

Using an Innovative Database and Machine Learning to Predict and Reduce Infant Mortality

In some places, and among some groups in the United States, infant mortality remains exceptionally high. For years, health care and community-based social services have delivered interventions designed to improve outcomes, but the interventions do not work equally well for all women and their children. Could an individualized approach of predicting risk and offering personalized intervention recommendations reduce the rate of infant mortality?

Disparities in Infant Mortality Rates at the Community Level

Infant mortality rates—indicators of population health—typically vary by level of economic development. Evidence shows that a 1-percent increase in gross domestic product (GDP) per

KEY FINDINGS

- Researchers created a unique data set that links individual-level vital statistics, electronic health records, community-based social service records, and other socio-environmental data describing ten years of births in Allegheny County, Pennsylvania.
- The research team designed and implemented machine learning algorithms and causal inference models to predict which women and their children were at highest risk of infant mortality, the interventions that women were most likely to use, and which interventions would most effectively reduce the risks for each woman and child.
- The interventions found to be most effective—broad pre-conception care, frequent prenatal care, doula support, and home visiting—aim to improve the health of the mother and result in lower mortality risk for the infant, especially when initiated before or early in pregnancy.
- Providers of health care and community-based social services can use the models with their patients or clients at high risk of infant mortality to tailor intervention options to their needs. Health care and social services can be better coordinated through better provider awareness of services across sectors and utilization of new tools.
- The models, methods, and tools developed are flexible and can be used by other localities and for other health conditions.

capita correlates with a 20-percent reduction in infant mortality.¹ However, despite the economic wealth of the United States overall, the infant mortality rates of some U.S. communities rival those of nations with less developed economies.

Overall, 5.7 of every 1,000 infants born alive died before the age of 1 in the United States in 2018. In contrast, European countries with similar levels of GDP per capita—such countries as Germany and the Netherlands—have much lower rates of infant mortality, around 3.2 per 1,000 births. However, a deeper look at certain places and among some groups shows that some U.S. communities have substantially higher rates of infant mortality. In some neighborhoods of Allegheny County in southwest Pennsylvania, for instance, the rate of infant mortality is more than 20

per 1,000 births, on par with such developing countries as Nicaragua and the Philippines.

In many communities, infant mortality persistently affects some groups disproportionately. Specifically, in Allegheny County, the disturbingly high rates and disparities in infant mortality among some groups have remained for many decades despite robust health care and social service systems. For example, over the past two decades, non-Hispanic black (hereafter, black) infant mortality was, on average, 3.2 times higher than non-Hispanic white (hereafter, white) infant mortality, a disparity 44 percent higher than that compared with the United States overall (see Figure 1).

Little is known about which health services and social service programs effectively reduce risk, for whom, and under which circumstances. Potentially beneficial health services include pre- and inter-conception care, early prenatal care, and doula support. Social services include an array of supports, from home visiting programs to substance abuse programs. Disconnects between the health care and social service systems indicate not only a lack of coordination between health and social services, but also a dearth of integrated data to help identify individuals at greatest risk.

When, What, and Where of Interventions

Interventions can take place in the time before pregnancy (*preconception* for a first pregnancy, or *interconception* for the time between pregnancies), during the pregnancy itself (*prenatal*), or within a year of the birth (*postnatal*).

For the purposes of this study, interventions were grouped into two main categories: health services and social services. Within these groups, specific services and programs are grouped into **health services**:

- pre- or interconception care (e.g., exams, vaccinations, genetic counseling, and contraceptive counseling within one year of conception)
- prenatal care (e.g., first prenatal visit during first trimester, one or more prenatal visits per month during pregnancy)
- doula support at birth.

Others are grouped into **social services**:

- home visiting programs (e.g., Healthy Start, Nurse Family Partnership, Early Head Start)
- pregnancy or postpartum support (e.g., Family Support Centers)
- breastfeeding and nutrition support (e.g., the Women, Infant, and Children [WIC] nutritional program)
- behavioral health services (e.g., drug and alcohol services, mental health support services).

