BMC Medical Ethics

RESEARCH

Open Access

Validity and reliability of the Oxford Utilitarianism Scale among obstetricians and gynecologists

Ayşe Figen Türkçapar[1*], Uğur Doğan[2] and Nüket Örnek Büken[3]

Abstract

Background Utilitarianism is a normative ethical theory that evaluates moral decisions based on their consequences, emphasizing the maximisation of overall well-being/outcomes for the greatest number. It serves as an important framework in clinical settings, where ethical dilemmas frequently arise. This study aimed to assess the validity and reliability of the Turkish version of the Oxford Utilitarianism Scale (OUS) among obstetricians and gynecologists (OB/GYNs).

Methods A total of 194 obstetricians and gynecologists participated in this online study and completed the OUS alongside a brief sociodemographic questionnaire. Construct validity was assessed using confirmatory factor analysis (CFA). Reliability was evaluated via internal consistency coefficients (Cronbach's alpha) and corrected item–total correlations. Additionally, measurement invariance analyses by sex were performed.

Results CFA confirmed a two-dimensional structure— impartial beneficence and instrumental harm—with good fit indices. High reliability was indicated by both internal consistency coefficients and corrected item-total correlations. The results of the measurement invariance analyses showed similar measurement properties across genders.

Conclusions The Turkish OUS demonstrated adequate construct validity and acceptable internal consistency in this OB/GYN sample. It can be used as a reliable and valid tool to support research, ethics education, and structured ethical reflection in obstetric practice.

Keywords Oxford Utilitarianism Scale, Utilitarianism, Medical ethics, Psychometric evaluation, Obstetricians and gynecologists, Obstetrics and gynecology, Ethical decision-making

*Correspondence:
Ayşe Figen Türkçapar
figenturkcapar@gmail.com; aysefigen.turkcapar@sbu.edu.tr
[1]Department of Obstetrics and Gynaecology, Ankara Gülhane Health Research Centre, Surgical Medical Sciences Section, Ankara, Turkey
[2]Department of Guidance and Counseling, Muğla Sıtkı Koçman University, Muğla, Turkey
[3]Faculty of Medicine, Department of History of Medicine and Medical Ethics, Hacettepe University, Ankara, Turkey

Introduction

Utilitarianism is a normative ethical theory—originating in moral philosophy and widely used in economics and public policy—that evaluates actions and decisions by their consequences, prioritising the maximisation of overall outcomes (e.g., well-being, welfare, or benefit) across affected individuals [1]. In its classical formulation, an action is morally right or optimal if it produces the greatest overall good for the greatest number, even when this involves personal sacrifice or trade-offs across stakeholders. This contrasts with deontological ethics, which holds that actions should conform to universal duties, rights, and rules, independent of their outcomes.

In normative ethics research, classical moral dilemma scenarios are commonly used to distinguish between utilitarian and deontological tendencies, helping to identify an individual's moral orientation [2]. Approving harm for the greater good is typically considered a utilitarian judgment, whereas disapproving of it is nonutilitarian or deontological [3]. However, real-life moral conflicts often require decisions about upholding or violating rights or obligations in pursuit of maximised outcomes [4]. In practice, many people do not make decisions that consistently align with a simple 'harm-for-greater-good' logic; prior research has highlighted important exceptions and limitations, including cases where utilitarian reasoning may neglect the welfare of all parties or be applied selectively depending on relationship closeness and context [5].

To enhance the understanding of utilitarian psychology, Kahane et al. (2018) proposed the '2D model' of moral utilitarianism and developed the Oxford Utilitarianism Scale (OUS) to assess its two dimensions [5]. In this model, instrumental harm (IH) is defined as causing harm, with the ultimate goal of achieving greater overall benefit. This perspective suggests that while principles such as honesty and protection of innocents are important, they may sometimes need to be violated to maximize overall well-being and benefit the majority. Impartial beneficence (IB) refers to the principle of promoting the greater good, even through personal sacrifices, based on the belief that everyone's well-being is equally important (e.g., donating a kidney). The OUS measures individual differences in adherence to these two utilitarian principles. Research has shown a positive association between empathic concern and high scores on the Impartial Beneficence subscale (OUS-IB). In contrast, high scores on the Instrumental Harm subscale (OUS-IH) are inversely related to empathic concern [5, 6].

