

RESEARCH REPORT

Learning Health Systems

Analyzing electronic medical records to extract prepregnancy morbidities and pregnancy complications: Toward a learning health system

Yitayeh Belsti¹ | Lisa Moran¹ | Aya Mousa¹ | Rebecca Goldstein^{1,2} |
Daniel Lorber Rolnik^{2,3} | Mahnaz Bahri Khomami¹ | Mihiretu M. Kebede⁴ |
Helena Teede^{1,2} | Joanne Enticott¹

¹Monash Centre for Health Research and Implementation (MCHRI), Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia

²Monash Health, Melbourne, Victoria, Australia

³Department of Obstetrics and Gynaecology, School of Clinical Sciences, Monash University, Melbourne, Victoria, Australia

⁴German Cancer Research Center (DKFZ), Heidelberg, Germany

Correspondence

Joanne Enticott, Faculty of Medicine, Nursing and Health Sciences, Monash Centre for Health Research and Implementation (MCHRI), Monash University, Level 1, 43-51 Kanooka Grove Clayton, Melbourne, VIC 3168, Australia.
Email: joanne.enticott@monash.edu

Abstract

Introduction: Preexisting and pregnancy-related medical conditions frequently co-occur, leading to multimorbidity (≥ 2 morbidities) in pregnant women, and much of this information is in semi-structured format in electronic medical records (EMRs). The aim was to advance the learning health system as a platform for automating information extraction from EMRs and to uncover the prevalence of common morbidities during pregnancy and their association with pregnancy-related complications.

Methods: This study included 48 502 pregnant women attending Monash Health maternity hospitals from 2016 to 2021. Natural language processing (NLP) was used to extract morbidities from semi-structured text in EMRs. Chi-squared tests were used to assess the association between morbidities of gestational diabetes mellitus (GDM) and other pregnancy complications. The *k*-means clustering algorithm identified clusters of comorbid conditions associated with GDM.

Results: The most common comorbidities during pregnancy were vitamin deficiency (14 019; 28.9%), overweight (13 918; 28.7%), obesity (11 026; 22.7%), anemia and other blood-related disorders (4821; 9.9%), mental health disorders (4314; 9.8%), asthma (4126; 8.5%), thyroid diseases (3576; 7.4%), endometrial disease (1927; 3.9%), cardiovascular disease (1525; 3.1%), and polycystic ovary syndrome (PCOS) (1464; 3.0%). While 22.5% of women had no medical conditions, 77.5% had one or more. Multimorbidity was associated with conditions including overweight, obesity, vitamin deficiency, thyroid disease, substance use, PCOS, GDM, and endometrial diseases. On cluster analysis, aged 35 years or older, overweight, vitamin deficiency, obesity, thyroid disease, asthma, uterine disease, other blood disorders, mental disorders, and PCOS were associated with GDM.

Conclusions: More than three-quarters of pregnant women in the Australian urban setting experienced one or more morbidities during pregnancy, which can be associated with adverse pregnancy outcomes. This project contributes to developing a

learning health system infrastructure to deliver high-value maternal health care while reducing costs.

KEYWORDS

comorbidities, gestational diabetes mellitus, learning health system, natural language processing

1 | INTRODUCTION

Pregnancy is a complex physiological process involving several critical stages, such as ovulation, fertilization, implantation, decidua formation, placentation, organogenesis, establishing the maternal-fetal interface, fetal growth, and, ultimately, parturition.¹ Completing each stage is vital for achieving a successful pregnancy. However, preexisting and pregnancy-related medical conditions are prevalent and can interfere with these stages, often occurring as simultaneous comorbidity in pregnant women.^{2–4}

Over time, the average maternal age for primigravidas increased from 28.4 years in 2011 to 29.7 years in 2021. For multigravidas, it rose from 31.3 years in 2011 to 32.2 years in 2021 in Australia, similar to other high-income countries.⁵ There is also an alarming rise in average prepregnancy body mass index (BMI) globally.⁶ This upward trend of prepregnancy risk factors is accompanied by significant physiological and metabolic alterations⁷ and the development of chronic diseases, including type 2 diabetes mellitus (T2DM), hypertension, and dyslipidemia.^{8,9} As a result, 56% of Australian women have at least one chronic condition.¹⁰

When a body that is already disturbed metabolically with chronic disease is exposed to pregnancy, which further alters body physiology,¹¹ the preexisting metabolic condition might be exacerbated, or a woman may develop new metabolic disturbances like gestational diabetes mellitus (GDM) or preeclampsia.^{12,13} GDM is a common, preventable cardiometabolic condition in pregnancy, affecting approximately 15% of women globally, with significant geographical variation and an increasing trend over time.^{14,15} If such women remain undiagnosed, unmonitored, and untreated, it will often lead to maternal, fetal, and newborn complications, including cesarean births, low birth weight, preterm birth, stillbirths, and even death.^{16,17} Therefore, it is essential to understand the distribution and interactions of common medical conditions (morbidities) and maternal characteristics in pregnancy. Identifying common contributing factors and adverse synergistic interactions between morbidities and maternal factors during pregnancy is essential to designing feasible and effective interventions. A key opportunity is to leverage electronic medical records (EMRs) to analyze the burden of these medical conditions, their interactions, and their impact on pregnancy outcomes.

Since maternal health is affected by different conditions emerging from a wide range of dimensions,^{18–20} health information from all parts of a woman's history could help inform the provision of optimal antenatal care.^{21,22} Despite the existing technological advances, multi-dimensional, multi-provider, and multidisciplinary cooperation and systems to provide comprehensive health information have not yet

been fully implemented by existing healthcare systems in many countries, including Australia.²³ A term first coined by the US Institute of Medicine (IOM) called the learning health system (LHS)²⁴ provides a model to facilitate health systems to move toward providing the best care, personalized for each woman. LHS models envision developing a continuous learning health system by aligning science, analytics, informatics, incentives, and culture for continuous improvement and innovation, looking for the best practice by actively participating in all elements and gaining new knowledge from daily experience.²⁵ In pregnancy care, such an LHS approach could conduct sophisticated analytics to extract relevant information from the electronic health records of a woman and provide this information in user-friendly formats to be used by antenatal care providers to tailor the best high-value care for the woman and her offspring.²⁶

In line with advanced technologies, modern medicine increases human understanding of disease and produces innovative treatment strategies. However, due to the complex nature of some diseases and variations in patient characteristics, experimental studies are costly to conduct and slow to put into practice.^{27,28} Therefore, the available health system is hungry for a system through which continuous and unbiased new knowledge flows and helps individual pregnancy care practice through daily experience.

The aim was to advance the LHS as a platform for automating information extraction from EMRs from an ethnically diverse cohort of pregnant women. The objective was to uncover the prevalence of and patterns between common morbidities during pregnancy and pregnancy-related complications and identify clusters of comorbid conditions associated with GDM, the most common metabolic complication in pregnancy.^{29,30} With rising maternal comorbidities and costs of maternity care in Australia,³¹ such information is timely, as investments and infrastructure for such LHS approaches will be necessary to inform and enable monitoring, prevention, and improved maternity care.

2 | METHODS

2.1 | Study population and data sources

The study utilized routinely gathered health data from pregnant women who received care at Monash Health maternity hospitals in Melbourne, Australia, between January 2016 and June 2021. Monash Health, one of Australia's most extensive public health networks, caters to over two million individuals throughout their lives. It encompasses eight hospitals and manages more than 12 000 births annually.

As part of Australia's universal, freely accessible public healthcare system, Monash Health serves women from over 78 countries with a large immigrant population and 65% of mothers born overseas. Study findings are reported in line with the REporting of studies Conducted using the Observational Routinely-collected Data (RECORD) statement.³²

2.2 | Data structure and terminologies

In this study, “morbidity” is defined as a preexisting medical condition, comorbidity as any coexisting medical condition, and multimorbidity as having two or more medical conditions. The term hypertension in this paper means preexisting medical conditions of hypertension. Pregnancy-induced hypertension and preeclampsia are defined as pregnancy-related complications. Semi-structured data: documented in free text form in the dataset are factors such as maternal medical conditions, antenatal medications, substance use during pregnancy, obstetric complications, labor complications, indication for induction, birth defects, and immediate post-labor complications. The data were recorded by clinicians, midwives, and data personnel working at the front line who are responsible for antenatal care. The standard operating procedure for this is that information deemed relevant to the current care is sought and entered into the medical record. This can be done by directly asking the woman questions, or it could be information volunteered by the woman herself. Semi-structured is defined as data that falls between structured and unstructured data, has self-descriptive properties, and is arranged with limited structural constraints, typically in the form of extended markup language.³³ To enhance data accessibility, usability, and knowledge discovery, we applied basic Natural Language Processing (NLP) to convert unstructured text into a structured format. Structured data: The variables “overweight” (25.0 to <30 kg/m²) and “obesity” (≥ 30 kg/m²) were derived from BMI data, and “advanced age,” defined as an age of 35 years or older, was created using maternal age data.

