ML and precision public health: Saving mothers and babies from dying in rural India

Morris K[1^], Huang VS[1^], Jain M[1], Ramesh BM[2], Kemp H[1], Blanchard J[2], Isac S[3], Sarkar B[3], Gothalwal V[3], Namasivayam V[3], Sgaier S[1,4,5*]

1. Surgo Foundation, Washington, DC, USA
2. Dept Community Health Sciences, University of Manitoba, Winnipeg, MB, Canada
3. Centre for Global Public Health, Dept Community Health Sciences, University of Manitoba, Winnipeg, MB, Canada & India Health Action Trust, Lucknow, Uttar Pradesh, India
4. Dept of Global Health & Population, Harvard T.H. Chan School of Public Health, Boston, MA, USA
5. Dept Global Health, University of Washington, Seattle, WA, USA

^: joint first authors
*: corresponding author (semasgaier@surgofoundation.org)

Introduction

1. Ambitious global health goals with limited resources call for more targeted intervention.
2. In India, high maternal and neonatal mortality prompted system-wide intervention (e.g., incentives) to encourage baby deliveries at hospital rather than home.
3. Initial success (hospital delivery up by 40% over 10 years) has plateaued; 20% of women still deliver at home. Current interventions rely mostly on data of "what" people do or have, rather than a true understanding of "why" people do them.
4. Application of multiple analysis techniques could potentially yield more holistic insights.

Method

Survey Data:
- 5,968 mothers in Uttar Pradesh, India, who had given birth within the last 60 days (2018)
- Sampling from community healthcare worker catchment areas, from 75 Districts and 600 Blocks

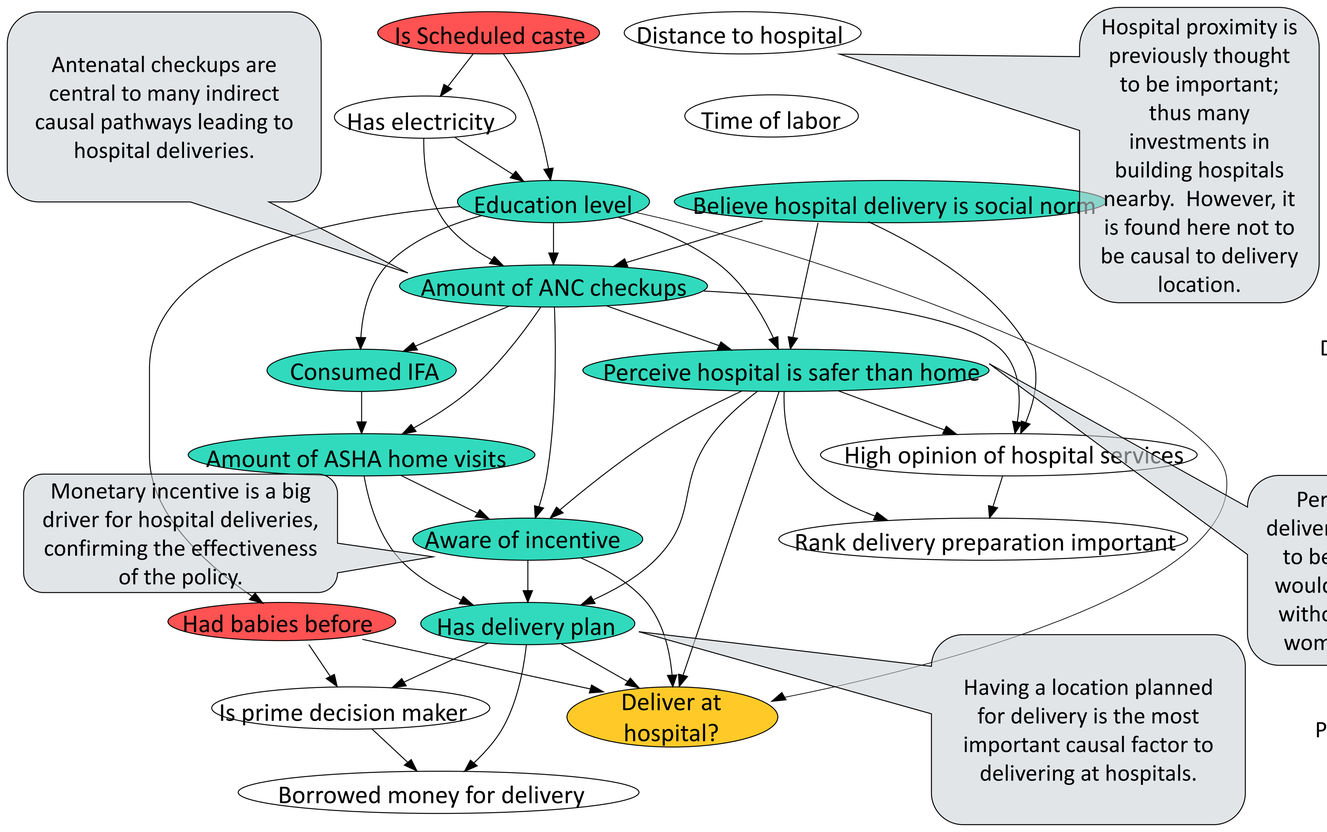
Three-pronged approach:
1. Collect data on perceptual and social drivers, not the typical approach of just demographics and behaviors.
2. Use Causal Bayesian network (BN) to learn the interactions between these factors, prescreened using OLS (GNS Healthcare's REFS™ software, which uses Markov Chain Monte Carlo for structural discovery).
3. Use Chi-square automatic interaction detection analysis (CHAID) for segmentation to inform policies.

