference (HC) (mm)</th>
<th>Please explain why you rated this image as you did*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td></td>
<td>No</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
<td>No</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
</tbody>
</table>

*Explain rating only for poor images (I.e., 3 & 4)

Please do not fill in column G, J and Q.
Appendix E. Usability testing script
Usability testing script for NAFTNet sonographic standardization of CDH prognostication project

Introduction:
- Nimrah Abbasi
  - Maternal-fetal medicine, Fetal Medicine Unit, Mount Sinai Hospital, Department of Ob/Gyn University of Toronto
  - MSc QIPS- Thesis project on standardization of sonographic prognostication of prenatally diagnosed CDH
- Aamir Siddiqi (Trice imaging)

Purpose of interview:
- Review objective of study
  - Establish baseline inter-rater agreement in measuring o/e LHR and validity (compared to an external reviewer)
  - Determine lung area measurement method with highest inter-rater agreement among participants in NAFTNet FETO consortium centers and NAFTNet centers outside of the consortium
  - Understand sources of variability in measurements/clinical practice
  - Importance of standardizing measurements among NAFTNet centers to ensure consistent prognostication and prenatal counseling, generalize neonatal outcomes, select candidates for prenatal intervention and particularly in the era of the Tracheal Occlusion To Accelerate Lung growth (TOTAL) randomized control trial (RCT) and future RCTs evaluating efficacy of prenatal treatment
  - Findings will be presented at NAFTNet biannual meeting
  - Standardized protocol for CDH prognostication to be developed based on findings and implemented across NAFTNet
- Review software for image downloading and obtaining measurements
- Review documents for submission (questionnaire, data collection sheet) + timeline
- Review methods of lung area measurement (anteroposterior, longest and trace method)
- Purpose of direct observation/usability testing
  - Identify key steps by “unbiased” observer
  - Subtle variations in sonographic assessment among fetal medicine experts
  - Expert “tips and tricks”
  - Selection criteria for optimal image
  - Evaluate ultrasound tool (TRICE imaging), ultrasound clips for measurement and potential use for future training
  - Anonymous
  - Non-evaluative
  - Brief (15-20 minutes)
Questions for reviewers:
1. Landmarks for optimal image selection (prioritize)
2. What makes a good image?
3. What makes a bad image?
4. Areas of uncertainty/lacking consensus
5. “Think aloud” for different lung area measurement methods and head circumference measurements
6. Preferred lung area measurement method (reviewer and center preferences)? Why?
7. Disadvantages/advantages of other methods (usability issues)?
8. Review questionnaire answers if time
Appendix F. Usability data collection sheet for FETO group
Data collection sheet for usability testing
Usability testing: FETO Group

Tasks:
Identify landmarks for lung area measurement

<table>
<thead>
<tr>
<th>Obtain lung area measurement using AP method</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landmarks:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obtain lung area measurement using longest method</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landmarks:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obtain lung area measurement using trace method</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landmarks:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Identify landmarks for head circumference
Preferred method: AP L Trace

Why:

Center preference: AP L Trace

Challenges with current measurement tools:

Additional data collection sheet

<table>
<thead>
<tr>
<th>Case 1 (excellent quality- all landmarks clear &amp; visible, lung up)</th>
<th>AP</th>
<th>Longest</th>
<th>Trace</th>
<th>Head</th>
<th>Comments on image quality</th>
</tr>
</thead>
</table>

| Case 2 (bad quality, lung down, shadowing and landmarks unclear) | | | | | |

| Case 3 (adequate quality, lung up, maternal obesity, some shadowing, lung borders less clear, oblique) | | | | | |
Appendix G. Usability data collection sheet for non-FETO group
Data collection sheet for usability testing
Usability testing: Non-FETO group

Tasks:
Identify landmarks for Lung area measurement
4 chamber view   Y   N
Axial section of the chest   Y   N
Clear lung borders   Y   N
Fetal position   Y   N
Image optimization   Y   N

Obtain lung area measurement using AP method   Y   N
AP diameter: Midclavicular, posterior wall of chest to atria, parallel to vertebrae & sternum   Y   N
Transverse diameter: Descending aorta to medial border of rib   Y   N

Obtain lung area measurement using longest method
Place callipers along longest diameters of lung   Y   N
Traverse structures other than lung   Y   N

Obtain lung area measurement using trace method
Trace entire perimeter of lung   Y   N
Include tissue other than lung   Y   N
Preferred method: AP L Trace

Why:

Center preference: AP L Trace

Challenges with current measurement tools:

