5-2020

Validity and Reliability of the Antepartum Gastrointestinal Symptom Assessment Instrument

Lisa Hanson
Marquette University, lisa.hanson@marquette.edu

Leona VandeVusse
Marquette University, leona.vandevusse@marquette.edu

Mauricio Garnier-Villarreal
Marquette University, mauricio.garniervillarreal@marquette.edu

Donna O. McCarthy
Marquette University, donnalee.mccarthy@marquette.edu

Teresa Jerofke-Owen
Marquette University, teresa.jerofke@marquette.edu

See next page for additional authors

Authors

Lisa Hanson, Leona VandeVusse, Mauricio Garnier-Villarreal, Donna O. McCarthy, Teresa Jerofke-Owen, Emily Malloy, and Heidi Paquette

Validity and Reliability of the Antepartum Gastrointestinal Symptom Assessment Instrument

Lisa Hanson
Marquette University, College of Nursing, Milwaukee, WI
Leona VandeVusse
Marquette University, College of Nursing, Milwaukee, WI
Mauricio Garnier-Villarreal
Marquette University, College of Nursing, Milwaukee, WI
Donna McCarthy
Marquette University, College of Nursing, Milwaukee, WI
Teresa Jerofke-Owen
Marquette University, College of Nursing, Milwaukee, WI
Emily Malloy
Marquette University, College of Nursing, Milwaukee, WI
Aurora Sinai Midwifery and Wellness Center, Milwaukee, WI
Heidi Paquette

Marquette University, College of Nursing, Milwaukee, WI

Abstract

Objective

To examine the psychometric properties of the nine-item Antepartum Gastrointestinal Symptom Assessment (AP-GI-SA) instrument.

Design

Single-group prospective design.

Setting

Urban prenatal clinic serving a diverse population.

Participants

Convenience sample of 45 pregnant women.

Methods

Participants completed the AP-GI-SA before a scheduled prenatal care appointment. We used Bayesian structural equation modeling to evaluate the construct validity of the scale and assessed known-groups validity. We assessed reliability through maximal reliability coefficient estimate and measured internal consistency with Cronbach's alpha coefficient.

Results

Participants completed the instrument in 2 minutes or less. Construct validity was supported by confirmatory factor analysis (posterior predictive p value = 0.49, gamma-hat = 0.970, and root mean square error of approximation = 0.065), which indicated that the single-factor model is a plausible data-generative model for GI symptoms. The maximal reliability coefficient of 0.75 and Cronbach's alpha coefficient of 0.67 supported reliability. Average AP-GI-SA scores were the highest for women in the third trimester. Of all nine GI symptoms, heartburn in the third trimester received the highest score.

Conclusion

Our findings provide preliminary support for the validity and reliability of the AP-GI-SA. The instrument may be used as a measure in intervention studies where GI symptoms of pregnancy are an outcome. The AP-GI-SA could also be useful in clinical settings to quickly evaluate GI symptoms.

Keywords

Symptom assessment and management, referred to as *symptom science*, is an important area of nursing science and clinical practice. In symptom science, gastrointestinal (GI) discomforts in pregnancy are understudied. Pregnancy-related GI discomforts result from the hormonal and physical changes that lead to the significant anatomic and physiologic changes necessary to support a growing fetus. Several of the symptoms commonly reported by pregnant women are characterized as GI. For example, human chorionic gonadotropin peaks between 9 and 12 weeks gestation and causes the most intense symptoms of nausea and vomiting (Gadsby et al., 2019, Jarvis and Nelson-Piercy, 2011, Patil et al., 2012). Nausea and vomiting occur in 70% to 80% of pregnant women (Bustos, Venkataramanan, & Caritis, 2017). Human chorionic gonadotropin levels decline in the second trimester, and nausea and vomiting cease for most women (American College of Obstetricians and Gynecologists, 2018, Lee and Saha, 2011).

As the size of the uterus increases, the intestines are displaced, which causes flatulence (Reedy, Bowers, & King, 2019). In the second and third trimesters, the predominance of progesterone secretion from the placenta slows GI peristalsis (de Milliano, Tabbers, van der Post, & Benninga, 2012) and causes constipation in approximately half of all pregnant women; this can lead to symptoms of bloating (Zielinski, Searing, & Deibel, 2015). Although diarrhea is a less common GI symptom than constipation, it may be physiologic near the time of labor (Zielinski et al., 2015). Progesterone is also responsible for the relaxation of the pyloric sphincter, which contributes to the heartburn that frequently occurs in the second and third trimesters (Ramu, Mohan, Rajasekaran, & Jayanthi, 2011). Heartburn is a symptom of chronic gastroesophageal reflux disease (GERD), which occurs in 30% to 50% of pregnant women (Zielinski et al., 2015). Women previously diagnosed with GERD may report an increase in symptom severity during pregnancy (Gerson, 2012), and symptoms may include an acid taste in the mouth, difficulty swallowing, regurgitation, a burning sensation, noncardiac chest pain (Zielinski et al., 2015), and belching (Kessing, Bredenoord, & Smout, 2014).