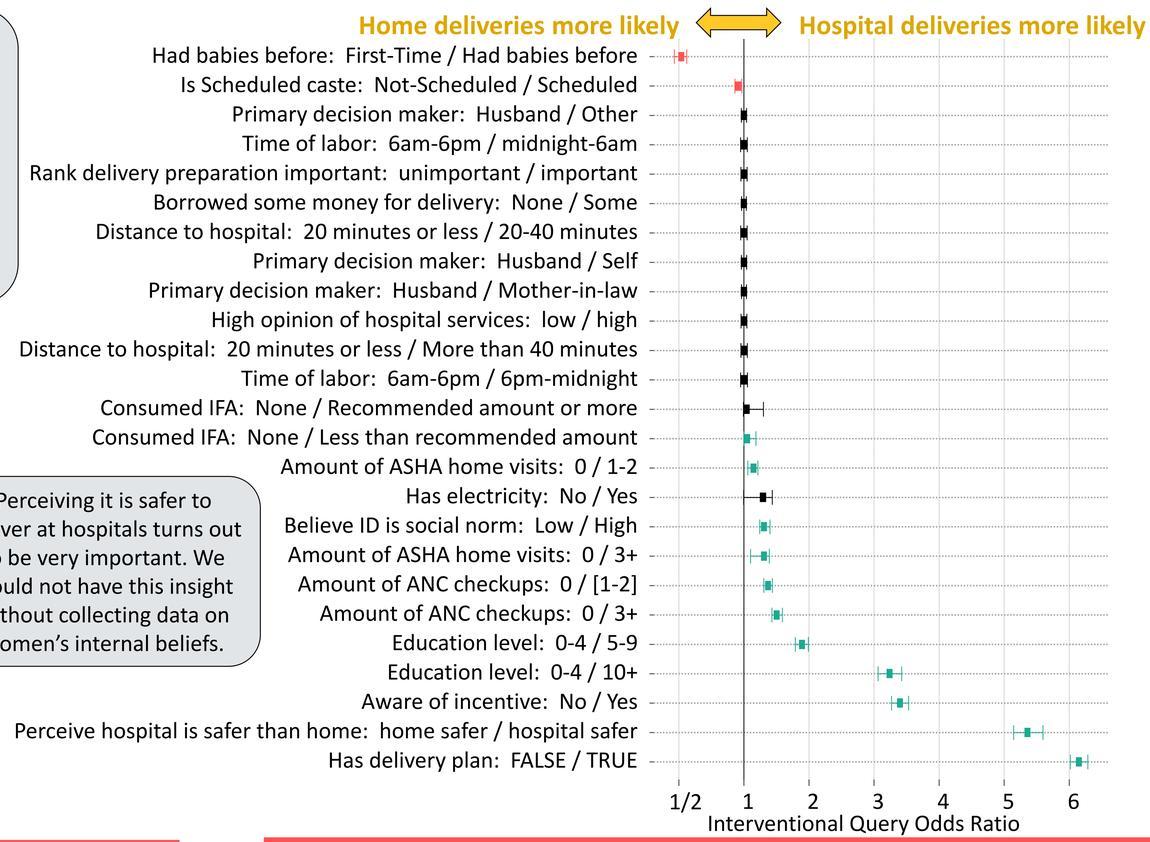
Result 1: Collected broad set of variables to include perception and social drivers

1. We collected perceptual, structural, and influencers in addition to demographics and behaviors to paint the more whole picture.
2. Significant predictors of hospital delivery from regression and caste were used as inputs for BN learning.