2.3 | Data cleaning and analysis

Analysis was performed in Python (version 3.10.5) using various packages. Since within the dataset, all columns record disease names only if the pregnant women have any condition and are left empty otherwise, we combined rule-based extraction with normalization and fuzzy matching.

The pandas library was used for data manipulation and analysis. Text cleaning, sentence tokenization, and word replacement were facilitated by the NLTK package (version 3.7)³⁴ and the *re* package (version 3.10.5)³⁵ for handling regular expressions. Both *CountVectorizer* and *TfidfVectorizer* from the *sklearn.feature_extraction.text* module were utilized for converting text documents into matrices of token counts and *Term Frequency-Inverse Document Frequency* (TF-IDF)

features, respectively. Topic modeling was performed using the Latent Dirichlet Allocation method from the *sklearn.decomposition* module, while the *sklearn.cluster* module's *KMeans* function enabled clustering analysis. The Spacy library, along with its *Matcher* module, was employed for advanced NLP tasks such as pattern matching. Data visualizations were created using the *matplotlib.pyplot* library and the *FreqDistVisualizer* from the *yellowbrick.text.freqdist* module.

The Natural Language Toolkit (NLTK) package was utilized for tasks such as tokenization, stemming, lemmatization, and parsing of texts. The “string” and “re” modules were employed for string formatting, searching, manipulation, and for dealing with regular expressions. The text underwent a cleaning process to remove punctuations, numbers, and whitespaces. Additionally, all words were converted to lowercase for consistency. Word tokenization was then performed to split the text into separate words or tokens, enabling a better understanding of the context. To enhance the quality of analysis, common stop words (e.g., “a,” “an,” “and,” “are,” “as,” “at,” “be,” “by,” “for,” “from,” “has,” “he,” “in,” “is,” “it,” “its,” “of,” “on,” “that,” “the,” “to,” “was,” “were,” “will,” “with,” etc.) that have little contribution to the overall meaning of the text and do not contribute significant meaning to the analysis were eliminated from the text data. To gain insights into the distribution and importance of words, the frequencies of the most common words were calculated. This analysis provided an overall understanding of the prominence and relevance of words in the text data. Finally, the text data were visualized using WordCloud, a technique that presents a quick overview of key terms and themes. The WordCloud representation allows for visual identification of the most prominent words, enabling viewers to grasp the main focus or subject of the text rapidly. The size of each word in the WordCloud is related to the frequency of the word (larger font indicates greater frequency), and it provided an important visual figure that enabled us to get an overarching insight into relevant themes and to identify important keywords before further analysis.

Keywords for each common medical condition were extracted by examining the top 1000 most frequently mentioned medical conditions. As an example, for mental disorders, the identified keywords included “anxiety,” “depression,” “panic,” “psychological,” “stress,” “bpad,” “bipolar,” “personality,” “borderline,” “schizophrenia,” “intellectually,” “bpd,” “fatigue,” “ptsd,” “psychosis,” “bulimia,” “psych,” “substance,” “suicidal,” “borderline,” and “schizoaffective.” Here, “bpad” refers to bipolar disorder; “bpd” to borderline personality disorder; and “ptsd” to post-traumatic stress disorder. These keywords were used to determine whether there was a diagnosis of a mental disorder in a given patient record. The “Spacy” matcher Python module was utilized for this purpose. Keywords used to extract each disease are indicated in the Data S1.

Due to the nature of the dataset, which does not include negative or uncertain contexts, we did not find it necessary to use more advanced NLP techniques to handle negative and uncertain contexts. We rather employed rule-based extraction using MedSpaCy with *TargetRule* for disease extraction. This method is straightforward and

effective for predefined lists of terms, but it may miss variations not covered by the rules. To avoid missing any conditions, we printed the top 1500 words available in the columns and extracted all relevant terms to capture variations. We used normalization to convert text to lowercase, ensuring consistency and reducing variability due to capitalization. We then applied fuzzy matching using the FuzzyWuzzy library to handle minor typos and variations. This approach is useful for matching terms that are similar but not exact, which is important for dealing with real-world data where spelling errors are common.

The frequencies of the most common words were calculated to gain insights into the distribution and importance of words. Common medical conditions were extracted from the text based on word frequency.

After extracting common medical conditions, their associations with GDM and common obstetric, labor, and fetal complications were calculated. The chi-square statistic was calculated to assess the degree of association between morbidities and each pregnancy complication. Morbidity comparisons were also performed between (a) pregnant women who had GDM and (b) pregnant women who had no GDM. A *p*-value of <0.05 was considered statistically significant. To understand the strength of the association, we used the Phi coefficient, which is derived from the chi-square statistic.

2.4 | Cluster analysis

Cluster analysis was performed to group a set of factors so that those in the same group (or cluster) are more similar to each other than those in different groups. The K-means clustering algorithm was applied, and the elbow method determined the number of clusters. As the most common cardiometabolic complication in pregnancy and the context of a research program focused on GDM risk identification, prevention, and management, the risks of GDM across various clusters were also assessed after creating isolated columns representing clusters.

3 | RESULTS

This analysis incorporated a total of 48 502 pregnancies. More than half of the participants had a BMI exceeding 25 kg/m², and 22.5% were over the age of 35 years. Common medical conditions were vitamin deficiency (*n* = 14 019; 28.9%), overweight (*n* = 13 918; 28.7%), obesity (*n* = 11 026; 22.7%), anemia and other blood-related disorders (*n* = 4821; 9.9%), mental health disorders (*n* = 4314; 9.8%), asthma (*n* = 4126; 8.5%), thyroid diseases (*n* = 3576; 7.4%), endometrial disease (*n* = 1927; 3.9%), cardiovascular disease (CVD) (*n* = 1525; 3.1%), PCOS (*n* = 1464; 3.0%), iron deficiency anemia (*n* = 1308; 2.7%), migraine (*n* = 1274; 2.6%), urinary tract infections (*n* = 984; 2.0%), diabetes mellitus (*n* = 743; 1.53%), musculoskeletal disease (*n* = 666; 1.4%), hypertension (*n* = 581; 1.2%), and kidney diseases (*n* = 392; 0.8%) (Figure 1).

Common complications included GDM (*n* = 10 343; 21.3%), induced birth (*n* = 14 623; 30.1%), cesarean section (*n* = 14 027; 28.9%), admission of newborn to special care nursery or antenatal care unit (*n* = 8848; 18.1%), macrosomia (*n* = 4308; 8.9%), preterm delivery (*n* = 4066; 8.4%), preeclampsia (*n* = 1698; 3.5%), shoulder dystocia (*n* = 1069; 2.2%), and birth defect (*n* = 2464; 5.1%).

3.1 | Frequency of single morbidity and multimorbidity

The number of conditions experienced during pregnancy ranged from none to nine. A relatively low proportion of women (*n* = 10 890; 22.5%) had no captured conditions, while (*n* = 18 143; 37.4%) had one medical condition and (*n* = 12 467; 25.7%) had contended with two conditions. The percentage continued to decline as the number of conditions increased, with (*n* = 5131; 10.6%) having three conditions, (*n* = 1385; 2.9%) having four, and (*n* = 1; <0.1%) having nine conditions (Figure S1).