Ethics, as a systematic study of moral principles, guides decision-making in moral dilemmas. In medical practice, the principle of beneficence emphasises promoting patient well-being and reducing suffering, but clinicians also face situations that require balancing benefits and harms across multiple stakeholders (e.g., mother–fetus conflicts, intensive care triage, or resource allocation) [7]. From a utilitarian perspective, ethically defensible decisions are those that maximise overall expected benefit—sometimes prioritising the welfare of many patients or the broader community over a single individual case—while making trade-offs explicit and proportionate. Accordingly, concepts such as 'medical appropriateness' can be interpreted not only as technical feasibility and benefit for an individual patient, but also as clinical decisions that are justified in terms of expected overall benefit within real-world constraints [8]. Because physicians' moral reasoning is shaped by professional norms and personal moral development (e.g., upbringing and religiosity), measuring utilitarian tendencies may help to characterise how clinicians approach ethically complex decisions and to inform ethics education and reflective practice [9].

Recent validity and reliability studies have demonstrated the structural consistency and psychometric validity of the OUS across diverse cultures [10, 11]. Two Turkish adaptation studies also supported the original two-dimensional structure, but they were conducted in non-physician samples [12, 13]. Ethical decision-making mechanisms can vary significantly in professional contexts, particularly in medical practice. Utilitarian reasoning in medicine is not merely an abstract preference: it can shape how clinicians evaluate conflicting interests (e.g., the mother's autonomy versus the fetus's welfare), interpret proportionality of harm, and respond to population-level constraints.

A brief, psychometrically sound measure such as the OUS could therefore be useful for (i) mapping patterns of ethical decision-making among clinicians, (ii) evaluating outcomes of medical ethics education, and (iii) informing reflective practice and policy discussions in high-stakes contexts such as obstetrics and perinatal care [14]. Obstetricians and gynecologists (OB/GYNs) represent a particularly unique professional group in this regard and are frequently confronted with decisions involving conflicting values such as maternal health and autonomy versus fetal well-being. The perinatal ethics professional responsibility model identifies situations where OB/GYNs must make utilitarian decisions and recommends that obstetric choices be evaluated within the framework of maternal autonomy, fetal benefit, and clinical appropriateness [14]. Empirical evidence regarding the psychometric performance of the OUS in OB/GYN specialists is lacking.

The aim of this study was to examine the psychometric properties of the Turkish version of the Oxford Utilitarianism Scale (OUS) among OB/GYN specialists by

evaluating its construct validity, internal consistency, and measurement invariance across sex.

Method

Participants

Data were collected between 1 February 2021 and 31 July 2021 via an anonymous, cross-sectional online survey. Because recruitment was conducted through an online voluntary survey, the sample may be affected by self-selection and coverage bias (e.g., physicians who are more active in digital networks may be overrepresented). In addition, although anonymity may reduce evaluation concerns, online self-report designs provide limited control over response authenticity and careless responding. The invitation letter and informed consent form, including a brief introduction and explanation of the study's purpose, were shared on the official website of the Turkish Society of Obstetrics and Gynaecology (TJOD), along with the survey link. Therefore, participation depended on both voluntary response and access to these professional communication channels, which may limit representativeness of the full OB/GYN population. The survey link remained accessible for six months. Additionally, the invitation and survey links were distributed via email and messaging services to TJOD members and to OB/GYN physicians in closed social media groups. The Turkish translation of the OUS used in this study was based on the version employed in Erzi's [13] study and was consistent with the translation used by Kumova [12]. The research project was reviewed and approved by Hacettepe University's Non-Interventional Clinical Research Ethics Committee on 19 November 2019 (decision no. 2019/27 – 07).

Sample size justification

This study was designed as a psychometric validation of a 9-item, two-factor scale (OUS) in a professional physician sample. In scale validation research, sample-size adequacy is typically evaluated in relation to model complexity, estimation method, and indicator quality rather than by a single universal cutoff [15, 16]. Considering the modest complexity of the CFA model and the subsequent multi-group invariance testing, the present sample ($N = 194$) was considered practically acceptable for primary model testing in this hard-to-reach OB/GYN population, while still requiring cautious interpretation for subgroup comparisons due to limited statistical power.