Additional data collection sheet

<table>
<thead>
<tr>
<th>Case</th>
<th>AP</th>
<th>Longest</th>
<th>Trace</th>
<th>Comments on image quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1 (excellent quality-all landmarks clear &amp; visible, lung up)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 2 (bad quality, lung down, shadowing and landmarks unclear)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 3 (adequate quality, lung up, maternal obesity, some shadowing, lung borders less clear, oblique)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix H. Usability data collection sheet for sonographers and clinical fellows in training
### Data collection sheet for usability testing

**Usability testing: Sonographers, clinical fellows in training**

**Tasks:**

<table>
<thead>
<tr>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify landmarks for lung area measurement</td>
<td>Identify landmarks for lung area measurement</td>
</tr>
<tr>
<td>4 chamber view</td>
<td>Y</td>
</tr>
<tr>
<td>Axial section of the chest</td>
<td>Y</td>
</tr>
<tr>
<td>Clear lung borders</td>
<td>Y</td>
</tr>
<tr>
<td>Fetal position</td>
<td>Y</td>
</tr>
<tr>
<td>Image optimization</td>
<td>Y</td>
</tr>
</tbody>
</table>

- Obtain lung area measurement using trace method
- Callipers placed over contralateral lung  |  Y  |  N |
- Include tissue other than lung  |  Y  |  N |

### Additional data collection sheet

<table>
<thead>
<tr>
<th>Case 1 (excellent quality- all landmarks clear &amp; visible, lung up)</th>
<th>Image frame number</th>
<th>Trace</th>
<th>Comments on image quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

107
Appendix I. NAFTA Net o/e LHR protocol PDSA 4
Proposed standardized protocol for observed-to-expected lung-to-head ratio (o/e LHR) measurement in antenatal prognostication of congenital diaphragmatic hernia (CDH) by the NAFTNet FETO consortium (adapted from Jani et al. (1)) (PDSA 4)

1. An axial view of the chest at the level of the four-chamber view of the heart with clear visualization of the atrioventricular valves, inter-ventricular and inter-atrial septum should be obtained.
2. The image should be optimized with adequate magnification of the fetal thorax, such that the chest occupies the whole screen. The image should be frozen prior to final magnification to ensure landmarks are clearly visible and that the entire chest is visualized. Adequate contrast between the lung and surrounding tissue should be demonstrated, which may be aided by use of a wide dynamic range and a higher frequency probe.
3. The foetus should be positioned in a supine or transverse position, with the vertebral body in the axial plane of the chest located between 3 to 9 o’clock and the lung contralateral to the side of diaphragmatic hernia close to the transducer.
4. Shadows from ribs over the lung area to be measured should be avoided by placing the transducer such that the ultrasound beam is parallel to the ribs and crosses the intercostal space.
5. Lung area should be measured using the trace method, as this method has been demonstrated to have the highest reproducibility. Callipers should be placed over the edges of the lung. Attention should be taken to avoid inclusion of cardiovascular structures and myocardium, vessels, particularly the descending aorta and bony structures including the medial portions of the ribs and vertebrae.
6. To obtain the lung-to-head ratio (LHR), the area of the lung (mm²) is divided by the head circumference (mm).
7. To obtain the o/e LHR, the LHR is divided by the expected mean LHR for gestational age using the calculator from www.totaltrial.eu.
8. The qualitatively best image that meets the above criteria should be used to calculate o/e LHR.

Appendix J. NAFTNet o/e LHR protocol PDSA 6 & 7
**Sonographic measurement of observed-to-expected lung-to-head ratio (o/e LHR) for antenatal prognostication of congenital diaphragmatic hernia (CDH) (PDSA 6 & 7)**

(Adapted from Jani et al. (1))

1. An axial view of the chest with visualization of a single rib on either side of the chest at the level of the four-chamber view of the heart with clear visualization of the atrioventricular valves, interventricular and inter-atrial septum should be obtained.

2. The image should be optimized with adequate magnification of the fetal thorax, such that the chest occupies the whole screen. The image should be frozen prior to final magnification to ensure landmarks are clearly visible and that the entire chest is visualized. Adequate contrast between the lung and surrounding tissue should be demonstrated, which may be aided by use of a wide dynamic range and a higher frequency probe.

3. The fetus should be positioned in a lateral decubitus position with the vertebral body in the axial plane of the chest located between 3 to 9 o’clock, and the lung contralateral to the side of CDH close to the transducer.