GI symptoms of pregnancy are time limited and generally do not produce long-term, disabling problems. However, they may have a negative effect on quality of life and daily activities. Women's experiences of the common GI discomforts of pregnancy and nonpharmacologic interventions to modulate these symptoms have rarely been explored scientifically. A reliable and valid scale to measure GI symptoms in pregnant women is needed for research in symptom management and in clinical practice to assess outcomes of interventions for these common symptoms.

> A tool for comprehensive assessment of gastrointestinal symptoms of pregnancy is needed for clinicians and researchers.

Review of the Literature

There is significant interest in the standardization of symptom measurement in clinical research. The Patient-Reported Outcome Measurement Information System (PROMIS) contains established, person-centered measures that are used to monitor a variety of health and illness symptoms (HealthMeasures, 2020). However, the PROMIS database includes no comprehensive tools with which to measure GI symptoms in pregnancy.

Patient self-report is an efficient way to assess symptoms in clinical settings and for research applications (Lackner et al., 2014). Prospective patient symptom diaries for persons with GI disease have been viewed as superior to retrospective recall questionnaires (Jones et al., 2019), but this approach has not been applied to pregnant women. Daily diaries add considerable burden to research participants, and they are often incomplete, may contain fabricated information, and may be an unreliable representation of symptoms (Lackner et al., 2014).

Mujagic et al. (2015) systematically reviewed 110 publications in which symptom assessment tools were used to assess abdominal pain in persons with irritable bowel syndrome. Among the studies reviewed, 18 different questionnaires were used, although psychometric data were reported in only 11. The measures used to assess the reliability and validity of instruments tested in these studies varied significantly. The authors concluded that the wide variety of instruments used made the outcomes difficult to compare among published studies.

Bolier, Kessing, Smout, and Bredenoord (2015) conducted a systematic review of instruments used to measure four well-defined dimensions of GERD: symptoms, treatment response, diagnosis, and quality of life. The authors extracted and analyzed 65 questionnaires and concluded that no single questionnaire assessed all four dimensions. Malfertheiner, Malfertheiner, Kropf, Costa, and Malfertheiner (2012) conducted a prospective cohort study to determine the frequency of GERD symptoms during pregnancy. Their sample included 840 participants: 510 pregnant women and 330 nonpregnant women (control group). The researchers used an established tool (the Reflux Disease Questionnaire) that contained 12 questions: six questions related to symptom frequency during the previous week and six related to symptom severity. Each item was rated on a six-point scale from *none* to *very severe*. Although less than 10% of the control group experienced GERD symptoms, Malfertheiner et al. (2012) noted a steady increase in the frequency and severity of GERD symptoms during pregnancy, from 26% of women in the first trimester to 51% in the third trimester. They concluded that among pregnant women, GERD is significantly undertreated and that symptom relief is inadequate. Malfertheiner et al. showed that GI symptom assessment during pregnancy is feasible and has potential for clinical and research purposes.

A tool for comprehensive assessment of GI symptoms during pregnancy is needed for clinicians and researchers. We identified one instrument with a subscale that had potential to be adapted for use during pregnancy. The Severity of Dyspepsia Assessment (SODA) Non-Pain Symptom subscale is a questionnaire developed by Rabeneck et al. (2001) to measure the severity of dyspepsia in persons with GI problems, such as peptic ulcers. The original SODA Non-Pain Symptom subscale included seven items—burping, heartburn, bloating, passing gas, sour taste, nausea, and bad breath—that were rated on a 5-point Likert-type scale in order of increasing severity. Although Rabeneck et al. (2001) focused on measuring dyspepsia from a disease point of view, this tool has the potential to be useful in the evaluation of GI symptoms in pregnant women. Rabeneck et al. (2001) tested the SODA Non-Pain Symptom subscale with 100 participants who had diagnoses of chronic GI problems. The subscale had a Cronbach's alpha coefficient estimate of internal consistency reliability of .90 in this initial study. However, only 6% of the study sample were women, and none were pregnant.

We adapted the SODA Non-Pain Symptom subscale by adding two extra items, diarrhea and constipation. The result is a nine-item modified instrument that addresses the symptoms frequently reported by healthy pregnant women. The Likert-type scale used in the original SODA Non-Pain Symptom subscale was maintained in the new instrument, named the Antepartum Gastrointestinal Symptom Assessment (AP-GI-SA; see Figure 1). Participants rated each symptom from 1 (*no problem*) to 5 (*very severe problem, markedly influences daily activities and/or requires rest*; Rabeneck et al., 2001, p. 763). The purpose of this preliminary study was to examine the psychometric properties of the nine-item AP-GI-SA instrument.