In the total cohort, (*n* = 37 612; 77.5%) had ≥1 comorbidity during pregnancy, of which (*n* = 18 143; 37.4%) had a single medical condition while (*n* = 19 469; 40.1%) had multimorbidity. Of 10 913 pregnant women aged 35 years and above, this increased to (*n* = 8731; 18.0%) having ≥1 comorbidity, with (*n* = 3932; 8.1%) experiencing a single morbidity, while (*n* = 4799; 9.9%) presented with multimorbidity. All 24 943 pregnant women with a BMI of ≥25 kg/m² had ≥1 comorbidity, with (*n* = 9525; 19.6%) having a single morbidity and (*n* = 15 418; 31.8%) with multimorbidity (Table 1).

3.2 | Association of morbidities with complications

The top 40 significant correlations are presented in Figures 2 and S2. It shows critical associations such as obesity with GDM, macrosomia, hypertensive disorders of pregnancy, preeclampsia, and cesarean section; preexisting diabetes with neonatal special care admission and type of delivery; and preexisting maternal hypertension with preeclampsia, among others.

The top three complications associated with each morbidity are presented in Figure 3. Overweight, obesity, vitamin deficiency, thyroid disease, substance use, PCOS, and endometrial diseases are among the morbidities most associated with GDM (Table 2).

3.3 | Cluster analysis

The text clustering analysis resulted in five distinct clusters based on the medical conditions of pregnant women. Figure 4 presents the cluster analysis visually.

In assessing the risk of GDM across various clusters, it was found that 29.4% of women in Cluster 0 developed GDM,

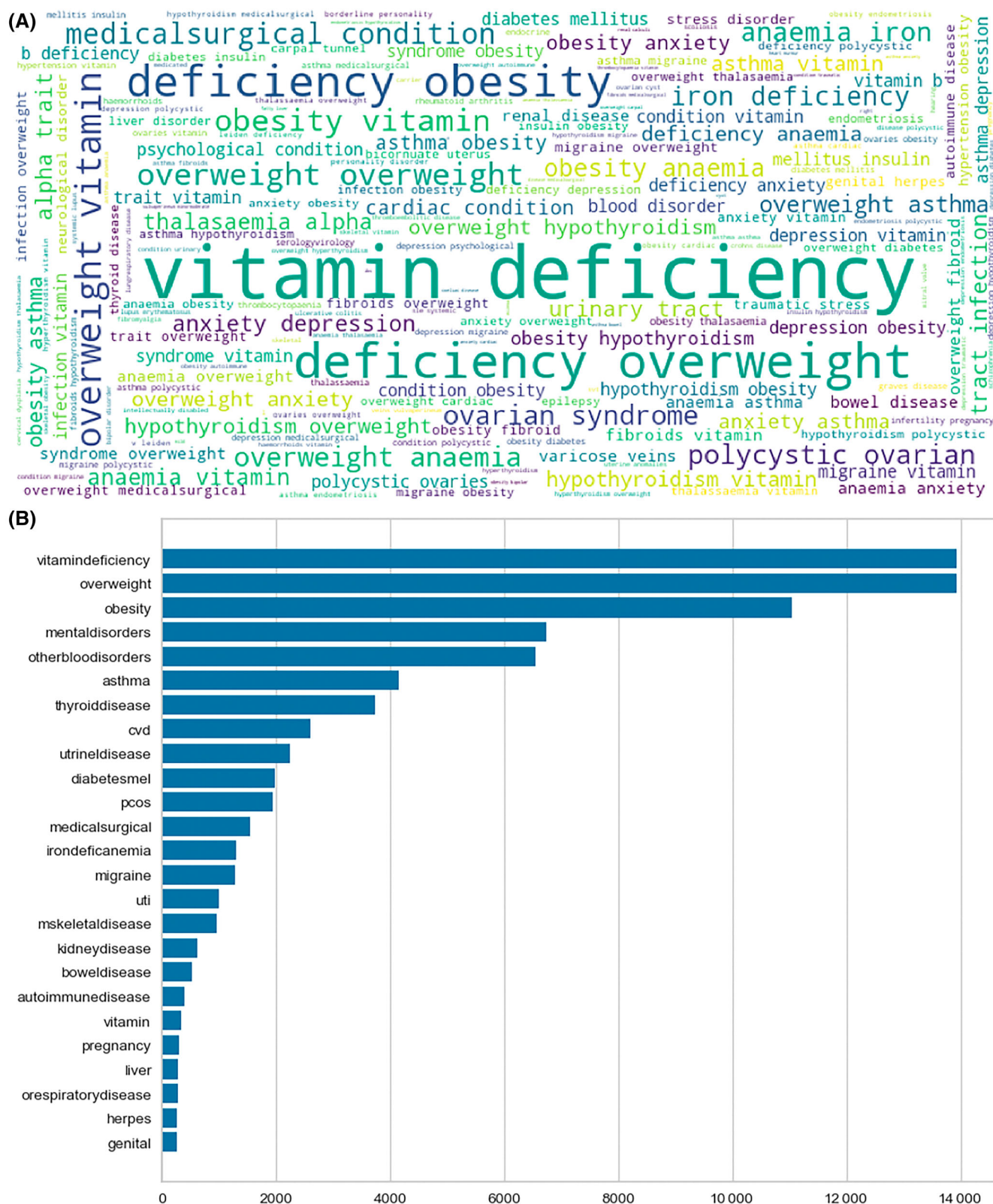


FIGURE 1 Common pregnancy comorbidities. (A) Top panel: WordCloud of common pregnancy comorbidities. (B) Bottom panel: Top 25 commonly occurring medical condition tokens.

17.5% in Cluster 1, 29.5% in Cluster 2, 21.5% in Cluster 3, and 14.6% in Cluster 4. There was no significant difference in the incidence of GDM between clusters 2 and 0 ($p = 0.86$).

However, statistically significant differences between the remaining cluster combinations indicated varying GDM risks among the clusters.

TABLE 1 The distribution of comorbidities across baseline characteristics of participants and selected obstetric complications.

Characteristics	Total births (N = 48 502)	No morbidity (N = 10 890)	Single morbidity (N = 18 143)	Multimorbidity ^a (N = 19 469)
Age (years)				
≤24	6176 (12.7%)	1347 (12.4%)	2348 (12.9%)	2481 (12.7%)
25–29	13 806 (28.5%)	3239 (29.7%)	5243 (28.9%)	5324 (27.3%)
30–34	17 607 (36.3%)	4122 (37.9%)	6620 (36.5%)	6865 (35.3%)
35–39	8830 (18.2%)	1823 (16.7%)	3191 (17.6%)	3816 (19.6%)
≥40	2083 (4.3%)	359 (3.3%)	741 (4.1%)	983 (5.0%)
BMI (kg/m ²)				
≤19.9	4169 (8.6%)	1976 (18.1%)	1503 (8.3%)	690 (3.5%)
20–24.9	19 144 (39.5%)	8798 (80.8%)	7042 (38.8%)	3304 (17.0%)
25–26.9	6632 (13.7%)	0 (0.0%)	2785 (15.4%)	3847 (19.8%)
27–29.9	7286 (15.0%)	0 (0.0%)	2969 (16.4%)	4317 (22.2%)
30–34.9	6276 (12.9%)	0 (0.0%)	2301 (12.7%)	3975 (20.4%)
35+	4749 (9.8%)	0 (0.0%)	1470 (8.1%)	3279 (16.8%)
Ethnicity				
Caucasian	1852 (3.8%)	166 (1.5%)	713 (3.9%)	973 (5.0%)
Oceania (not White-Australian or White-New Zealander)	21 256 (43.8%)	4057 (37.3%)	7683 (42.3%)	9516 (48.9%)
Middle-Eastern, North African, or Sub-Saharan African	2967 (6.1%)	549 (5.0%)	1159 (6.4%)	1259 (6.5%)
Southern and Central Asian	14 549 (30.0%)	3108 (28.5%)	5588 (30.8%)	5853 (30.1%)
South-East and North-East Asian	7778 (16.0%)	2989 (27.4%)	2970 (16.4%)	1819 (9.3%)
Other	100 (0.2%)	21 (0.2%)	30 (0.2%)	49 (0.3%)
Parity				
0	19 357 (39.9%)	4583 (42.1%)	7166 (39.5%)	7608 (39.1%)
1	17 618 (36.3%)	4231 (38.9%)	6626 (36.5%)	6761 (34.7%)
≥2	11 527 (23.8%)	2076 (19.1%)	4351 (24.0%)	5100 (26.2%)
Substance use	6033 (12.4%)	878 (8.1%)	2091 (11.5%)	3064 (15.7%)
Preeclampsia	1698 (3.5%)	192 (1.8%)	586 (3.2%)	920 (4.7%)
Pregnancy-induced Hypertension	1113 (2.3%)	102 (0.9%)	409 (2.3%)	602 (3.1%)
Gestational diabetes mellitus	10 343 (21.3%)	1608 (14.8%)	3768 (20.8%)	4967 (25.5%)
Preterm birth	4066 (8.4%)	783 (7.2%)	1413 (7.8%)	1870 (9.6%)
Induction of labor	14 623 (30.1%)	2962 (27.2%)	5257 (29.0%)	6404 (32.9%)
Cesarean delivery	14 027 (28.9%)	2415 (22.2%)	5064 (27.9%)	6548 (33.6%)
Birth defect	2464 (5.1%)	537 (4.9%)	918 (5.1%)	1009 (5.2%)
Admission to a special nursery or neonatal intensive care unit	8848 (18.1%)	1604 (14.7%)	3117 (17.2%)	4127 (21.2%)
Shoulder dystocia	1069 (2.2%)	210 (1.9%)	415 (2.3%)	444 (2.3%)
Macrosomia	4308 (8.9%)	701 (6.4%)	1721 (9.5%)	1886 (9.7%)