4. Shadows from ribs over the lung area to be measured should be avoided by placing the transducer such that the ultrasound beam is parallel to the ribs and crosses the intercostal space.

5. Lung area should be measured using the trace method, as this method has been demonstrated to have the highest reproducibility(2). Callipers should be placed over the perimeters of the lung only. Attention should be taken to avoid inclusion of cardiovascular structures and myocardium, vessels, particularly the descending aorta and bony structures including the medial portions of the ribs and vertebrae.

6. The head circumference should be measured in the standard biparietal axial plane of the fetal head at the level of the cavum septi pelluciidi, thalami and posterior horns of the lateral ventricles with visualization of the midline falx dividing the brain into two symmetrical hemispheres. The cerebellum and posterior fossa should not be visualized in this plane. Head circumference can be measured by using the elliptical measurement tool placed around the outside of the skull bones.

7. The o/e LHR is preferred over the LHR as it independent of gestational age and has been shown to be a significant predictor of mortality in both left and right sided CDH, irrespective of liver herniation [1, 3]. To obtain o/e LHR, the LHR obtained by dividing the area of the lung (mm²) by the head circumference (mm) is divided by the expected mean LHR for gestational age using the calculator from www.totaltrial.eu.

8. In practice, 2-3 measurements should be obtained. The qualitatively best image that meets the above criteria should be used to calculate o/e LHR. If there are several such images, then average of measurements should be used.

9. An axial sweep of the chest should be obtained from which a still image meeting the above criteria should be selected for lung area assessment. Additionally, a sweep may help appreciate intrathoracic herniation of liver and other abdominal viscera. Identification of hepatic vasculature with color Doppler may further help quantify degree of liver herniation.
Appendix K. NAFTNet o/e LHR protocol PDSA 8
Sonographic measurement of observed-to-expected lung-to-head ratio (o/e LHR) for antenatal prognostication of congenital diaphragmatic hernia (CDH) (PDSA 8)

(Adapted from Jani et al. (1))

The o/e LHR is a significant predictor of neonatal morbidity and mortality, independent of gestational age and in both left- and right-sided CDH, irrespective of liver herniation (1,3). To obtain o/e LHR, the LHR obtained by dividing the area of the lung (mm²) by the head circumference (mm) is divided by the expected mean LHR for gestational age (GA) using the calculator from www.totaltrial.eu.

Severity may be classified depending on o/e LHR and presence of intrathoracic liver herniation (4):

<table>
<thead>
<tr>
<th>Pulmonary hypoplasia</th>
<th>o/e LHR</th>
<th>Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild</td>
<td>&gt;45% + intrabdominal liver</td>
<td>&gt;75%</td>
</tr>
<tr>
<td>Moderate</td>
<td>26-35% + intrathoracic liver</td>
<td>30-60%</td>
</tr>
<tr>
<td>Severe</td>
<td>15-25% + intrathoracic liver</td>
<td>30%</td>
</tr>
<tr>
<td>Extreme</td>
<td>&lt;15% + intrathoracic liver</td>
<td>&lt;0%</td>
</tr>
</tbody>
</table>

The following steps should be followed to obtain an o/e LHR:

1. An axial view of the chest should be obtained at the level of the four-chamber view of the heart with visualization of a single rib on each side of the chest. Atrioventricular valves, inter-ventricular and inter-atrial septum should be visualized as best possible.

2. The image should be optimized with adequate magnification of the fetal thorax, such that the entire chest is visualized and occupies the entire screen. The image should be frozen prior to final magnification to ensure landmarks are clearly visible. Use of a high frequency probe and wide dynamic range may help improve contrast between lung and surrounding tissue.

3. The fetus should be in a lateral decubitus position with the vertebral body at 3 or 9 o’clock, with the contralateral lung close to the transducer. Avoid rib shadowing by scanning through an intercostal space.

4. The contralateral lung area should be measured using the trace method, as this method has the highest inter-rater agreement(2). Calipers should be placed over the perimeters of the lung only. Cardiovascular structures, mediastinal vessels and bony structures including the medial portions of the ribs and vertebrae should not be included.

5. Head circumference should be measured at the level of the cavum septi pellucidi, thalami and posterior horns of the lateral ventricles. The midline falx should be seen dividing the brain into two symmetrical hemispheres. The cerebellum and posterior fossa should not be seen. The elliptical measurement tool may be used with calipers placed around the skull bones.

6. 2-3 measurements should be obtained. The qualitatively best image that meets criteria should be used. If there is more than 1 optimal image, then the average o/e LHR should be used.