	Rating Scale				
	No Problem	Mild problem, can be ignored	Moderate problem, cannot be ignored but does not	Severe problem, influences concentration on daily activities	Very severe problem, markedly influences daily

			influence daily activities		activities and/or requires rest
Burping/belching	1	2	3	4	5
Heartburn	1	2	3	4	5
Bloating	1	2	3	4	5
Passing gas	1	2	3	4	5
Sour taste	1	2	3	4	5
Nausea	1	2	3	4	5
Bad breath	1	2	3	4	5
Diarrhea	1	2	3	4	5
Constipation	1	2	3	4	5

Figure 1. Antepartum Gastrointestinal Symptom Assessment. Adapted from "SODA (severity of dyspepsia assessment): A new effective outcome measure for dyspepsia-related health," by L. Rabeneck, K. G. Cook, K. Wristers, J. Souchek, T. Menke, & N. P. Wray, 2001, *Journal of Clinical Epidemiology*, *54*, p. 763, using the Non-Pain Symptom Scale from the SODA tool. Copyright 2001 by Elsevier. Used with permission from Elsevier.

Methods

Design

We used a prospective, single-group design. Participants completed the AP-GI-SA before routine prenatal visits. The study was approved by the Marquette University institutional review board and the clinic administrator.

Setting and Sample

The study setting was a privately-owned urban prenatal clinic serving a diverse population. For this preliminary study of the AP-GI-SA, the convenience sample consisted of 45 adult pregnant women who spoke and read English and self-identified as not having GI, medical, or pregnancy complications.

Procedures

The first two authors trained research assistants in the study procedures. The research assistants worked with the clinic receptionist to identify pregnant women who were approached in the waiting room of the clinic and asked if they would be willing to participate in the study. Written informed consent was not required by the institutional review board for this study; completion of the instrument served as consent. Women who verbally consented independently completed the AP-GI-SA and a short demographic questionnaire. After returning the completed questionnaires to the research assistants, each participant received a $10 thank you gift.

Measures

The participants were asked to rate their experience with each of the nine symptom items and the severity of each symptom during the past week. Demographic information, including age, expected date of birth, race, and ethnicity, was also collected by questionnaire.

One aspect of construct validity is known-groups comparisons (Rodrigues, Adachi, Beattie, Lau, & MacDermid, 2019). In pregnancy, we would clinically expect women's experience of the GI symptoms

of pregnancy to increase or decrease based on the trimester of pregnancy. We hypothesized that (a) participants in the first trimester would report more nausea, (b) participants in the second and third trimesters would report more heartburn than those in the first trimester, and (c) participants in the third trimester would report more constipation than those in the first or second trimesters. We did not make hypotheses to predict the occurrence of the remaining six GI symptoms.

Our findings provide preliminary support for the validity and reliability of the nine-item Antepartum Gastrointestinal Symptom Assessment.

Analysis

Using only study numbers as identifiers, we entered data from questionnaires into an Excel spreadsheet and imported it into R (Version 3.3.3; R Core Team, 2017). The construct validity of the AP-GI-SA was assessed by evaluating the factorial structure of the scale items to measure the construct of GI symptomatology with Bayesian confirmatory factor analysis (Kline, 2016). Bayesian inference is recommended for small sample sizes because, unlike frequentist statistical inference (null hypothesis significance testing), it does not assume that the data are representative of a larger population (Smid, McNeish, Miočević, & van de Schoot, 2020). The fit of the model was determined through global fit measures, such as the posterior predictive p value, gamma-hat, and root mean square error of approximation, to evaluate if the factor structure is an accurate representation of the relationships among variables. The Bayesian general linear model multiple-group mean comparison was used for the known-group comparisons (Bürkner, 2017, Merkle and Rosseel, 2018).

The reliability of the AP-GI-SA was evaluated with the maximal reliability coefficient using Bayesian Structural Equation Modeling. The maximal reliability coefficient estimates the reliability of a scale by calculating different weights for the items in the scale and, therefore, provides the maximum possible reliability for a linear combination of the scale items (Raykov, 2012). The Cronbach's alpha coefficient was used to evaluate the internal consistency based on interitem correlation.

Results

The participants (N = 45) had a mean age of 25 years (standard deviation [SD] = 5.42), were primarily Black (51.1%), and were primarily in the third trimester of their pregnancies (n = 24; see Table 1). Participants completed the instrument in 2 minutes or less, and none had any questions or requested clarification of any of the items on the AP-GI-SA. One participant was excluded from analysis because she answered only two of the nine items; therefore, data from 44 AP-GI-SA completed instruments were analyzed. Two of the surveys contained partial missing data (between 1 and 4 values each). These missing data were handled with data augmentation through the estimation process, which results in high parameter recoverability as full information maximum likelihood (Enders, 2010).

