

Viewpoint

# Transporting an Artificial Intelligence Model to Predict Emergency Cesarean Delivery: Overcoming Challenges Posed by Interfacility Variation

Joshua Guedalia<sup>1\*</sup>, MBA, PhD; Michal Lipschuetz<sup>1,2\*</sup>, RN, PhD; Sarah M Cohen<sup>2</sup>, MPH; Yishai Sompolinsky<sup>2</sup>, MD; Asnat Walfisch<sup>2</sup>, MD; Eyal Sheiner<sup>3</sup>, MD; Ruslan Sergienko<sup>4</sup>, MA; Joshua Rosenbloom<sup>2</sup>, MD; Ron Unger<sup>1</sup>, PhD; Simcha Yagel<sup>2</sup>, MD; Hila Hochler<sup>2</sup>, MD

<sup>1</sup>The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel

<sup>2</sup>Division of Obstetrics & Gynecology, Hadassah Medical Organization and Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem, Israel

<sup>3</sup>Department of Obstetrics and Gynecology, Soroka University Medical Center, Ben-Gurion University of the Negev, Beer-Sheva, Israel

<sup>4</sup>Department of Public Health, School of Public Health, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

\*these authors contributed equally

**Corresponding Author:**

Michal Lipschuetz, RN, PhD

Division of Obstetrics & Gynecology

Hadassah Medical Organization and Faculty of Medicine

Hebrew University of Jerusalem

Churchill Avn. 8

Jerusalem, 9765415

Israel

Phone: 972 25841111

Email: [michal.lipschuetz@gmail.com](mailto:michal.lipschuetz@gmail.com)

## Abstract

Research using artificial intelligence (AI) in medicine is expected to significantly influence the practice of medicine and the delivery of health care in the near future. However, for successful deployment, the results must be transported across health care facilities. We present a cross-facilities application of an AI model that predicts the need for an emergency caesarean during birth. The transported model showed benefit; however, there can be challenges associated with interfacility variation in reporting practices.

(*J Med Internet Res* 2021;23(12):e28120) doi: [10.2196/28120](https://doi.org/10.2196/28120)

**KEYWORDS**

machine learning; algorithm transport; health outcomes; health care facilities; artificial intelligence; AI; ML; pregnancy; birth; pediatrics; neonatal; prenatal

## Introduction

The integration of artificial intelligence (AI) into health care is expected to significantly influence the practice of medicine [1-4]. Machine learning (ML) as a modeling strategy is an attractive option for characterizing and predicting complex biological phenomena [5].

Critics of AI applications note that the applications are primarily based on retrospective research, with insufficient focus devoted to “real-life” implementation and verification of reproducibility in clinical practice [5,6]. For example, an ML prediction algorithm developed in an urban tertiary care center with a

diverse patient population may be unsuitable for a community hospital treating a homogenous population according to local protocols.

Therefore, transporting AI models across health care facilities is critical to effectively translating AI research into medical practice [7]. In this study, we aimed to investigate the validation of a model to predict the need for an emergency caesarean during birth, the critical challenge stemming from interfacility variation in subjective measurements, and to devise a method to address this challenge.

## Methods

In brief, we developed 2 ML models to predict the risk for emergency caesarean delivery (for a detailed description of the methods and model features, see [Multimedia Appendix 1](#) and [\[8\]](#)). The first model was designed to be used at admission to the labor and delivery unit (admission model); the second model was designed for use during labor, integrating additional data that accumulate as labor progresses (labor progression model). These additional data supplementing the model allow for more accurate prediction. Both models will alert the staff of the likelihood that a parturient might require an emergency caesarean delivery, allowing for the preparation of staff and patient.