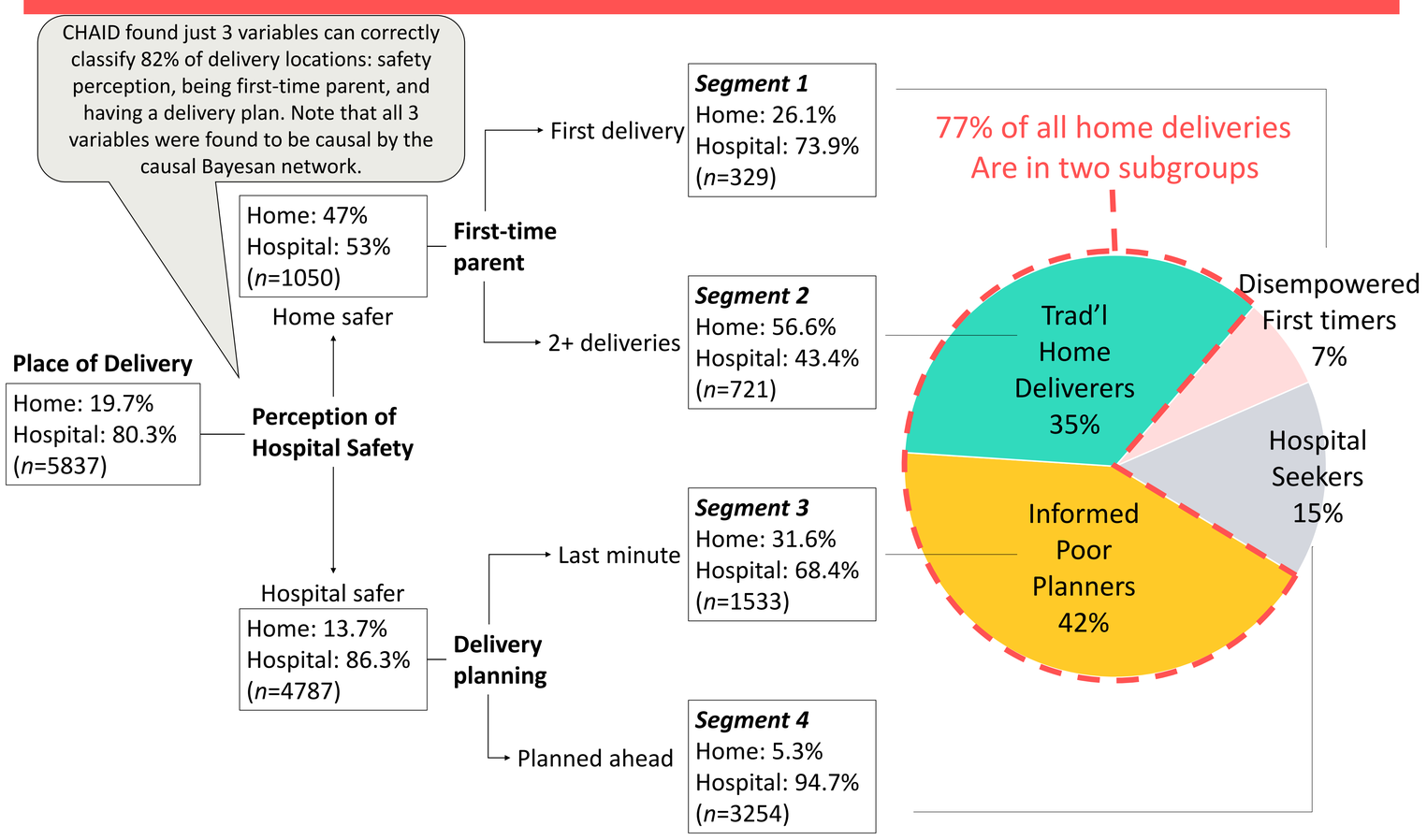
Variable	Response Options	Regression	Bayesian Network
Internal Beliefs			
Opinion of Hospital Facilities	7-item composite; Low vs. High	X	
Opinion of Hospital Services	6-item composite; Low vs. High	X sig	X
Rank Importance of Hospital Delivery	Important vs. Unimportant	X sig	X
Risk Perception of Childbirth	Low vs. High	X	
Worry about Delivery Problems	Little vs. Lot	X	
Perception of Hospital Safety	Hospital safer vs. Home safer	X sig	X
Nurse Gives Injection to Make Delivery Easier	Agree vs. Disagree	X	
Hospital is Not Necessary if Birth Attendant is Good	Agree vs. Disagree	X	
Hospital is Not Necessary if Past Home Delivery	Agree vs. Disagree	X	
Pregnant Women Attract Evil Spirits	Agree vs. disagree	X	
False Beliefs about ANC checkups	3-item composite; Few vs. Many	X	
Barriers to ANC checkups	5-item composite; Few vs. Many	X	
Knowledge of IFA	Percent correct recall (0-100)	X	
Agency	10-item composite; 1-5 Likert scale	X	
Insecurity	2-item composite; 1-5 Likert scale	X	
Conscientiousness	3-item composite; 1-5 Likert scale	X	
Empathy	1-5 Likert scale	X	
Openness	2-item composite; 1-5 Likert scale	X	
Optimism	1-5 Likert scale	X	
Neuroticism	1-5 Likert scale	X	
Demographics			
Education	0-4 years, 5-9 years, 10-12 years, 13+ years	X sig	X
Parity (# of pregnancies where baby is viable)	1, 2, 3, 4+	X sig	X
Religion	Hindu vs. Other	X	
Caste	ST, SC, OBC, none of these	X	X
Income	Little vs. Lot	X	
Financial Insecurity	2-item composite; 1-5 Likert scale	X	
Electricity in home	Yes vs. No	X sig	X
Household type	Nuclear vs. Joint/Other	X	
Structural			
Social Norms	Low vs. High	X sig	X
Hospital Distance	0-20 min vs. 21-40 min vs. 40+ min	X sig	X
Labor Start Time	Middle of the night vs. Day vs. Evening	X sig	X
Money Borrowed	None vs. Some	X sig	X
Incentive Awareness	Yes vs. No	X sig	X
Influencers			
Discussed Delivery Location with ASHA	True vs. False	X	
Primary Decision Maker	Self vs. Husband vs. Mother-in-law vs. Other	X sig	X
People for Social Support	Few vs. Many	X	
Number of ASHA Home Visits	None vs. 1-2 vs. 3-4 vs. 5+	X sig	X
Behavior			
Pregnancy Registration	Not registered vs. 1st vs. 2nd vs 3rd trimester	X	
Delivery Plan	Planned ahead vs. Last minute decision	X sig	X
Number of ANC Checkups	0 - 9	X sig	X
Take IFA during pregnancy	None vs. <Recommended vs. >Recommended	X sig	X