^aHaving ≥2 comorbidities.

4 | DISCUSSION

We extracted preexisting and pregnancy-related medical conditions in pregnant women, and much of this information was found in semi-structured format in EMRs. Overall, the extracted information showed that the burden of morbidity was high. The most common morbidities during pregnancy included vitamin deficiency, overweight, obesity,

anemia and other blood-related disorders, mental health disorders, asthma, thyroid diseases, endometrial disease, CVD, PCOS, and iron deficiency anemia. Associations such as obesity with GDM, macrosomia, hypertensive disorders of pregnancy, preeclampsia, and cesarean section; preexisting diabetes with neonatal special care admission and type of delivery; and preexisting maternal hypertension with preeclampsia, among others, were observed.

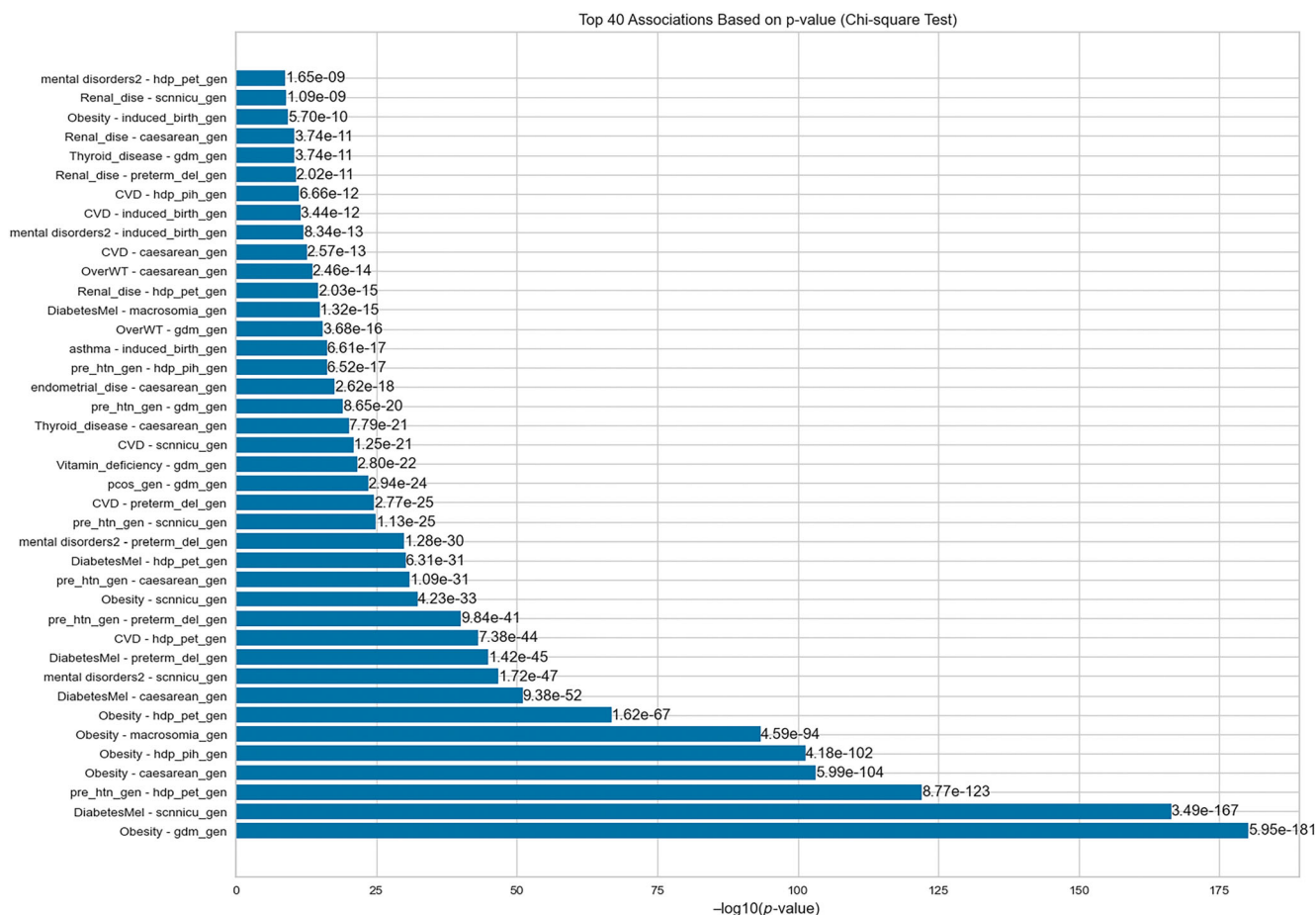


FIGURE 2 Top 40 associations based on p -value (Chi-squared test). The level of statistical significance decreases from bottom to top. CVD, cardiovascular disease; D.M., diabetes mellitus; GDM, gestational diabetes mellitus; PCOS, polycystic ovary syndrome; PIH, pregnancy-induced hypertension; SCN_NICU, admission of newborn to special care nursery or antenatal care unit.

To advance our large public antenatal services in Australia toward an LHS infrastructure that can automate information extraction from EMRs, this work was codeveloped with our multidisciplinary clinician coauthors, including an obstetrician (D.L.R.), a midwife (M.B.K.), two endocrinologists (H.T., R.G.), and a dietitian (L.M.), who provide expert care to pregnant women. This study outlines the service and patient outputs that frontline clinicians require in a timely manner from an LHS infrastructure. Ideally, the work conducted in this project would be automated within an LHS to provide regular updates on these outputs, allowing clinicians and administrators to monitor outcomes and implement both systemic (organization-wide) and individual interventions as needed. Clinicians have also indicated that disaggregating data by each treating clinician or team would be valuable, provided it occurs in an environment that supports professional development and is not used for punitive purposes.³⁶ Ultimately, embedded analytics within an LHS would utilize this data to enable predictive risk stratification of high-risk patients and real-time clinical decision support for tailored patient interventions. The curated data were used as inputs for developing and temporally validating our risk prediction models,^{29,37,38} which are intended to enhance decision-making in personalized maternity care within our LHS.

The number of medical conditions during pregnancy varied from none to as many as nine. Our results indicated that a significant proportion, approximately 77.5%, of pregnant women experience at least one morbidity during pregnancy. Of these, 37.4% had a single medical condition, while 40.1% had multimorbidity. This finding is consistent with an emerging body of evidence that recognizes the high prevalence of morbidities among pregnant women.³⁹ In contrast, other studies have shown the prevalence of multimorbidity ranging from 0.83% to 46.2%, which is lower than our findings.^{40–43} This might be due to differences in the characteristics of the populations studied, inclusion and exclusion criteria, and the spectrum of medical conditions considered in defining morbidity.