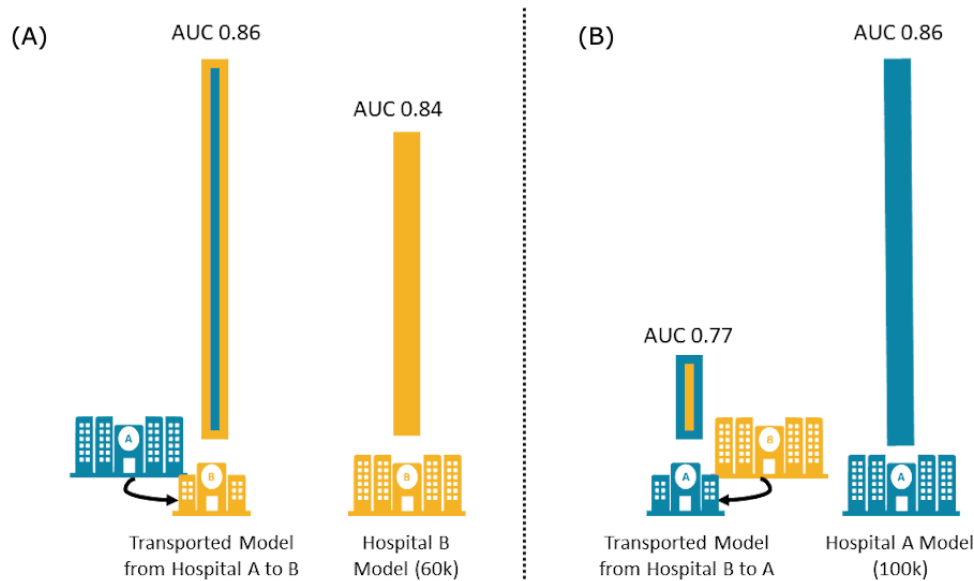
The models were trained using data from approximately 100,000 births at Hospital A. We extracted multiple data features from individual parturient electronic medical records (EMRs), totaling approximately 11 million data points. The institutional review boards at Hadassah Hebrew University Medical Center and Soroka Medical Center approved the study.

Both models were able to predict the need for emergency caesarean delivery, with the admission model achieving an area under the curve (AUC) of 0.82 and the labor progression model showing an increased performance, with an AUC of 0.86.

Having created and trained an ML-based model at a given health care facility, model transport can provide a smaller facility its benefits, without the large stored medical records or the expense and expertise required for development. However, care must be taken to monitor how the transport may affect the performance of the models, given differences in populations or settings.

We compared the prediction performance of the models trained and tested at Hospital A when transported to a second facility, Hospital B, where they were tested on data from approximately 60,000 births. Both the admission and labor progression models transported from Hospital A showed comparable prediction performance at Hospital B. [Figure 1A](#) illustrates the transport and performance of the labor progression model (see [Multimedia Appendix 2](#) for the hospital characteristics and [Multimedia Appendix 3](#) for the AUCs and 95% CIs of all models).

**Figure 1.** (A) Comparing the performance of Hospital A labor progression model (in blue) transported to Hospital B (yellow/blue bar) versus Hospital B local model (in yellow) and (B) Comparing the performance of Hospital B labor progression model (in yellow) transported to Hospital A (blue/yellow bar) versus Hospital A local model (in blue). AUC: area under the curve.



We then reversed the process and retested the success of transporting the models, by training the models at Hospital B and testing the prediction accuracy at Hospital A. Although the admission model trained at Hospital B provided similar levels of prediction at Hospital A, the labor progression model showed a reduced level of prediction (AUC 0.77 vs AUC 0.84; [Figure 1B](#)). We examined the model features to determine the cause of this decreased performance (see [Multimedia Appendix 1](#)).

Two important measurements of labor progression are fetal head station and cervical dilation. Fetal head station denotes the fetal descent within the maternal pelvis based on the position of the fetal head in centimeters above (-) or below (+) the maternal ischial spines [\[9\]](#). Cervical dilation refers to the opening of the maternal uterine cervix, in centimeters, from closed cervix (0 cm) to full cervical dilation (10 cm). These 2

measurements represent the primary features of the progress of the birth; how rapidly descent and dilation progress depends on several factors, including parturient parity, medical history, pelvic anatomy, the size of the fetus, and the position of the fetus at the time of labor [\[10\]](#). Results are operator-dependent, and measurements can vary between facilities based on local protocols and practice habits [\[11\]](#).