Instruments

A sociodemographic data form, developed by researchers based on previous studies, was used to collect information on participants' age, gender, marital status, years of specialization and professional title.

Oxford Utilitarianism Scale (OUS)

The Oxford Utilitarianism Scale (OUS) is a 9-item self-report measure assessing utilitarian tendencies along two dimensions: Impartial Beneficence (IB; 5 items) and Instrumental Harm (IH; 4 items) [5]. Items are rated on a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree). Total scores are computed by summing item responses (range 9–63), with subscale scores calculated as the sum of the relevant items (IB range 5–35; IH range 4–28); higher scores indicate stronger endorsement of the corresponding utilitarian dimension (See Appendix 1). The Turkish translation used in this study followed prior Turkish adaptations [12, 13].

Data analysis

Statistical analyses were conducted in JASP version 0.18.3 [16] (R-based) [17]. Construct validity was evaluated using confirmatory factor analysis (CFA) to test the hypothesised two-factor structure (IB and IH). Reliability was examined as internal consistency, using Cronbach's alpha for the total scale and each subscale, and corrected item–total correlations to assess the degree to which each item contributed to its scale score (acceptable ITC ≥ 0.30). Inter-item correlations were computed to describe item coherence. Measurement invariance across sex was assessed via multi-group CFA in a hierarchical sequence (configural, metric, and scalar invariance). Relationships between age, years of specialization, and OUS scores were examined using Pearson correlation coefficients. Additionally, potential differences across age groups (≤39, 40–49, ≥50 years) were examined using one-way ANOVA.

In addition, observed mean differences between men and women were explored with Welch's t-tests (two-tailed) for the total score and subscales; these comparisons were considered secondary to the invariance framework. Model fit was interpreted using commonly recommended criteria: $\chi^2/df < 5$; RMSEA < 0.08; GFI > 0.90; CFI ≥ 0.95; SRMR ≤ 0.08; and TLI ≥ 0.90 [18–20]. Maximum likelihood estimation was used for CFA and invariance analyses. In line with psychometric guidance, model fit and parameter interpretability were prioritised over rigid adherence to a single sample-size rule (see standard CFA guidance).

Results

Descriptive statistics

The sample comprised 194 obstetrics and gynecology specialists, with ages ranging from 29 to 71 years (mean age = 47.8 ± 8.9 years). Among the participants, 109 were female (56.2%), and 85 were male (43.8%). In terms of marital status, 165 participants (85.1%) were married, 13 (6.7%) were single, and 16 (8.2%) were divorced. The professional roles included 123 specialist physicians (63.4%),

Table 1 The mean, standard deviation, skewness, and kurtosis values for the total Oxford Utilitarianism Scale (OUS) score, as well as for its two subscales, Impartial Beneficence and Instrumental Harm

	Total (*N*:194)			**Male (*N*:85)**			**Female (*N*:109)**		
	TS(OUS)	**OUS(IB)**	**OUS(IH)**	**TS(OUS)**	**OUS(IB)**	**OUS(IH)**	**TS(OUS)**	**OUS(IB)**	**OUS(IH)**
Mean	34.747	22.294	12.454	34.565	21.776	12.788	34.890	22.697	12.193
Std. Dev	7.950	5.251	5.092	7.372	4.633	5.143	8.403	5.676	5.060
Skewness	-0.209	-0.381	0.252	-0.172	-0.141	0.049	-0.242	-0.558	0.419
Kurtosis	0.024	-0.062	-0.729	-0.230	-0.609	-0.780	0.119	0.154	-0.595
Minimum	12.000	8.000	4.000	17.000	11.000	4.000	12.000	8.000	4.000
Maximum	55.000	35.000	25.000	53.000	31.000	25.000	55.000	35.000	24.000

TS(OUS): total Oxford Utilitarianism Scale score (sum of 9 items; range 9–63). OUS(IB): Impartial Beneficence subscale (sum of 5 items; range 5–35). OUS(IH): Instrumental Harm subscale (sum of 4 items; range 4–28). Std. Dev.: standard deviation