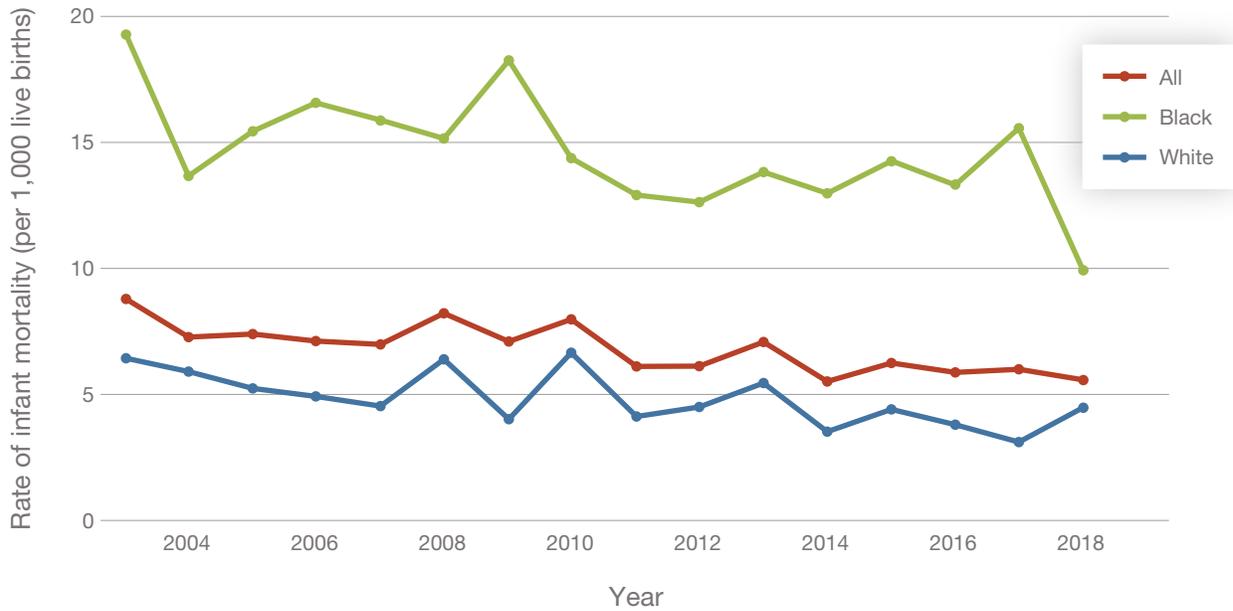
Timing of Most Infant Deaths

Nearly half of infant mortalities in Allegheny County occur within 24 hours of birth, suggesting that the most-important risk factors are prenatal and maternal rather than postnatal. Therefore, solutions are not likely to lie in safe sleep interventions or in programs that teach new parents certain behaviors to use in the first days and weeks after birth. Rather, the prenatal and maternal factors leading to these deaths include life course issues, such as poverty and systemic racism, and difficult-to-change behavioral issues, such as diet and smoking (both before and during pregnancy). The persistence of elevated infant mortality rates necessitates new approaches to mitigate the effects of these factors on birth outcomes.

Study Approach

The Richard King Mellon Foundation, a charitable institution located in Allegheny County, recognized the persistent problem of infant mortality in the Pittsburgh area and sought a new way to tackle the

FIGURE 1
 Infant Mortality Trends in Allegheny County



issue by enlisting the help of academic, medical, and community partners. The resulting project combined the efforts of researchers from the RAND Corporation, the University of Pittsburgh, and the Magee Women’s Research Institute (MWRI), in collaboration with a broad coalition of community-based stakeholders focused on maternal and child health.