Table 1. Demographic Characteristics of Participants (N = 44)

Variable	Results
Age in years	M = 25, SD = 5.42 (range = 19–29)
Gestational age in weeks	M = 27.3, SD = 9.91 (range = 6.1–40.7)
First trimester (up to 12 weeks), n (%)	6 (13.3)
Second trimester (13–27 weeks), n (%)	14 (31.1)

Third trimester (28–40+ weeks), *n* (%)	24 (84.4)
Race/ethnicity, *n* (%)	
American Indian or Alaskan Native	0
Black	23 (51.1)
Hispanic	12 (26.7)
Pacific Islander or Asian	4 (8.9)
White	4 (8.9)
Biracial	1 (2.2)

Note. There were 45 participants in the original sample. One participant did not complete the instrument. *M* = mean; *SD* = standard deviation.

Known-Group Comparisons

The participants' mean severity scores for the nine GI symptom items contained in the AP-GI-SA are shown in Table 2. These results are presented graphically in Figure 2, in which the bold horizontal line in each box indicates the average AP-GI-SA score for participants by trimester. An analysis of the correlation between AP-GI-SA and weeks of gestation resulted in an $r = 0.307$ (95% credible interval [CI] [–0.016, 0.629]), which shows a moderate linear relation between GI symptoms and gestational age. The average AP-GI-SA scores for each of the nine GI symptoms by trimester are shown in Figure 3. With the exception of burping and sour taste, scores on the other seven GI symptoms measured by the AP-GI-SA increased slightly as gestational age increased.

Table 2. Mean Severity Scores on the Nine Items of the Antepartum Gastrointestinal Symptom Assessment (*N* = 44)

Symptom	Mean Score (*SD*)
Burping/belching	1.36 (0.57)
Heartburn	2.00 (1.03)
Bloating	1.95 (1.01)
Passing gas	1.78 (0.88)
Sour taste	1.31 (0.68)
Nausea	2.23 (1.31)
Bad breath	1.26 (0.58)
Diarrhea	1.45 (0.79)
Constipation	1.84 (1.14)
Total average	1.69 (0.48)

Note. The Likert-type scale used for self-report of severity of each symptom during the past week ranged from 1 to 5, in order of increasing severity, from *no problem* to *very severe*. Total average score is calculated as the average from all the scale items. *SD* = standard deviation.
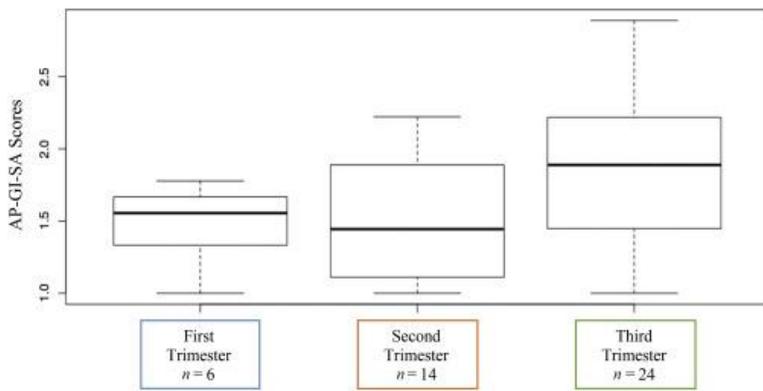
Figure 2. Antepartum Gastrointestinal Symptom Assessment (AP-GI-SA) total scores by trimester.
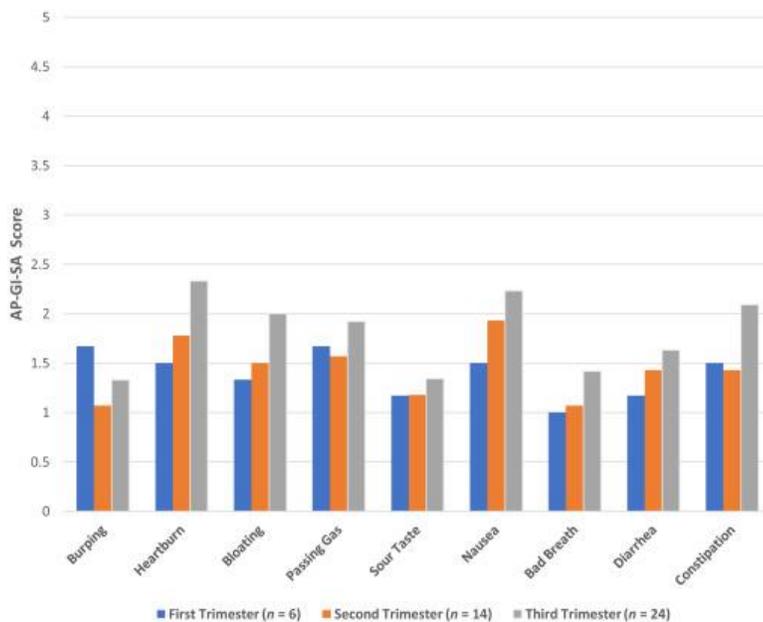

Figure 3. Average Antepartum Gastrointestinal Symptom Assessment (AP-GI-SA) scores for each of the nine gastrointestinal symptoms by trimester.