We identified a difference between the 2 facilities in fetal head station measurements used by the labor progression model. Specifically, we found that the dispersion and central tendency of this variable, as stratified to cervical dilation, differed between the 2 hospitals: Data from Hospital A were widely distributed across the full -3 to +3 scale, while those from Hospital B were more concentrated around -2 to +2. This difference may explain the reduced performance when transporting from Hospital B,

while no reduction in performance was observed when transporting from Hospital A.

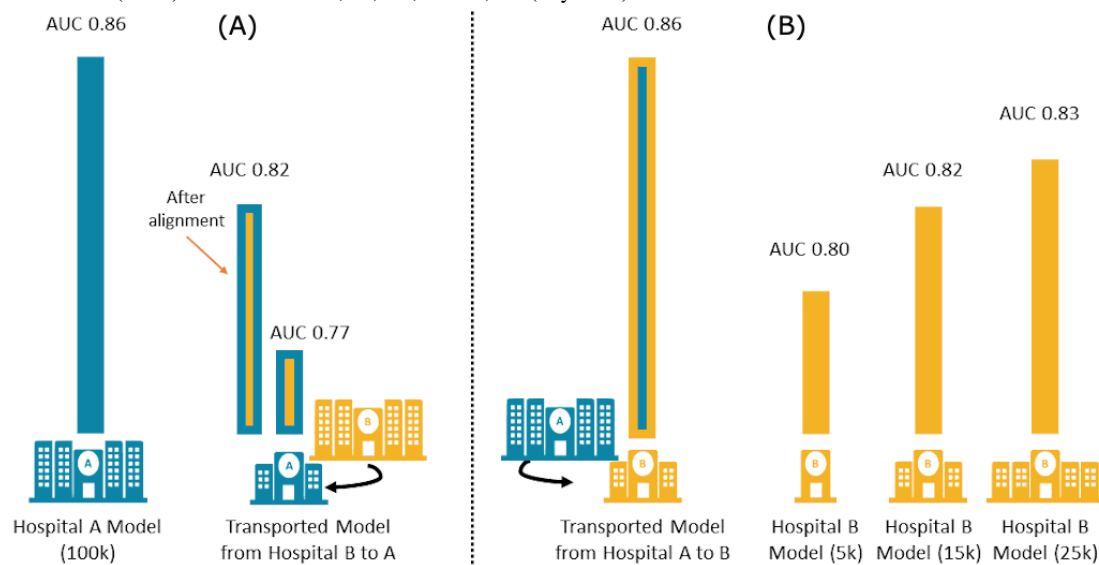
In order to overcome this disparity, we evaluated the patterns of distribution of fetal head station as distributed across the dilation. We aligned the station within the distribution of the cervical dilation in order to encompass both approaches. This partly adjusted for the variation and improved the cross-facility prediction (AUC 0.82; [Figure 2A](#); see [Multimedia Appendix 1](#) and [Multimedia Appendix 3](#) for the AUCs and 95% CIs of all models).

This difference highlights the difficulties introduced by discrepancies in reporting practices between facilities.

Alignment can resolve some disparities, but here, it only partly recouped model performance.

To further evaluate whether our labor progression model could potentially benefit an even smaller facility, we simulated a hospital with a smaller EMR. The 100,000-case Hospital A model transported to Hospital B showed better performance (AUC 0.86) than a Hospital B model based on small samples of 5000 (AUC 0.80), 15,000 (AUC 0.82), and 25,000 (AUC 0.83) cases, emphasizing the benefit that can accrue to a smaller facility from a model trained at a larger facility and that the additional benefit decreases as the size of the available local EMR grows ([Figure 2B](#)).