Together, the research team set out to develop an integrated and personalized tool to predict risk and recommend interventions that can be used to coordinate care provided by health and social service providers before, during, and after pregnancy. The first in a multistep process, they built a database amassing ten years of data on births and infant deaths in Allegheny County, data on individual medical and other factors related to the outcomes, use of health and social services, and other socio-environmental factors. Then, using machine learning algorithms and causal inference modeling, they set out to develop tools that would be able to predict infant mortality risk and assess the effectiveness of each intervention. In combination, these tools would enable the identification of those at greatest risk and the personalized recommendation of the interventions that best reduce the risk.

The Database

The Infant Mortality Prediction System with the Intervention Management (or IMPreSIv) database combines data in a unique way that paints a holistic picture of the stressors on mothers’ and children’s health and the interventions available to reduce those stressors. The system does this by linking individual-level vital statistics data with electronic health records, community-based social service records, and other socio-environmental data. We are not aware of any other database containing this combination and scale of information at the individual level.

Birth and death certificate vital statistics records describe all the births and infant deaths in Allegheny County during the study period, from 2003 to 2013. Electronic health records from local health care systems describe maternal and child health risk factors from diagnostic histories, ultrasound measurements, test results, prescriptions, and procedures.

The participation in or use of health and social services is described in the vital statistics, electronic health records, and community-based social service records. Vital statistics describe when prenatal care was initiated, how frequently it was used, and whether doula support was used. Social service records describe participation in interventions delivered outside health care settings, such as the WIC nutritional program, sub-

stance abuse programs, and home visiting. Additional publicly available data describe the social, demographic, economic, and environmental factors (both supportive [such as access to quality nutrition] and detrimental [such as exposure to pollution]) that influence women's and children's health.

The Models

With data from IMPreSIV, the researchers constructed three sets of models that would

1. predict infant mortality risk
2. estimate the likelihood that each woman would participate in the intervention
3. determine the effects of each intervention on the risk of infant mortality.

The research team used machine learning algorithms to flexibly describe how a variety of factors might have influenced infant mortality risk and intervention participation. These algorithms allowed for insights not previously possible, and enhanced the models' predictive accuracy. Causal inference models incorporated the predicted risk and participation probabilities to determine the effects of each intervention. Together, the results provide meaningful and interpretable insights for health care and social service providers.

Results

Combining the database with the models gave the team an impressive ability to answer each of the following questions:

- Who is at risk?
- Which interventions are women most likely to use?
- Which interventions reduce risk the most?

Who Is at Risk?

The infant mortality prediction model identifies not only which women are at greatest risk for infant mortality, but also when they are at greatest risk. The team was also able to identify important risk predictors. Statistically significant predictors include prenatal observations, such as abnormal ultrasound measurements, postnatal observations (e.g., birth weight), parental characteristics (e.g., education and substance

abuse), and socio-environmental risk factors (e.g., local poverty and air quality).

In addition, the team used machine learning algorithms to predict infant mortality risk at different points in time (preconception, throughout pregnancy, and postnatal) and with varying levels of data availability. With predictions of their patient's risk, both health care and social service providers (who have access to different data) can better understand how to provide needed care and support to their patients and clients. And no matter when the patient interacts with a provider, the models can adjust to accurately describe their risk using the information available at the time.

Which Interventions Are Women Most Likely to Use?

Understanding which interventions are likely to be used by which women is important because risk will not be affected if participation does not increase. Furthermore, in the Allegheny County context, care coordination might be an important driver of participation. Studies from other contexts show that care coordination across systems improves maternal and child health outcomes,² and our study shows that, in Allegheny County, there appears to be little coordination and overlap between the use of health and social services. Figure 2 shows that, although the use of different social services is correlated (green indicates high correlation of use), the women who use social services (e.g., home visiting) do not also use health services (e.g., early prenatal care), and vice versa.

With the objective of maximizing participation, the team used machine learning algorithms to develop predictions of which interventions are likely to be used by women. With this information, health and social service providers can coordinate recommended interventions and increase participation. Certain interventions, such as doula support, might be particularly effective at bridging the health care and social service systems.³

Which Interventions Reduce Risk the Most?