This high level of morbidity, combined with advanced maternal age, contributes to the rapidly rising rates of pregnancy complications, including induction of labor and cesarean delivery.⁴⁴ The health burden and cost of these conditions and the related pregnancy impact are substantial, with recent evidence showing a 60% increase in Australian public maternity costs over the last 5 years alone.⁴⁴ This is unsustainable, and understanding these morbidities will assist with prevention. For example, lifestyle interventions, which are cost-effective and recommended for implementation for prevention, are

Top 3 Complications for Each Comorbidity Based on p-value (Chi-square Test)

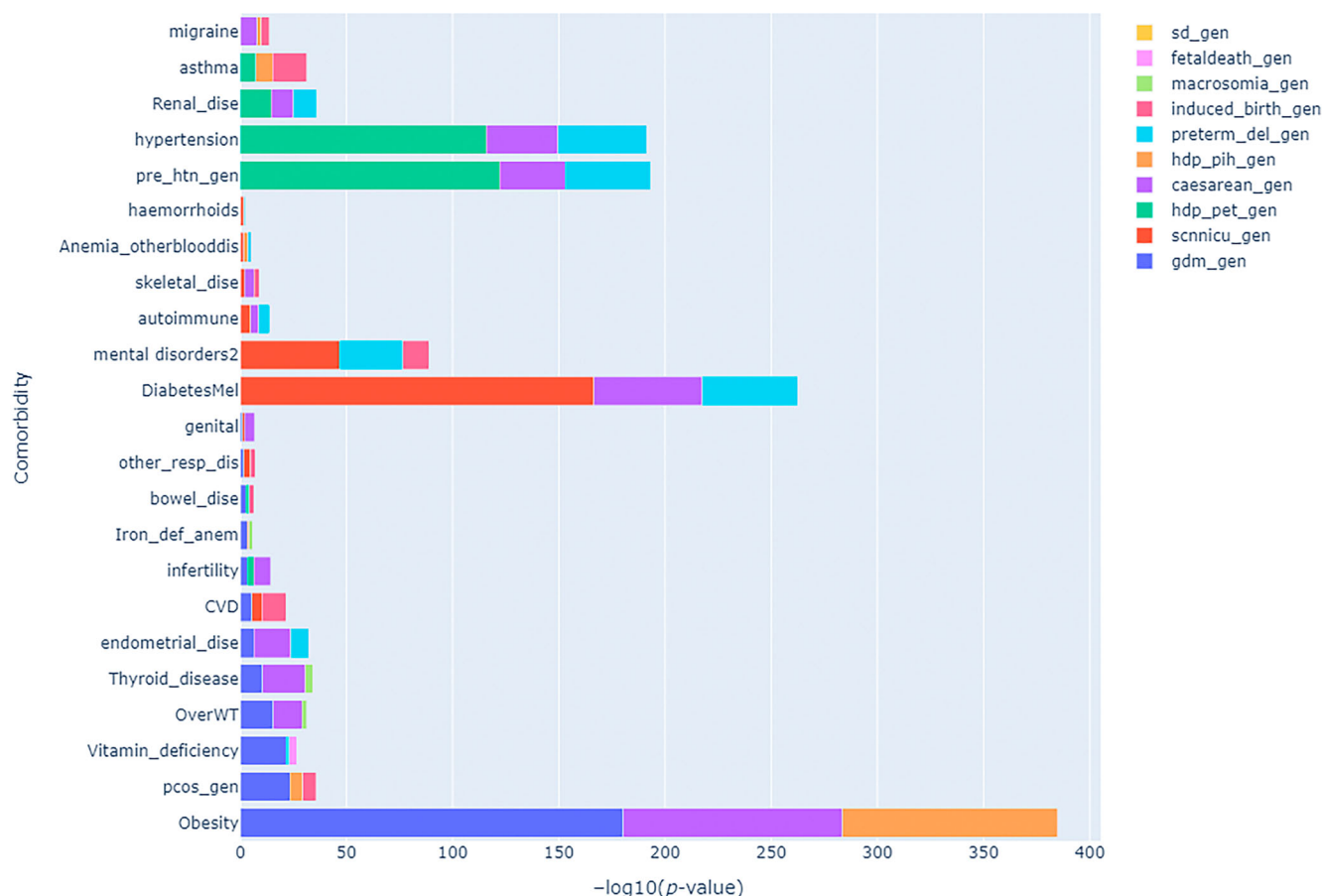


FIGURE 3 The top three complications associated with each comorbidity. CVD, cardiovascular disease; D.M., diabetes mellitus; GDM, gestational diabetes mellitus; PCOS, polycystic ovary syndrome; PIH, pregnancy-induced hypertension; SCN_NICU, admission of newborn to special care nursery or antenatal care unit.

proven to reduce many of these complications in pregnancy.^{15,31,45} Implementation remains limited, with a persistent focus on costly treatments. Appreciation of these data also recognizes the need to significantly optimize healthcare services during pregnancy to protect Australia's mothers and babies.

The variability in the number of morbidities during pregnancy was striking, with 22.5% of women not experiencing any medical conditions. While this suggests many pregnancies may be categorized as low-risk, 37.4% of women had one medical condition, and 25.7% had two, necessitating medical attention.⁴⁶ About 14% had three or more medical conditions, and this number is significantly higher when looking at subgroups with high BMI and increased age. This suggests that to achieve the most significant impact, these groups are targeted with more comprehensive antenatal care capable of managing multiple concurrent conditions.⁴⁷ It has been reported that advanced maternal age is more likely to be associated with a BMI of 25 or higher, as well as with other comorbid conditions.^{17,48}

The NLP methods applied to the EMRs can enhance the learning health system loop by transforming semi-structured clinical data into structured, actionable data, which enables the extraction of valuable

clinical insights that can be used for decision support, research, and improving patient care.⁴⁹ It enables the unlocking of previously hidden clinical information, increases data heterogeneity, and aids in predicting and preventing diseases.⁵⁰ Integrating NLP with machine learning models, such as logistic regression and BERT, has shown high efficiency in diagnosing disease groups and can be used for preliminary health assessments.⁵¹ Deep learning and convolutional neural networks further enhance the ability to detect relationships between diseases, symptoms, and treatments, thereby facilitating better clinical decision-making.⁵² By developing annotation-free NLP pipelines, health systems can automatically derive precise clinical codes from pathology reports, thus enabling a self-learning system that continuously improves from historical and emerging data.⁵³

4.1 | Implications of the study and the learning health system

Our study offers valuable insights into the prevalence and associations of morbidities and maternal characteristics with obstetric

TABLE 2 Medical conditions and their association with gestational diabetes mellitus.

Medical conditions	Gestational diabetes		Chi-squared	p-value
	No	Yes		
Overweight ^a				
No	27 542 (56.8)	7042 (14.5)	66.4	< 0.001
Yes	10 617 (21.9)	3301 (6.8)		
Obesity ^a				
No	30 570 (63.0)	6907 (14.2)	822.8	< 0.001
Yes	7589 (15.6)	3436 (7.1)		
Common mental disorders				
No	34 716 (71.6)	9472 (19.5)	3.6	0.056
Yes	3443 (7.1)	871 (1.8)		
Polycystic ovary syndrome				
No	37 164 (76.6)	9874 (20.4)	103.3	< 0.001
Yes	995 (2.1)	469 (1.0)		
Cardiovascular disease				
No	36 991 (76.6)	9986 (20.6)	3.9	0.046
Yes	1168 (2.4)	357 (0.7)		
Substance use				
No	33 079 (68.2)	9390 (19.4)	125.1	<0.001
Yes	5080 (10.5)	953 (2.0)		
Thyroid disease				
No	35 502 (73.2)	9424 (19.4)	43.7	<0.001
Yes	2657 (5.5)	919 (1.9)		
Urinary tract infections				
No	37 386 (77.1)	10 132 (20.9)	0.0	0.958
Yes	773 (1.6)	211 (0.4)		
Hypertension				
No	37 792 (77.1)	10 129 (20.9)	83.4	< 0.001
Yes	367 (0.8)	214 (0.4)		
Vitamin deficiency				
No	27 527 (56.8)	6956 (14.3)	94.2	<0.001
Yes	10 632 (21.9)	3387 (7.0)		
Asthma				
No	34 871 (71.9)	9505 (19.6)	2.7	0.100
Yes	3288 (6.8)	838 (1.7)		
Other respiratory disease				
No	37 978 (78.3)	10 273 (21.2)	6.1	0.013
Yes	181 (0.4)	70 (0.1)		
Anemia and other blood disorders				
No	34 399 (70.9)	9282 (19.1)	1.4	0.2
Yes	3760 (7.8)	1061 (2.2)		
Iron deficiency anemia				
No	37 079 (76.4)	10 115 (20.9)	0.6	0.440
Yes	1080 (2.2)	228 (0.5)		
Endometrial disease				
No	36 732 (75.7)	9843 (20.3)	25.3	<0.001
Yes	1427 (2.9)	500 (1.0)		