7. An axial sweep of the chest should be obtained from which a still image meeting the above criteria should be selected for lung area assessment. Additionally, a sweep may help appreciate intrathoracic herniation of liver and other abdominal viscera. Identification of hepatic vasculature with color Doppler may further help quantify degree of liver herniation.

8. Training should occur under the supervision of a sonographer with sufficient expertise in measuring o/e LHR. A minimum of 5 images meeting the above criteria should be submitted for review by a local expert to ensure accurate measurement technique.

Comment [16]: Protocol adapted based on results from usability testing on NAFTNet o/e LHR protocol performed with representative end users.

Comment [17]: Wording unclear and text heavy. Heuristic violation(s): Minimalist design and use of the user’s language.

Comment [18]: Incorrect interpretation and wording. Image should be optimized with appropriate focus, depth and presence of necessary landmarks prior to magnification. Heuristic violation(s): Prevention of error.

Comment [19]: Usability testing reveals users still incorporating structures other than lungs on assessments after provision of protocol.

Comment [20]: Competency assessment included. Recommendation based on an international consensus statement for obstetric and gynecological ultrasound competency assessment in residency training.
Appendix L. NAFTNet protocol for sonographic measurement of o/e LHR for antenatal prognostication of CDH
Sonographic measurement of observed-to-expected lung-to-head ratio (o/e LHR) for antenatal prognostication of congenital diaphragmatic hernia (CDH)
(Adapted from Jani et al. (1))

o/e LHR is our best predictor of neonatal morbidity and mortality, independent of gestational age and can be used with left and right CDH (1, 3). To obtain the o/e LHR, lung area (mm$^2$) is divided by the head circumference (HC) [mm] and is expressed as a percentage of the expected mean LHR for gestational age (GA) using the calculator from www.totaltrial.eu.

CDH Severity and prognosis is determined by o/e LHR and intrathoracic liver herniation (4):

<table>
<thead>
<tr>
<th>Pulmonary hypoplasia</th>
<th>o/e LHR</th>
<th>Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild</td>
<td>&gt;45% + intrabdominal liver</td>
<td>&gt;75%</td>
</tr>
<tr>
<td>Moderate</td>
<td>26-33% + intrathoracic liver</td>
<td>30-60%</td>
</tr>
<tr>
<td>Severe</td>
<td>15-25%</td>
<td>30%</td>
</tr>
<tr>
<td>Extreme</td>
<td>&lt;15%</td>
<td>~0%</td>
</tr>
</tbody>
</table>

The following steps should be followed for an optimal o/e LHR:

1. Obtain an axial view of the chest at the level of the 4-chamber view of the heart with visualization of a single rib on either side of the chest. Atrioventricular valves, inter-ventricular and inter-atrial septi should be visualized as best possible.

2. The fetus should be in a left lateral decubitus position with its vertebral body at 3 or 6 o’clock and the contralateral lung (to the CDH) closest to the transducer.

3. Avoid rib shadowing by scanning through an intercostal space.

4. Optimize and magnify the image of the fetal thorax, such that the entire chest is visualized and occupies the full screen. Optimize the image prior to final magnification to ensure landmarks are clearly visible. A high frequency probe (e.g. 5-9 MHz) and wide dynamic range may help to improve contrast between the lung and surrounding tissues.

5. Measure the contralateral lung area using the trace method, which has the highest inter-rater agreement (2).

6. Place the calipers over the lung perimeter only. Do not include cardiovascular structures, mediastinal vessels or bony structures, including the medial portions of the ribs or vertebral.

7. Take an ultrasound sweep of the fetal chest, from which an optimal still image should be selected for lung area assessment. A sweep may also help identify intrathoracic herniation of liver and other abdominal viscera. Identification of hepatic vasculature with color Doppler may further help quantify liver herniation.

8. 2-3 measurements should be obtained. Use the qualitatively best image(s) that meet above criteria. If there is more than 1 optimal image, then use the average o/e LHR.

9. Measure HC at the level of the cavum septi pellucidi, thalami and posterior horns of the lateral ventricles. The midline falx should be seen dividing the brain into two symmetrical hemispheres. The cerebellum and posterior fossa should not be seen. An elliptical measurement tool may be used with calipers placed around outer edges of the skull bones.

10. Training should occur under the supervision of a sonographer with sufficient expertise in measuring o/e LHR. A minimum of 5 images meeting the above criteria should be submitted for review by a local expert to ensure accurate measurement technique (5).