We analyzed data in relation to the a priori hypotheses for known-group comparisons of symptom prevalence by trimester. Because the number of participants who were in the first trimester was small (n = 6), we excluded the first hypothesis about nausea occurring more frequently in early pregnancy. Heartburn symptoms had a greater mean for the third trimester (mean [M] = 2.21, SD = 1.10) compared with the second trimester (M = 1.79, SD = 0.97). This difference, however, was very small (M_D = 0.42, SE = 0.35, 95% CI [−0.27, 1.11]). Constipation was also more frequent in the third trimester (M = 2.08, SD = 1.32) than the second (M = 1.50, SD = 0.76), and this difference was also small (M_D = 0.58, SE = 0.40, 95% CI [−0.22, 1.40]). A similar pattern was found with nausea. Participants in the third trimester (M = 2.50, SD = 1.25) had higher scores than those in the second (M = 2.14, SD = 1.51), with a small mean difference (M_D = 0.36, SE = 0.43, 95% CI [−0.50, 1.19]).

Confirmatory Factor Analysis

When testing the factorial structure of a single factor defining GI symptoms, we found a posterior predictive p value (ppp) of .495, indicating good overall fit, because the perfect fit is measured as ppp =

.5. The fit indices indicated good overall fit with greater gamma-hat (M = 0.970, 90% CI [0.926, 1.000]) and low root mean square error of approximation (M = 0.065, 90% CI [0.000, 0.131]). These indices suggest that the one-factor structure is a good representation of the underlying factor that predicts the responses for the nine symptom items (Garnier-Villarreal & Jorgensen, 2019). The standardized factor loadings (see Table 3) ranged from 0.229 to 0.736; passing gas had the lowest factor loading, and constipation had the largest. Based on the 95% CIs, we found that for three items (heartburn, passing gas, and sour taste) the CI includes 0, indicating that we are less than 95% certain of them having positive factor loadings. These three items also had the lowest factor loadings. We found that the GI symptoms factor explained 23% of the variance, on average, in the indicators, with explained variance (R^2) ranges from 5% to 54%, showing that one factor on GI symptoms is s good representation of the nine scale items.

Table 3. Factor Loadings for the Confirmatory Factor Analysis (N = 44)

Item	Unstandardized Factor Loading (SD)	95% Credible Interval	Standardized Factor Loading	R^2
Burping/belching	0.243 (0.119)	[0.016, 0.489]	0.401	0.161
Heartburn	0.409 (0.229)	[−0.042, 0.860]	0.375	0.141
Bloating	0.461 (0.217)	[0.027, 0.887]	0.434	0.188
Passing gas	0.210 (0.194)	[−0.165, 0.587]	0.229	0.052
Sour taste	0.252 (0.153)	[-0.039, 0.550]	0.351	0.123
Nausea	0.580 (0.273)	[0.035, 1.113]	0.421	0.177
Bad breath	0.435 (0.119)	[0.212, 0.681]	0.687	0.472
Diarrhea	0.372 (0.171)	[0.050, 0.726]	0.444	0.198
Constipation	0.916 (0.228)	[0.498, 1.377]	0.736	0.542

Note. SD = standard deviation.

Reliability

The reliability of AP-GI-SA scale was evaluated to determine how accurately it measured the GI symptoms as a function of the nine items. The Cronbach's alpha internal consistency estimate for the nine-item instrument was 0.67. This coefficient indicates that, on average, 75% of the information across the nine items defined the overall factor of GI symptoms.

Discussion

Our findings provide preliminary support for the validity and reliability of the AP-GI-SA to evaluate GI symptoms experienced by pregnant women. All participants completed the AP-GI-SA without assistance or questions, which supports face validity of the instrument. The factor analysis indicated that all the scale items fit well to define the latent factor of symptomatology. The passing gas item had the lowest standardized factor loading, indicating that its association with the latent factor is lower than the other items. Smaller factor loadings are more likely a result of a small sample size rather than inadequacy of the items.