(Continues)

TABLE 2 (Continued)

Medical conditions	Gestational diabetes		Chi-squared	p-value
	No	Yes		
Migraine				
No	37 182 (76.7)	10 046 (20.3)	1.0	0.314
Yes	977 (2.0)	297 (0.6)		
Kidney disease				
No	37 829 (78.0)	10 281 (21.2)	6.8	0.009
Yes	330 (0.7)	62 (0.1)		
Bowel disease				
No	37 903 (78.1)	10 301 (21.2)	8.9	0.0028
Yes	256 (0.5)	42 (0.1)		
Genital herpes				
No	37 899 (78.1)	10 289 (21.2)	2.9	0.085
Yes	260 (0.5)	54 (0.1)		
Hemorrhoids				
No	37 963 (78.3)	10 282 (21.2)	0.8	0.384
Yes	196 (0.4)	61 (0.1)		
Autoimmune disease				
No	37 897 (78.1)	10 270 (21.2)	0.0	0.887
Yes	262 (0.5)	73 (0.2)		
Musculoskeletal disease				
No	37 628 (77.6)	10 208 (21.0)	0.4	0.534
Yes	531 (1.1)	135 (0.3)		
Infertility				
No	38 077 (78.5)	10 300 (21.2)	12.0	<0.001
Yes	82 (0.2)	43 (0.1)		
Multimorbidity				
No	23 612 (48.7)	5356 (11.0)	344.3	<0.001
Yes	14 547 (30.0)	4987 (10.0)		
Advanced age (≥35 years)				
No	30 476 (62.8)	7113 (14.7)	573.8	<0.001
Yes	7683 (15.8)	3230 (6.7)		

^aOverweight and obesity information were determined from prepregnancy BMI data and were mutually exclusive.

complications, thereby enriching the field of obstetric research. Despite the increasing prevalence of morbidity risk factors like advanced maternal age and obesity, healthcare services are generally designed around managing individual chronic conditions.⁵⁴ Since the collective impact of multimorbidity differs significantly from that of a single condition, effective preconception management and maintaining a coordinated and focused healthcare approach throughout the antenatal period are required to prevent unfavorable outcomes. Evidence from non-pregnancy contexts suggests collaborative, multidisciplinary approaches effectively manage multimorbidity.⁵⁴ These include guidelines and strategies for dealing with common clusters of conditions, necessitating a deep understanding of their co-occurrence and potential issues

like drug-drug interactions.^{55,56} The development and implementation of such strategies are vital, given the increasing prevalence of multimorbidity risk factors. The ability to gather valuable data from semi-structured EMRs using NLP could support shared decision-making and save resources as unnecessary time-intensive tasks and repeated tests are avoided. Critical multimorbidity data are made readily available.^{26,57}

The structured data from this study enabled us to analyze frequency, trends, and associations between risk factors and pregnancy complications and explore new insights; beyond associations, we also conducted clustering analyses, which are vital exploratory tools that can uncover new patterns and insights. The findings from frequency associations, trend analyses, and

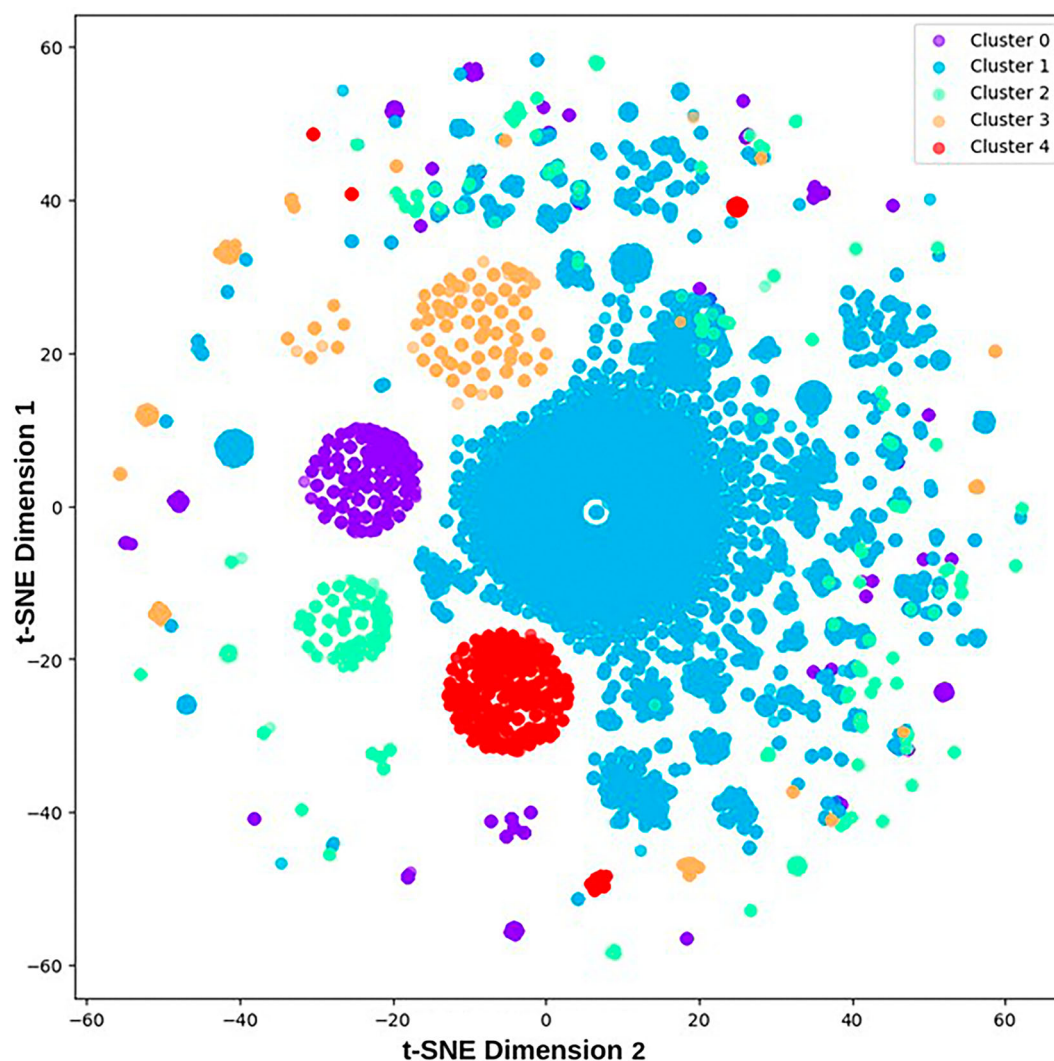


FIGURE 4 The cluster analysis results scatter plot shows those who share similar comorbidities during their pregnancies (each point representing a pregnant woman). Cluster 0: obesity, vitamin deficiency, asthma, other blood disorders, advanced age, mental disorders, thyroid disease, PCOS, migraine, urinary tract infection; Cluster 1: mental disorders, other blood disorders, asthma, thyroid disease, overweight, vitamin deficiency, uterine disease, PCOS, iron deficiency anemia, obesity; Cluster 2: advanced age, overweight, vitamin deficiency, obesity, thyroid disease, asthma, uterine disease, other blood disorders, mental disorders, PCOS; Cluster 3: overweight, vitamin deficiency, thyroid disease, other blood disorders, asthma, mental disorders, PCOS, uterine disease, migraine, cardiovascular disease; Cluster 4: vitamin deficiency, other blood disorders, thyroid disease, asthma, mental disorders.

clustering results helped to advance our precision medicine platform for personalized pregnancy care. This study enabled us to identify potential morbidities within the semi-structured EMRs that previously were not being considered in our existing work program, which focused on risk prediction for GDM and improving pregnancy and birth outcomes, therefore improving the number and accuracy of potential predictors. Finally, the study allowed us to consider interacting mechanisms and gateways that may link clustered conditions to adverse outcomes in pregnancy. Finally, this study is part of a more extensive program that builds an LHS to deliver high-value maternal health care and personalized pregnancy care.^{26,57}

4.2 | Strengths and limitations

This study involved an extensive evaluation of morbidities in a large sample size of pregnant women by leveraging EMRs and NLP text mining. However, the study has some limitations. Firstly, measurement bias may arise from inaccuracies or errors during data collection, especially as the data were routinely collected for clinical purposes. Furthermore, the statistical measures reflect the relationship between variables and do not necessarily represent causal associations. Hence, these findings, while helpful in advancing our understanding of morbidities in pregnancy, warrant cautious interpretation.