**Figure 2.** (A) Comparing the performance of Hospital B labor progression model (in yellow) transported to Hospital A versus Hospital B model after alignment adjustments transported to Hospital A (blue/yellow bars) versus Hospital A local model (in blue) and (B) Comparing the performance of Hospital A labor progression model transported to Hospital B (yellow/blue bar) versus Hospital B local models trained on progressively larger local electronic medical record (EMR) data sets of 5000, 15,000, and 25,000 (in yellow). AUC: area under the curve.



## Conclusions

In conclusion, integrating ML applications into clinical medicine will require validation and transportation between medical facilities [7,12-14]. We demonstrated that ML may be applied to clinical practice and to obstetrics in particular. A big data-driven ML algorithm can be successfully transported, and a data-poor center can benefit from work performed in a larger facility.

However, transportation requires careful investigation of specific features and consideration of variations in local populations, protocols, and reporting to calibrate the system fit [7,12]. Nevertheless, model predictions are heavily dependent on the data used in training and by the variations in recording practices and protocols operative in a given health care facility. We observed that the more detailed labor progression model, when

trained without accounting for reporting differences, provided a lower AUC than the admission model. Although the progression model contained more detailed information on the progression of the labor and intrahospital showed benefit over the admission model, the benefit provided was lost when transporting the model to a different hospital: The transported model performance was inferior to that of the simpler model. Interfacility variation between health care centers may introduce unexpected effects into a prediction model. Generalizability and transportability among medical facilities necessitate overcoming biases via external validation and adapting the model to local protocols [15].

Successful translation of AI research into practice depends on transport across health care facilities. This can individualize health care, improve outcomes, and reduce complications across broader populations.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Additional methodology.

[\[DOCX File , 544 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Demographic parameters of the two hospitals.

[\[DOCX File , 14 KB-Multimedia Appendix 2\]](#)

## Multimedia Appendix 3

AUROC of the different models.

[\[DOCX File , 14 KB-Multimedia Appendix 3\]](#)