The causal inference models estimated

- the effects of each health and social service intervention on the risk of infant mortality
- the differences in each intervention's effects by women's characteristics (e.g., teenage mothers,

FIGURE 2
Correlation of Intervention Participation



NOTES: The X's indicate when the correlation between the two interventions is not significant at $p < 0.05$. FSC = Family Support Center.

black women, smokers, those with history of preterm deliveries)

- the combined effects of multiple interventions (e.g., benefits from WIC with home visiting).

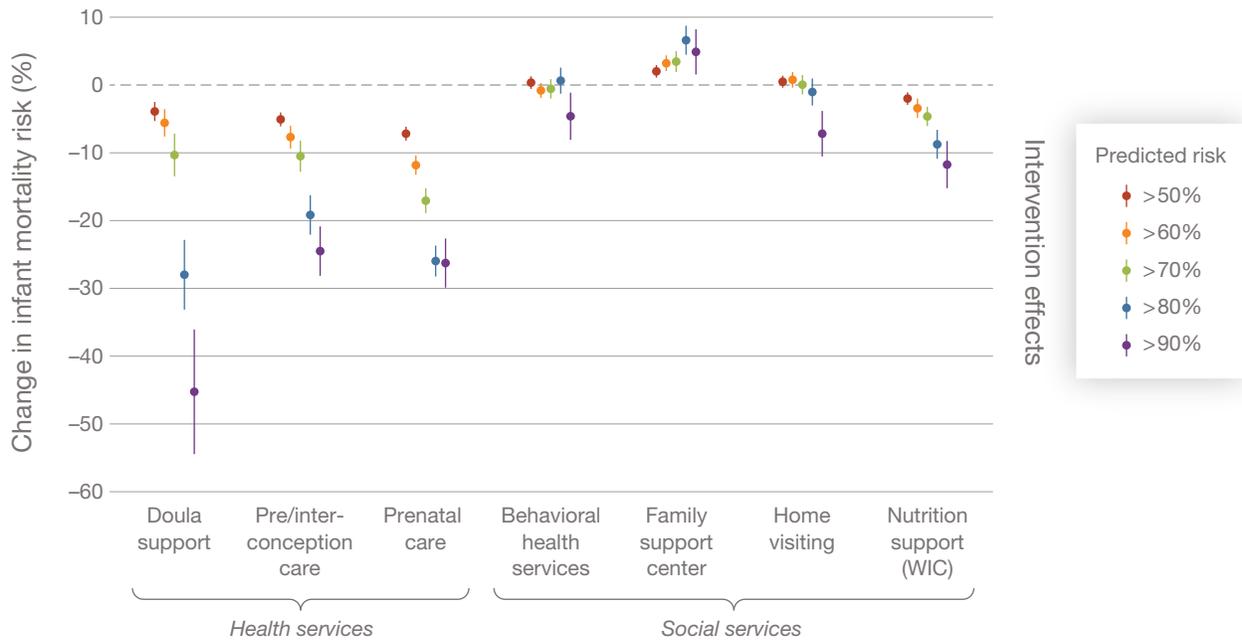
From these models, we found that multiple health services (e.g., doula support) and social services (e.g., home visiting) aimed at improving maternal and child health are effective, and other social services (e.g., mental health services) can also be effective depending on the risk level. Figure 3 shows the effects (dots) and 95-percent confidence intervals (lines) for each of the types of health and social service interventions by level of predicted risk (color). If estimates and the 95-percent confidence intervals are below the dotted line at 0 and do not cross it, the effects are statistically significant. Generally, the most-effective interventions we examined were those that targeted women's health before pregnancy, between pregnancies, or early in their pregnancies. The most-effective interventions were broad preconception care, frequent prenatal care, doula support, nutritional support, and home visiting. The effectiveness of each intervention tended to be highest

among those with the greatest predicted risk of infant mortality.