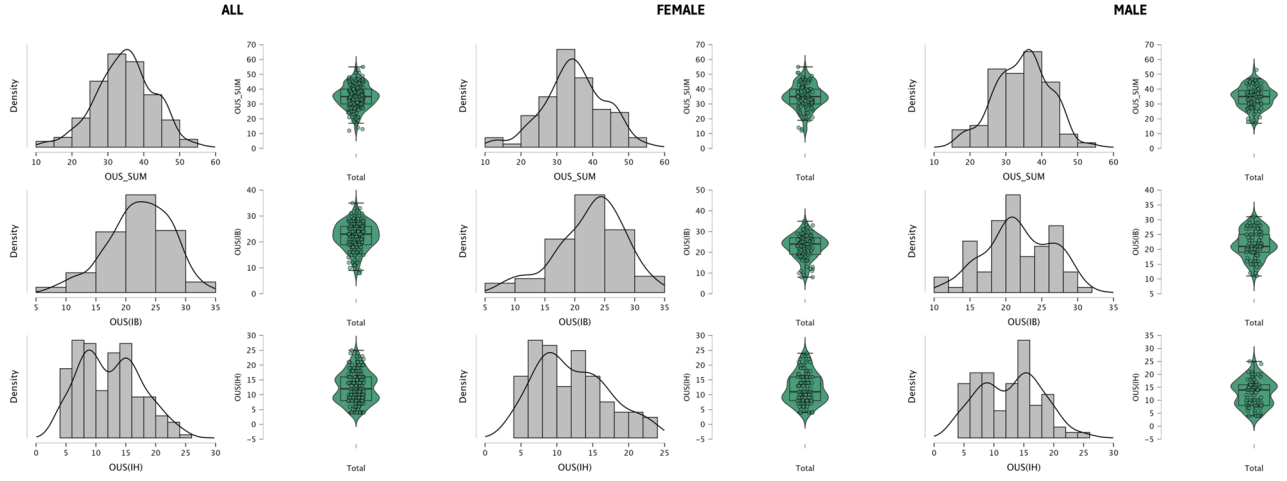

Fig. 1 The distribution of the data is visualized via histograms and boxplots

seven attending physicians (3.6%), and 64 (33.0%) holding academic positions. Five participants (2.6%) were assistant professors, 14 (7.2%) were associate professors, and 41 (21.1%) were full professors. The years of graduation from medical school ranged from 1976 to 2017, and the years of residency ranged from 1982 to 2021, with an average specialization duration of 18.0 ± 9.1 years. Regarding years of specialisation, 47 participants (24.2%) had 10 years or less, 63 (32.5%) had 11–20 years, and 84 (43.3%) had 21 years or more. When asked about the influence of religion on professional decision-making, 95 participants (49.0%) reported that religion never influenced their decisions; 42 (21.6%) said it influenced their decisions rarely; 41 (21.1%) said it influenced their decisions sometimes; 6 (3.1%) said it influenced their decisions often; and 10 (5.2%) said it influenced their decisions always.

Before reporting the CFA findings, descriptive statistics were calculated for the total OUS scores for the male, female, and combined groups. The mean, standard deviation, skewness, and kurtosis values for the total Oxford Utilitarianism Scale (OUS) score, as well as for its two subscales, Impartial Beneficence and Instrumental Harm, are shown in Table 1 across the entire sample and by gender group (male and female). The findings indicate that the mean and standard deviation values for the female group are slightly greater than those for the male group and the total sample. However, the means across all groups are fairly similar, suggesting comparable variability between groups.

The distribution of the data is visualized via histograms and boxplots in Fig. 1. In the overall sample, the distribution of OUS scores was slightly positively skewed, indicating an almost normal distribution pattern. The mean total score was 34.747 (SD = 7.950). The mean scores for the Impartial Beneficence and Instrumental Harm subscales were 22.294 (SD = 5.251) and 12.454 (SD = 5.092), respectively. Descriptive statistics by sex suggested broadly similar distributions. To empirically examine observed group differences, Welch's t-tests indicated no statistically significant differences between men and women for the total OUS score ($t = -0.29$, $df \approx 189.30$, $p = .775$), IB ($t = -1.24$, $df \approx 191.58$, $p = .215$), or IH ($t = 0.81$, $df \approx 179.24$, $p = .422$). These results are consistent with the invariance analyses reported below, supporting meaningful comparisons across sex. Pearson correlation analysis revealed no

statistically significant relationship between age and the total OUS score (r = .06, p = .43), the Impartial Beneficence subscale (r = .08, p = .29), or the Instrumental Harm subscale (r = .02, p = .84). To examine potential age-related differences, participants were categorised into three age subgroups: ≤39 years (n = 34), 40–49 years (n = 77), and ≥50 years (n = 83). One-way ANOVA revealed no statistically significant differences across age groups for the Total OUS score (F(2,191) = 0.37, p = .70), Impartial Beneficence (F(2,191) = 0.03, p = .97), or Instrumental Harm (F(2,191) = 1.06, p = .35).