## References

1. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med* 2020 Sep 24;3(1):126 [FREE Full text] [doi: [10.1038/s41746-020-00333-z](https://doi.org/10.1038/s41746-020-00333-z)] [Medline: [34556791](https://pubmed.ncbi.nlm.nih.gov/34556791/)]
2. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019 Apr 04;380(14):1347-1358. [doi: [10.1056/NEJMr1814259](https://doi.org/10.1056/NEJMr1814259)] [Medline: [30943338](https://pubmed.ncbi.nlm.nih.gov/30943338/)]
3. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015 Jul 17;349(6245):255-260. [doi: [10.1126/science.aaa8415](https://doi.org/10.1126/science.aaa8415)] [Medline: [26185243](https://pubmed.ncbi.nlm.nih.gov/26185243/)]
4. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 2018 Nov 27;19(6):1236-1246 [FREE Full text] [doi: [10.1093/bib/bbx044](https://doi.org/10.1093/bib/bbx044)] [Medline: [28481991](https://pubmed.ncbi.nlm.nih.gov/28481991/)]
5. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019 Oct 29;17(1):195 [FREE Full text] [doi: [10.1186/s12916-019-1426-2](https://doi.org/10.1186/s12916-019-1426-2)] [Medline: [31665002](https://pubmed.ncbi.nlm.nih.gov/31665002/)]
6. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020 Mar 25;368:m689 [FREE Full text] [doi: [10.1136/bmj.m689](https://doi.org/10.1136/bmj.m689)] [Medline: [32213531](https://pubmed.ncbi.nlm.nih.gov/32213531/)]
7. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015 Mar;68(3):279-289 [FREE Full text] [doi: [10.1016/j.jclinepi.2014.06.018](https://doi.org/10.1016/j.jclinepi.2014.06.018)] [Medline: [25179855](https://pubmed.ncbi.nlm.nih.gov/25179855/)]
8. Guedalia J, Lipschuetz M, Novoselsky-Persky M, Cohen SM, Rottenstreich A, Levin G, et al. Real-time data analysis using a machine learning model significantly improves prediction of successful vaginal deliveries. *Am J Obstet Gynecol* 2020 Sep;223(3):437.e1-437.e15. [doi: [10.1016/j.ajog.2020.05.025](https://doi.org/10.1016/j.ajog.2020.05.025)] [Medline: [32434000](https://pubmed.ncbi.nlm.nih.gov/32434000/)]
9. Dupuis O, Silveira R, Zentner A, Dittmar A, Gaucherand P, Cucherat M, et al. Birth simulator: reliability of transvaginal assessment of fetal head station as defined by the American College of Obstetricians and Gynecologists classification. *Am J Obstet Gynecol* 2005 Mar;192(3):868-874. [doi: [10.1016/j.ajog.2004.09.028](https://doi.org/10.1016/j.ajog.2004.09.028)] [Medline: [15746684](https://pubmed.ncbi.nlm.nih.gov/15746684/)]
10. Evbuomwan O, Chowdhury YS. *Physiology, Cervical Dilation*. Treasure Island, FL: StatPearls Publishing; 2021.
11. Graseck A, Tuuli M, Roehl K, Odibo A, Macones G, Cahill A. Fetal descent in labor. *Obstet Gynecol* 2014 Mar;123(3):521-526. [doi: [10.1097/AOG.0000000000000131](https://doi.org/10.1097/AOG.0000000000000131)] [Medline: [24499749](https://pubmed.ncbi.nlm.nih.gov/24499749/)]
12. Song X, Yu ASL, Kellum JA, Waitman LR, Matheny ME, Simpson SQ, et al. Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. *Nat Commun* 2020 Nov 09;11(1):5668 [FREE Full text] [doi: [10.1038/s41467-020-19551-w](https://doi.org/10.1038/s41467-020-19551-w)] [Medline: [33168827](https://pubmed.ncbi.nlm.nih.gov/33168827/)]
13. Hassanzadeh H, Nguyen A, Karimi S, Chu K. Transferability of artificial neural networks for clinical document classification across hospitals: A case study on abnormality detection from radiology reports. *J Biomed Inform* 2018 Sep;85:68-79 [FREE Full text] [doi: [10.1016/j.jbi.2018.07.017](https://doi.org/10.1016/j.jbi.2018.07.017)] [Medline: [30026067](https://pubmed.ncbi.nlm.nih.gov/30026067/)]
14. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng* 2010 Oct;22(10):1345-1359. [doi: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191)]
15. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014 Aug 01;35(29):1925-1931 [FREE Full text] [doi: [10.1093/eurheartj/ehu207](https://doi.org/10.1093/eurheartj/ehu207)] [Medline: [24898551](https://pubmed.ncbi.nlm.nih.gov/24898551/)]

## Abbreviations

- AI:** artificial intelligence  
**AUC:** area under the curve  
**EMR:** electronic medical record  
**ML:** machine learning

*Edited by G Eysenbach; submitted 22.02.21; peer-reviewed by N Barda, T Kahlon, JA Benítez-Andrades; comments to author 02.07.21; revised version received 25.08.21; accepted 29.10.21; published 10.12.21*

*Please cite as:*

*Guedalia J, Lipschuetz M, Cohen SM, Sompolinsky Y, Walfisch A, Sheiner E, Sergienko R, Rosenbloom J, Unger R, Yagel S, Hochler H*

*Transporting an Artificial Intelligence Model to Predict Emergency Cesarean Delivery: Overcoming Challenges Posed by Interfacility Variation*

*J Med Internet Res 2021;23(12):e28120*

*URL: <https://www.jmir.org/2021/12/e28120>*

*doi: [10.2196/28120](https://doi.org/10.2196/28120)*

*PMID:*

©Joshua Guedalia, Michal Lipschuetz, Sarah M Cohen, Yishai Sompolinsky, Asnat Walfisch, Eyal Sheiner, Ruslan Sergienko, Joshua Rosenbloom, Ron Unger, Simcha Yagel, Hila Hochler. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 10.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.