Some interventions are particularly effective for those with certain characteristics (e.g., black, teenage) and certain risk factors (e.g., obesity, history of smoking), suggesting that education and outreach could play a major role among women with these combinations of characteristics and risk factors. Figure 4 shows the effects (dots) and 95-percent confidence intervals (lines) for each of the types of health and social service interventions by maternal characteristics (noted by different colors). Generally, health services were effective among women with each of these characteristics, except doula support among first pregnancies and prenatal care among teenage women and women who ever smoked. Among the social services, WIC was effective among women with each of these characteristics except teenage mothers, and home visiting was most effective for black and teenage women.

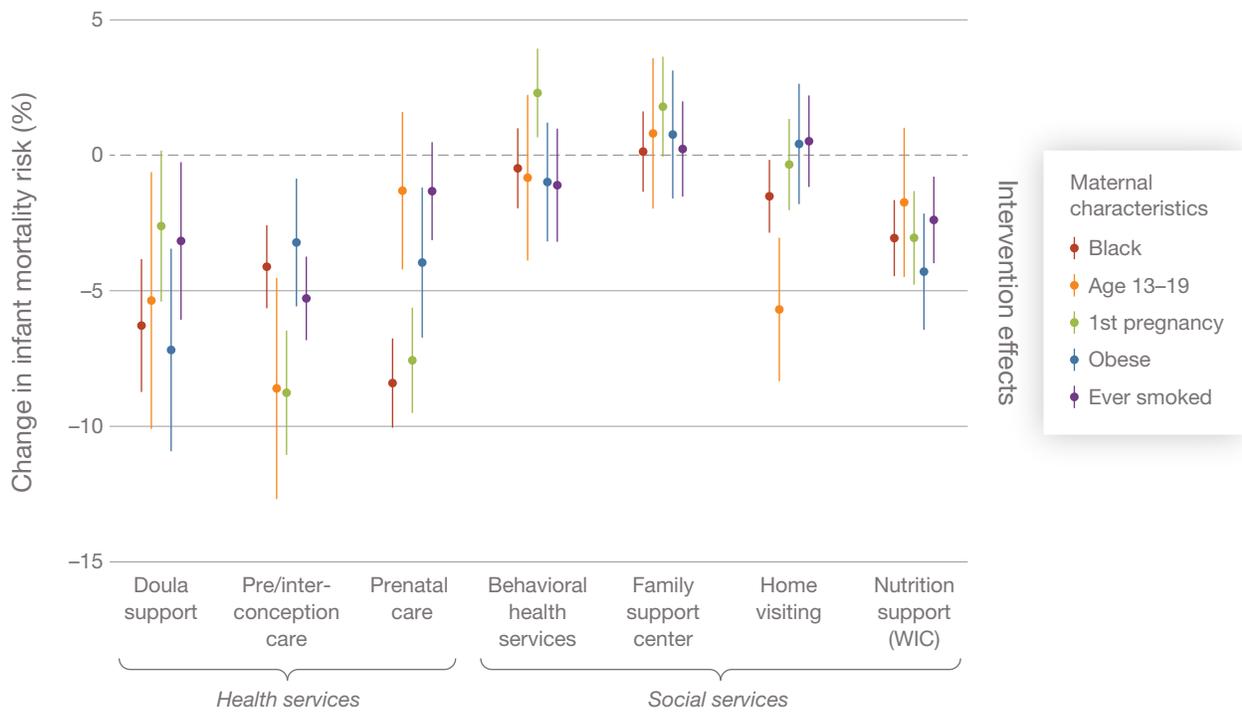
Additional results showed that certain combinations of interventions demonstrated synergistic effects. The most-effective combinations were those that more

FIGURE 3
Effects of Health and Social Service Interventions by Predicted Risk



NOTE: The dots represent the point estimates and the lines represent the 95-percent confidence intervals.

FIGURE 4
Effects of Health and Social Service Interventions by Maternal Characteristics



NOTE: The dots represent the point estimates (for those with greater than 50 percent predicted risk), and the lines represent the 95-percent confidence intervals (which are larger here because of the smaller sample sizes of women with the specified characteristics).

holistically address the individual needs of women, such as housing assistance and home visiting.