Result 2: Bayesian network finds direct causes to prioritize intervention while refuting factors previously thought to be important





Result 3: Segmentation provides basis for who to target for interventions



Discussion

1. A lack of high-quality data could result in precision targeting recommendations that are incomplete, inaccurate or biased. Precision public health approach requires better data collection to encompass a broad set of potential drivers and barriers to behavior. For example safety perception is a strong determinant for delivery location. Our example followed a toolkit called CUBES (to Change behavior, Understand Barriers, Enablers, and Stages of change) that we had previously developed
2. Multiple methodologies (both traditional and ML) together provide a more holistic view of the outcome of interest: methods could range from well-practiced (i.e., regression and CHAID decision tree) to emerging approaches not typically seen in global health (i.e., causal Bayesian network)
3. Causal Bayesian network allows us to determine direct and indirect causal factors and their relative level of impact should they be intervened on. Factors previously thought to be important (e.g., distance to hospital) turn out to be not causal to delivery location.
4. CHAID or other tree methods could provide the basis to make policy implementation more practical. Decision trees can prioritize segmenting variables. We found that 77% of all home deliveries occur with women in two subgroups determined by just three variables: perception of hospital safety, being a first-time parent, and whether there was a plan for delivery location.
5. Careful collaboration between researchers, policy makers and implementers is required throughout the analytical process to mitigate data quality issues, erroneous interpretation, and to improve the quality of the modelling output.
6. We have applied this approach to two other use cases in low income settings, demonstrating the generalizability of adopting a precision public health framework: (A) tuberculosis care seeking in Chennai, India (B) Sexual and reproductive health in Madhya Pradesh, India