The responses to the nine GI symptom items generally reflected the expected pattern of symptoms during pregnancy, although this should be interpreted with caution because the number of participants per trimester was very small. Known-groups validity testing supported two of the three a priori

hypotheses. Although it has been established that 50% to 90% of pregnant women experience some degree of nausea during pregnancy (Jarvis & Nelson-Piercy, 2011), fewer of the first trimester participants (n = 6) reported nausea compared with those in the second (n = 14) and third (n = 24) trimesters. Heartburn occurs in 30% to 50% of women during pregnancy and may reach a high of 80% in some populations (Ramu et al., 2011). Consistent with the literature, heartburn was reported more often by participants in the second and third trimesters. Constipation was also reported more frequently in the third trimester.

The two GI symptoms that participants reported as problems that affected their daily activities, heartburn and nausea, are very common discomforts of pregnancy (Clark et al., 2013, Heitmann et al., 2015). Overall, the scores for each of the nine GI symptoms items were less than 2.5 (range = 1–2.33, median = 1.55), indicating that the symptoms had minimal effects on the pregnant participants' daily activities. This may be attributable to use of a convenience sample. It is possible that women who felt unwell due to GI symptoms (or other pregnancy discomforts) declined to participate. Some of the findings were unexpected, partly because asking pregnant women about the nuances of GI symptoms (sour taste, burping, passing gas) is not routine in clinical practice. However, these distinctions may be useful in research applications of the tool.

A strength of this analysis for the study was the use of Bayesian statistical analysis and the maximal reliability coefficient. The Cronbach's alpha estimate (0.67) indicates that the nine items in the AP-GI-SA are interrelated, without being redundant of one another. They each present unique information related to the GI symptoms. Specific cutoffs for Cronbach's alpha scores remain controversial because their interpretation is more fluid than a specific criterion (Cho & Kim, 2015). Bayesian Confirmatory Factor Analysis allowed us to more accurately measure reliability and validity of the AP-GI-SA instrument administered to this small sample (Barbaranelli et al., 2015, Raykov, 2012). The maximal reliability coefficient indicates moderately high reliability of the AP-GI-SA instrument (maximal reliability = 0.75). The findings of this study provide preliminary evidence of the validity and reliability of the AP-GI-SA as a measure of GI symptoms experienced by pregnant women.

> The Antepartum Gastrointestinal Symptom Assessment may be useful in clinical settings to quickly evaluate gastrointestinal symptoms.

Limitations

Our study had several limitations. The number of women who were approached to participate in the study but who ultimately declined to be part was not recorded. Therefore, the study participation rate cannot be reported. Marital and employment characteristics were not collected from participants. Because women self-selected participation, it is plausible that more or fewer of them had GI symptoms than nonparticipants. The small sample size limited the number of participants who were in each pregnancy trimester. Response bias is a potential limitation of a self-administered symptom tool. Although there was no way to directly connect the AP-GI-SA survey responses to a specific participant after she completed it, there is still a chance that women may have felt embarrassed to admit that they were having ongoing issues with potentially personal GI issues, such as passing gas or bad breath. Vomiting was not included in the revised AP-GI-SA but has since been added to form a 10-item instrument.

Implications

Potential research applications are broad for the AP-GI-SA. It can be applied as a measure in intervention studies where GI symptoms of pregnancy are an outcome. The revised 10-item AP-GI-SA is in use in a double-blind, randomized, placebo-controlled trial to determine the efficacy of probiotics to reduce antenatal Group B *Streptococcus* colonization (ClinicalTrials.gov identifier: NCT03696953). The AP-GI-SA instrument was placed on an iPad (Apple, Cupertino, CA) and is administered before initiation of the probiotic or placebo and repeated again 12 weeks later. This facilitates direct data entry into a secure data collection system used for the study and allows participants privacy in completion of the AP-GI-SA.

Future research with a larger, more diverse sample of pregnant women would offer greater statistical power to reexamine the factor analysis. The AP-GI-SA could be examined for reliability and validity in samples of pregnant women who have documented acute or chronic GI conditions for use in determining efficacy of treatments. The relationship between gestational age and the reporting of symptoms could be assessed more completely in future research applications of the AP-GI-SA.

The AP-GI-SA may be used in clinical settings to quickly evaluate GI symptoms.

Used as a previsit assessment, the AP-GI-SA could help a provider understand which symptoms the woman is experiencing, including the severity of her symptoms. Scores could be used to focus clinical therapies. Data collected from the tool could be used in subsequent prenatal visits to evaluate treatment approaches used to modulate these symptoms and improve quality of life.

Conclusion

Our preliminary findings show that the AP-GI-SA is an efficient and easy-to-use tool to measure GI symptoms in healthy pregnant women. Our findings also provide preliminary evidence for the validity and reliability of the AP-GI-SA. Therefore, the AP-GI-SA will be a useful tool for future research and clinical practice to assess interventions aimed at modifying GI symptoms in healthy pregnant women. Dissemination of these findings will allow other researchers to consider evaluation of this revised tool for use in other healthy populations.