5 | CONCLUSIONS

Three-quarters of pregnant women experienced at least one medical condition, of which 40% had two or more. In the context of rapidly rising pregnancy interventions, costs, and adverse outcomes, this study emphasizes the importance of considering the entire health profile of patients when predicting and managing risk and the need for public health measures and comprehensive prenatal and pregnancy care. GDM was common during pregnancy, affecting more than 1-in-5, and our results highlight the strong associations between several comorbidities and the incidence of GDM, which will guide prevention efforts. This work was codeveloped with our multidisciplinary clinicians, and is an important step in the pipeline that will ultimately deliver automated information extraction from EMRs within an LHS to drive better care and outcomes.

AUTHOR CONTRIBUTIONS

Y.B., J.E., H.T., L.M., and A.M. were involved in acquiring data, conceptualization, methodology, software, data curation, formal analysis, and writing—the original draft. All the remaining authors were involved in interpreting the data and critically reviewing or revising the manuscript for important intellectual content. All authors approved the final version for publication.

FUNDING INFORMATION

There was no funding for this project. Y.B. is supported by a Monash Graduate Scholarship (MGS) and Monash International Tuition Scholarship (MITS). A.M. and H.T. are funded by National Health and Medical Research Council (NHMRC) Fellowships.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All data and material access requests can be forwarded to the corresponding author's email address.

ETHICS STATEMENT

Ethics approval was obtained from Monash Health with application number of RES-21-0000-183 L. The research has been conducted in adherence with the Code of Ethics of the World Medical Association, also known as the Declaration of Helsinki.

REFERENCES

1. Cha J, Sun X, Dey SK. Mechanisms of implantation: strategies for successful pregnancy. *Nat Med*. 2012;18(12):1754-1767. doi:10.1038/nm.3012
2. Whelan LJ. Comorbid disorders in pregnancy. In: Tintinalli JE, Stapczynski JS, Ma OJ, Yealy DM, Meckler GD, Cline DM, eds. *Tintinalli's Emergency Medicine: A Comprehensive Study Guide*. 8th ed. McGraw-Hill Education; 2016 Available from: <http://accessmedicine.mhmedical.com/content.aspx?aid=1132929954>
3. Reece J, McCauley M, McCaw-Binns A, White SA, Samms-Vaughan M, van den Broek N. Maternal morbidity: a longitudinal study of women's health during and up to 22 months after pregnancy in Jamaica. *Psychol Health Med*. 2020;25(6):687-702. doi:10.1080/13548506.2019.1691243
4. Pati S, Puri P, Sinha R, Panda M, Pati S. Profile of comorbidity and multimorbidity among women attending antenatal clinics: an exploratory cross-sectional study from Odisha, India. *J Family Med Prim Care*. 2022;11(5):1980-1988. doi:10.4103/jfmpc.jfmpc-1855-21
5. Australian Institute of Health and Welfare. Australia's mothers and babies, maternal age. 2023 Accessed at January 10 2024. Available from: <https://www.aihw.gov.au/reports/mothers-babies/australias-mothers-babies/contents/overview-and-demographics/maternal-age>
6. Australian Institute of Health and Welfare. Australia's mothers and babies, maternal body mass index. 2023 Accessed at January 10 2024. Available from: <https://www.aihw.gov.au/reports/mothers-babies/australias-mothers-babies/contents/antenatal-period/maternal-body-mass-index>
7. Barajas-Martínez A, Easton JF, Rivera AL, et al. Metabolic physiological networks: the impact of age. *Front Physiol*. 2020;11:587994. doi:10.3389/fphys.2020.587994
8. El Assar M, Angulo J, Vallejo S, Peiró C, Sánchez-Ferrer CF, Rodríguez-Mañas L. Mechanisms involved in the aging-induced vascular dysfunction. *Front Physiol*. 2012;3:132. doi:10.3389/fphys.2012.00132
9. Chen Y, Zhang XP, Yuan J, et al. Association of body mass index and age with incident diabetes in Chinese adults: a population-based cohort study. *BMJ Open*. 2018;8(9):e021768. doi:10.1136/bmjopen-2018-021768
10. Australian Institute of Health and Welfare. The health of Australia's females, how healthy are Australia's females? 2023 Accessed at January 10 2024. Available from: <https://www.aihw.gov.au/reports/men-women/female-health/contents/how-healthy>
11. Kazma JM, van den Anker J, Allegaert K, Dallmann A, Ahmadzia HK. Anatomical and physiological alterations of pregnancy. *J Pharmacokinet Pharmacodyn*. 2020;47(4):271-285. doi:10.1007/s10928-020-09677-1
12. McIntyre HD, Catalano P, Zhang C, Desoye G, Mathiesen ER, Damm P. Gestational diabetes mellitus. *Nat Rev Dis Primers*. 2019;5(1):47. doi:10.1038/s41572-019-0098-8
13. Rana S, Lemoine E, Granger JP, Karumanchi SA. Preeclampsia: pathophysiology, challenges, and perspectives. *Circ Res*. 2019;124(7):1094-1112. doi:10.1161/CIRCRESAHA.118.313276
14. Wang H, Li N, Chiveste T, et al. IDF diabetes atlas: estimation of global and regional gestational diabetes mellitus prevalence for 2021 by International Association of Diabetes in Pregnancy Study Group's criteria. *Diabetes Res Clin Pract*. 2022;183:109050. doi:10.1016/j.diabres.2021.109050
15. Teede HJ, Bailey C, Moran LJ, et al. Association of antenatal diet and physical activity-based interventions with gestational weight gain and pregnancy outcomes: a systematic review and meta-analysis. *JAMA Intern Med*. 2022;182(2):106-114. doi:10.1001/jamainternmed.2021.6373
16. Carolan MC, Davey MA, Biro M, Kealy M. Very advanced maternal age and morbidity in Victoria, Australia: a population based study. *BMC Pregnancy Childbirth*. 2013;13(1):80. doi:10.1186/1471-2393-13-80
17. Australian Institute of Health and Welfare. Older mothers in Australia 2019, summary. 2021 Accessed at June 12 2023. Available from: <https://www.aihw.gov.au/reports/mothers-babies/older-mothers-in-australia-2019/summary>
18. WHO. Maternal health. 2022 Accessed at June 23 2022. Available from: <https://www.who.int/health-topics/maternal-health>