Putting the Evidence into Action

The research team is now working to mobilize lessons learned to provide real results for the women and children of Allegheny County. Tools are being developed and incorporated in health care providers' systems to provide integrated and continuous assessments of risk. Health care providers are receiving education about available community-based social services, and tools are being developed to track patient utilization of these services. Similarly, tools are being developed for community-based social service providers to inform them of their clients' risk and to suggest additional interventions to reduce that risk.

Because no intervention is effective if not used, recommendations need to leverage the data and predicted likelihood of use. Providers also need to use their interactions with patients to understand their thoughts regarding potential intervention options.

Using these results, health and social service providers might want to consider the following when making intervention recommendations for their patients before, during, or after pregnancy:

- Make intervention recommendations as early as possible.
- Tailor the recommendations based on the predicted risk, individual characteristics, and predicted participation.
- Potentially, recommend multiple interventions to maximize effectiveness.

For patients, mobile apps have been developed and are being tested. The apps will leverage the IMPreSIV database and aim to inform patients and provide a safe space for networking with other pregnant women and mothers. In addition, the apps will survey users to collect additional data to integrate with the IMPreSIV database. The research team is also working to validate the models and approaches for use in other health care systems.

To successfully launch these tools, our qualitative assessments of the local environment identified the need for community stakeholders to take the following additional steps.

- **Educate health care and community-based social service providers about available and**

effective interventions. Successful referrals to interventions often depend on health care and community providers' knowledge of programs, services, and supports in each other's domains.

- **Support outreach and engagement activities to increase participation for women who are at risk of poor birth outcomes.** Client-centered approaches could help maximize engagement.
- **Support the introduction or expansion of interventions that have demonstrated effectiveness in other settings and evaluate their effects in Allegheny County.** For instance, emerging evidence indicates that enhanced group prenatal care improves birth outcomes.
- **Focus maternal education and outreach efforts on the link between risk behaviors and poor outcomes.** Maternal smoking remains a concern, and substance abuse disorders have been growing among pregnant women.

Conclusion

This work provides actionable insights that are leading to tools that health care and community-based social services providers can use to improve their patients' and clients' health. Ultimately, the unique data and methods provide novel insights regarding maternal and child health, particularly to prevent the tragic outcome of infant mortality.

The integrative framework used in this project could be applied to other places and other complex problems by leveraging cross-sectoral data and using machine learning. Although the data are specific to Allegheny County, the methodological approach could be used in efforts to reduce infant mortality in other cities, states, and localities. Similarly, our approaches could be adapted to tackle other health outcomes besides infant mortality, such as cardiovascular disease, obesity, diabetes, and many others.

Notes

¹ Our World in Data, “Child Mortality vs GDP per capita, 2016,” webpage, undated. As of September 16, 2020: <https://ourworldindata.org/grapher/child-mortality-gdp-per-capita>

² J. A. Van Dijk, L. Anderko, and F. Stetzer, “The Impact of Prenatal Care Coordination on Birth Outcomes,” *Journal of Obstetric, Gynecologic and Neonatal Nursing*, Vol. 40, No. 1, 2011, pp. 98–108.

³ Emilie Lamberg Jones and Steven R. Leuthner, “Interdisciplinary Perinatal Palliative Care Coordination, Birth Planning, and Support of the Team,” in Erin M. Denney-Koelsch and Denise Côté-Arsenault, eds., *Perinatal Palliative Care: A Clinical Guide*, Cham: Springer, 2020, pp. 333–355.

This brief describes work done in RAND Social and Economic Well-Being and documented in *Examining Interventions to Address Infant Mortality in Allegheny County, Pennsylvania*, by Dana Schultz, Susan L. Lovejoy, and Evan D. Peet, RR-A858-1, 2021 (available at www.rand.org/t/RR-A858-1). To view this brief online, visit www.rand.org/t/RBA858-1. The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. RAND’s publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND**[®] is a registered trademark.

Limited Print and Electronic Distribution Rights: This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.

© 2021 RAND Corporation