References

American College of Obstetricians and Gynecologists, 2018. American College of Obstetricians and Gynecologists. **ACOG practice bulletin no. 189: Nausea and vomiting of pregnancy.** *Obstetrics & Gynecology*, 131 (1) (2018), pp. e15-e30, 10.1097/AOG.0000000000002456

Barbaranelli et al., 2015. C. Barbaranelli, C.S. Lee, E. Vellone, B. Riegel. **The problem with Cronbach's alpha: Comment on Sijtsma and van der Ark (2015).** *Nursing Research*, 64 (2) (2015), pp. 140-145, 10.1097/NNR.0000000000000079

Bolier et al., 2015. E.A. Bolier, B.F. Kessing, A.J. Smout, A.J. Bredenoord. **Systematic review: Questionnaires for assessment of gastroesophageal reflux disease.** *Diseases of the Esophagus*, 28 (2) (2015), pp. 105-120, 10.1111/dote.12163

Bürkner, 2017. P. Bürkner. **brms: An R package for Bayesian multilevel models using stan.** *Journal of Statistical Software*, 80 (1) (2017), pp. 1-28, 10.18637/jss.v080.i01

Bustos et al., 2017. M. Bustos, R. Venkataramanan, S. Caritis. **Nausea and vomiting of pregnancy—What's new?** *Autonomic Neuroscience: Basic and Clinical*, 202 (2017), pp. 62-72, 10.1016/j.autneu.2016.05.002

Cho and Kim, 2015. E. Cho, S. Kim. **Cronbach's coefficient alpha: Well known but poorly understood.** *Organizational Research Methods*, 18 (2) (2015), pp. 207-230, 10.1177/1094428114555994

Clark et al., 2013. S. Clark, B. Hughes, S.S. McDonald. **The impact of nausea and vomiting of pregnancy on quality of life: Report of a national consumer survey and recommendations for improving care.** *Obstetrical & Gynecological Survey*, 68 (9) (2013), pp. S1-S10, 10.1097/OGX.0b013e3182a8784d

de Milliano et al., 2012. I. de Milliano, M.M. Tabbers, J.A. van der Post, M.A. Benninga. **Is a multispecies probiotic mixture effective in constipation during pregnancy? 'A pilot study'.** *Nutrition Journal*, 11 (2012), p. 80, 10.1186/1475-2891-11-80

Enders, 2010. C.K. Enders. **Applied missing data analysis.** Guilford Press, New York, NY (2010)

Gadsby et al., 2019. R. Gadsby, V. Rawson, E. Dziadulewicz, B. Rousseau, H. Collings. **Nausea and vomiting of pregnancy and resource implications: The NVP Impact Study.** *British Journal of General Practice*, 69 (680) (2019), pp. e217-e223, 10.3399/bjgp18X700745

Garnier-Villarreal and Jorgensen, 2019. M. Garnier-Villarreal, T.D. Jorgensen. **Adapting fit indices for Bayesian structural equation modeling: Comparison to maximum likelihood.** *Psychological Methods*, 25 (1) (2019), pp. 46-70, 10.1037/met0000224

Gerson, 2012. L.B. Gerson. **Treatment of gastroesophageal reflux disease during pregnancy.** *Gastroenterology & Hepatology*, 8 (11) (2012), pp. 763-764

HealthMeasures, 2020. HealthMeasures. **PROMIS®.** Northwestern University, Evanston, IL (2020). Retrieved from http://www.healthmeasures.net/explore-measurement-systems/promis

Heitmann et al., 2015. K. Heitmann, L. Holst, A. Lupattelli, C. Maltepe, H. Nordeng. **Treatment of nausea in pregnancy: A cross-sectional multinational web-based study of pregnant women and new mothers.** *BMC Pregnancy and Childbirth*, 15 (321) (2015), pp. 3-13, 10.1186/s12884-015-0746-2

Jarvis and Nelson-Piercy, 2011. S. Jarvis, C. Nelson-Piercy. **Management of nausea and vomiting in pregnancy.** *British Medical Journal*, 342 (2011), pp. 1-8, 10.1136/bmj.d3606

Jones et al., 2019.
M.P. Jones, S. Walter, A. Faresjo, E. Grodzinsky, L. Kjellstrom, L. Viktorsson, ..., A. Andreasson. **Gastrointestinal recall questionnaires compare poorly with prospective patient diaries for gastrointestinal symptoms: Data from population and primary health centre samples.** *European Journal of Gastroenterology & Hepatology*, 31 (2) (2019), pp. 163-169, 10.1097/MEG.0000000000001296