19. CDC. CDC global health—maternal and child health. 2019 Available from: <https://www.cdc.gov/globalhealth/mch/index.htm>
20. Theall KP, Johnson CC. Environmental influences on maternal and child health. *Int J Environ Res Public Health*. 2017;14(9):1088. doi:10.3390/ijerph14091088
21. Kissoon N, Dugani S, Bhutta ZA. Maternal and child health: gains, but a long journey ahead. *CMAJ Can Med Assoc J*. 2015;187(16):E471-E472. doi:10.1503/cmaj.150725
22. Costello A, Naimy Z. Maternal, newborn, child and adolescent health: challenges for the next decade. *Int Health*. 2019;11(5):349-352. doi:10.1093/inthealth/ihz051
23. Budrionis A, Bellika JG. The learning healthcare system: where are we now? A systematic review. *J Biomed Inform*. 2016;64:87-92. doi:10.1016/j.jbi.2016.09.018
24. Institute of Medicine (US) Roundtable on Evidence-Based Medicine. In: Olsen L, Aisner D, McGinnis JM, eds. *The Learning Healthcare System: Workshop Summary*. National Academies Press. US; 2007. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK53494/>
25. Activity ALHS, Care R on V and SDH, Medicine I of. Roundtable on Value & Science-Driven Health Care. *Observational Studies in a Learning Health System: Workshop Summary*. National Academies Press; 2013 Available from: <https://www.ncbi.nlm.nih.gov/books/NBK201307/>
26. Callander EJ, Teede H, Enticott J. Value in maternal care: using the learning health system to facilitate action. *Birth*. 2022;49(4):589-594. doi:10.1111/birt.12684
27. Dr Wallace P. Learning Healthcare Project. 2015 Available from: <https://learninghealthcareproject.org/dr-paul-wallace-interview/>
28. Learning Healthcare Project. Need for LHS. 2015. Available from: <https://learninghealthcareproject.org/background/need-for-lhs/>
29. Teede HJ, Harrison CL, Teh WT, Paul E, Allan CA. Gestational diabetes: development of an early risk prediction tool to facilitate opportunities for prevention. *Aust N Z J Obstet Gynaecol*. 2011;51(6):499-504. doi:10.1111/j.1479-828X.2011.01356.x
30. Cooray SD, Boyle JA, Soldatos G, et al. Development, validation and clinical utility of a risk prediction model for adverse pregnancy outcomes in women with gestational diabetes: the PerSonal GDM model. *EClinicalmedicine*. 2022;1(52):101637. doi:10.1016/j.eclinm.2022.101637
31. Callander E. Variation in the cost of maternity care: how Australia's financing and delivery of care affects costs to the government and costs to women. *Women Birth*. 2019;1(32):S15-S16. doi:10.1016/j.wombi.2019.07.196
32. RECORD. Reporting Guidelines. 2023 Available from: <https://www.record-statement.org/>
33. Ganguly R, Chakraborty S. Semi-structured patient data in electronic health record. In: Dey N, ed. *Data-Driven Approach for Bio-Medical and Healthcare*. Springer Nature; 2023:219-233. doi:10.1007/978-981-19-5184-8_12
34. NLTK. Natural Language Toolkit. 2023 Available from: <https://www.nltk.org/>
35. Python documentation. re—regular expression operations. 2023. Available from: <https://docs.python.org/3/library/re.html>
36. Enticott J, Johnson A, Teede H. Learning health systems using data to drive healthcare improvement and impact: a systematic review. *BMC Health Serv Res*. 2021;21(1):200. doi:10.1186/s12913-021-06215-8
37. Belsti Y, Moran L, Du L, et al. Comparison of machine learning and conventional logistic regression-based prediction models for gestational diabetes in an ethnically diverse population; the Monash GDM machine learning model. *Int J Med Inform*. 2023;179:105228. doi:10.1016/j.ijmedinf.2023.105228
38. Cooray SD, De Silva K, Enticott JC, et al. Temporal validation and updating of a prediction model for the diagnosis of gestational diabetes mellitus. *J Clin Epidemiol*. 2023;164:54-64. doi:10.1016/j.jclinepi.2023.08.020
39. Bateman BT, Mhyre JM, Hernandez-Diaz S, et al. Development of a comorbidity index for use in obstetric patients. *Obstet Gynecol*. 2013;122(5):957-965. doi:10.1097/AOG.0b013e3182a603bb
40. Kersten I, Lange AE, Haas JP, et al. Chronic diseases in pregnant women: prevalence and birth outcomes based on the SNIp-study. *BMC Pregnancy Childbirth*. 2014;19(14):75. doi:10.1186/1471-2393-14-75
41. Admon LK, Winkelman TN, Heisler M, Dalton VK. Obstetric outcomes and delivery-related health care utilization and costs among pregnant women with multiple chronic conditions. *Prev Chronic Dis*. 2018;8(15):E21. doi:10.5888/pcd15.170397
42. Lee SI, Azcoaga-Lorenzo A, Agrawal U, et al. Epidemiology of pre-existing multimorbidity in pregnant women in the UK in 2018: a population-based cross-sectional study. *BMC Pregnancy Childbirth*. 2022;22(1):120. doi:10.1186/s12884-022-04442-3
43. Nakanishi K, Saijo Y, Yoshioka E, et al. Association between maternal multimorbidity and preterm birth, low birth weight and small for gestational age: a prospective birth cohort study from the Japan Environment and Children's study. *BMJ Open*. 2023;13(3):e069281. doi:10.1136/bmjopen-2022-069281
44. Callander EJ, Fenwick J, Donnellan-Fernandez R, et al. Cost of maternity care to public hospitals: a first 1000-days perspective from Queensland. *Aust Health Rev*. 2019;43(5):556-564. doi:10.1071/AH18209
45. Lloyd M, Teede H, Bailey C, Callander E, Ademi Z. Projected return on investment from implementation of a lifestyle intervention to reduce adverse pregnancy outcomes. *JAMA Netw Open*. 2022;5(9):e2230683. doi:10.1001/jamanetworkopen.2022.30683
46. Margerison-Zilko CE, Strutz KL, Li Y, Holzman C. Stressors across the life-course and preterm delivery: evidence from a pregnancy cohort. *Matern Child Health J*. 2017;21(3):648-658. doi:10.1007/s10995-016-2151-5
47. Roberts CL, Ford JB, Algert CS, et al. Population-based trends in pregnancy hypertension and pre-eclampsia: an international comparative study. *BMJ Open*. 2011;1(1):e000101. doi:10.1136/bmjopen-2011-000101
48. ACOG. Pregnancy at age 35 years or older. 2023 Available from: <https://www.acog.org/en/clinical/clinical-guidance/obstetric-care-consensus/articles/2022/08/pregnancy-at-age-35-years-or-older>
49. Bolodurina I, Shukhman A, Legashev L, Grishina L, Zhigalov A. Extracting and processing of Russian unstructured clinical tests for a medical decision support system. *Eng Proc*. 2023;33(1):41. doi:10.3390/engproc2023033041
50. Beaney T, Jha S, Alaa A, et al. Comparing natural language processing representations of coded disease sequences for prediction in electronic health records. *J Am Med Inform Assoc*. 2024;31(7):1451-1462. doi:10.1093/jamia/ocae091
51. Niu H, Omiaoum OA, Langston MA, et al. EHR-BERT: a BERT-based model for effective anomaly detection in electronic health records. *J Biomed Inform*. 2024;1(150):104605. doi:10.1016/j.jbi.2024.104605
52. Lee S, Lee J, Park J, et al. Deep learning-based natural language processing for detecting medical symptoms and histories in emergency patient triage. *Am J Emerg Med*. 2024;1(77):29-38. doi:10.1016/j.ajem.2023.11.063
53. Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: a review. *Trends Anaesth Crit Care*. 2021;1(38):4-9. doi:10.1016/j.tacc.2021.02.007
54. Kappel C, Rushton M, Johnson C, et al. Clinical experience of patients referred to a multidisciplinary cardio-oncology clinic: an observational

- cohort study. *Curr Oncol*. 2019;26(3):e322-e327. doi:[10.3747/co.26.4509](https://doi.org/10.3747/co.26.4509)
55. Freund T, Kunz CU, Ose D, Szecsenyi J, Peters-Klimm F. Patterns of multimorbidity in primary care patients at high risk of future hospitalization. *Popul Health Manag*. 2012;15(2):119-124. doi:[10.1089/pop.2011.0026](https://doi.org/10.1089/pop.2011.0026)
56. Cornell JE, Pugh JA, John W, et al. Multimorbidity clusters: clustering binary data from multimorbidity clusters: clustering binary data from a large administrative medical database. *Appl Multivar res*. 2008;12(3):163-182.
57. Enticott JC, Melder A, Johnson A, et al. A learning health system framework to operationalize health data to improve quality care: an Australian perspective. *Front Med*. 2021;27(8):730021. doi:[10.3389/fmed.2021.730021](https://doi.org/10.3389/fmed.2021.730021)

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Belsti Y, Moran L, Mousa A, et al.

Analyzing electronic medical records to extract prepregnancy morbidities and pregnancy complications: Toward a learning health system. *Learn Health Sys*. 2025;9(2):e10473. doi:[10.1002/lrh2.10473](https://doi.org/10.1002/lrh2.10473)