Kessing et al., 2014. B.F. Kessing, A.J. Bredenoord, A.J. Smout. **The pathophysiology, diagnosis and treatment of excessive belching symptoms.** *American Journal of Gastroenterology*, 109 (8) (2014), pp. 1196-1203, 10.1038/ajg.2014.165

Kline, 20160. .B. Kline. **Principles and practice of structural equation modeling.** (4th ed.), Guilford Press, New York, NY (2016)

Lackner et al., 2014. J.M. Lackner, J. Jaccard, L. Keefer, R. Firth, A.M. Carosella, M. Sitrin, . . . IBSOS Research Group. **The accuracy of patient reported measures for GI symptoms: A comparison**

of real time and retrospective reports. *Neurogastroenterology & Motility*, 26 (12) (2014), pp. 1802-1811, 10.1111/nmo.12466

Lee and Saha, 2011. N.M. Lee, S. Saha. **Nausea and vomiting of pregnancy.** *Gastroenterology Clinics of North America*, 40 (2) (2011), pp. 309-334, 10.1016/j.gtc.2011.03.009

Malfertheiner et al., 2012. S.F. Malfertheiner, M.V. Malfertheiner, S. Kropf, S.-D. Costa, P. Malfertheiner. **A prospective longitudinal cohort study: Evolution of GERD symptoms during the course of pregnancy.** *BMC Gastroenterology*, 12 (2012), p. 131, 10.1186/1471-230X-12-131

Merkle and Rosseel, 2018. E.C. Merkle, Y. Rosseel. **blavaan: Bayesian Structural Equation Models via parameter expansion.** *Journal of Statistical Software*, 85 (4) (2018), pp. 1-30, 10.18637/jss.v085.i04

Mujagic et al., 2015. Z. Mujagic, D. Keszthelyi, Q. Aziz, W. Reinisch, E.G. Quetglas, F. De Leonardis, ..., A.A. Masclee. **Systematic review: Instruments to assess abdominal pain in irritable bowel syndrome.** *Alimentary Pharmacology & Therapeutics*, 42 (9) (2015), pp. 1064-1081, 10.1111/apt.13378

Patil et al., 2012. C.L. Patil, E.T. Abrams, A.R. Steinmetz, S.L. Young. **Appetite sensations and nausea and vomiting in pregnancy: An overview of the explanations.** *Ecology of Food and Nutrition*, 51 (2012), pp. 394-417, 10.1080/03670244.2012.696010

R Core Team, 2017. R Core Team. **R: A language and environment for statistical computing.** *R Foundation for Statistical Computing*, Vienna, Austria (2017). Retrieved from https://www.R-project.org

Rabeneck et al., 2001. L. Rabeneck, K.F. Cook, K. Wristers, J. Souchek, T. Menke, N.P. Wray. **SODA (severity of dyspepsia assessment): A new effective outcome measure for dyspepsia-related health.** *Journal of Clinical Epidemiology*, 54 (2001), pp. 755-765, 10.1016/S0895-4356(00)00365-6

Ramu et al., 2011. B. Ramu, P. Mohan, M.S. Rajasekaran, V. Jayanthi. **Prevalence and risk factors for gastroesophageal reflux in pregnancy.** *Indian Journal of Gastroenterology*, 30 (3) (2011), pp. 144-147, 10.1007/s12664-010-0067-3

Raykov, 2012. T. Raykov. **Scale construction and development using structural equation modeling.** R.H. Hoyle (Ed.), *Handbook of structural equation modeling*, Guilford Press, New York, NY (2012), pp. 472-494

Reedy et al., 2019. N.J. Reedy, E.R.E. Bowers, T.L. King. **Pregnancy related conditions.** T.L. King, M. Brucker, K. Osborne, C.M. Jevitt (Eds.), *Varney's midwifery* (6th ed.), Jones & Bartlett, Burlington, MA (2019), pp. 755-815

Rodrigues et al., 2019. I.B. Rodrigues, J.D. Adachi, K.A. Beattie, A. Lau, J.C. MacDermid. **Determining known-group validity and test-retest reliability in the PEQ (personalized exercise questionnaire).** *BMC Musculoskeletal Disorders*, 20 (1) (2019), p. 373, 10.1186/s12891-019-2761-3

Smid et al., 2020. S.C. Smid, D. McNeish, M. Miočević, R. van de Schoot. **Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review.** *Structural Equation Modeling: A Multidisciplinary Journal*, 27 (1) (2020), pp. 131-161, 10.1080/1070551.2019.1577140

Zielinski et al., 2015. R. Zielinski, K. Searing, M. Deibel. **Gastrointestinal distress in pregnancy: Prevalence, assessment, and treatment of 5 common minor discomforts.** *Journal of Perinatal & Neonatal Nursing*, 29 (1) (2015), pp. 23-31, 10.1097/JPN.0000000000000078