Item-level descriptive statistics were subsequently calculated and are presented in Table 2. Examining the item means within the Impartial Beneficence dimension revealed that the highest mean (M = 5.278) was for item OUS3: 'Morally speaking, people should care about the well-being of all human beings on the planet equally; they should not give special priority to the well-being of those who are physically or emotionally close to them.' The lowest mean within this dimension was for item OUS4 (M = 3.742): 'Failing to help someone is as morally bad as actively harming them.' In the instrumental harm dimension, the highest mean was found for item OUS7 (M = 3.474): 'If achieving the greatest good for all people requires political oppression for a short and limited time, then political oppression is acceptable.' The lowest mean was for item OUS6 (M = 2.794): 'It is morally right to harm an innocent person if harming them is the only way to help several other innocent people.'

The standardized factor loadings of the items are represented in the estimate column of Table 2. As shown in the table, all the items had statistically significant factor loadings, indicating strong relationships with their respective latent constructs. Additionally, interitem correlations were computed to evaluate the relationships between individual items and their contribution to the scale's overall structure. The correlation heatmap is presented in Fig. 2, which visualizes the correlations between scale items. The color gradient represents the strength of the correlations: darker colors indicate stronger correlations, whereas lighter colors represent weaker correlations.

Positive interitem correlations were observed for the Impartial Beneficence items (OUS1, OUS2, OUS3, OUS4, and OUS5). Notably, strong positive correlations were observed between OUS2 and OUS3 (r = .464) and between OUS3 and OUS4 (r = .447). The correlation between OUS1 and OUS5 (r = .188) was lower, although still positive, suggesting that OUS5 may be less closely associated with the other items in this subscale. Similarly, positive correlations were also observed for the Instrumental Harm items (OUS6, OUS7, OUS8, and OUS9). Notably, high correlations were found between OUS6 and OUS8 (r = .596) and between OUS7 and OUS8 (r = .573). The correlation between OUS6 and OUS9 (r = .211) was lower than that between the other variables but still positive. Overall, the pattern of positive interitem correlations across both subscales suggests that the scale has a coherent and reliable structure.

Validity and reliability analysis of the Oxford Utilitarianism Scale (OUS)

Confirmatory factor analysis (CFA)

A confirmatory factor analysis was conducted to assess the construct validity of the OUS. The results, presented in Table 2, confirmed the two-factor structure of the OUS. The model fit indices indicated an acceptable fit: χ^2 = 29.13, df = 25, χ^2/df = 1.17, RMSEA = 0.039, CFI = 0.982, TLI = 0.974, and SRMR = 0.050. Furthermore, all factor loadings were statistically significant. The highest factor loading was observed for Item 8, which states that it is acceptable to torture an innocent person if it is necessary to obtain information that will prevent a bomb from killing hundreds of people. The lowest factor loading was found for Item 5: 'If one can help others by donating, then it is morally wrong to keep money that one doesn't truly need.'

Table 2 Item-level descriptive statistics

	Items	Mean	Sd	ITC	IIDM	Est.	SE	Lower	Upper	Std. Est.	R^2
Impartial Beneficence	OUS1	3.789	1.71	0.379	0.599	0.766	0.14	0.491	1.04	0.449	0.201
	OUS2	5.041	1.399	0.416	0.584	0.962	0.144	0.679	1.245	0.689	0.475
	OUS3	5.278	1.575	0.394	0.591	0.772	0.122	0.533	1.011	0.491	0.241
	OUS4	3.742	1.887	0.451	0.561	1.366	0.192	0.989	1.743	0.726	0.527
	OUS5	4.443	1.574	0.35	0.611	0.629	0.128	0.378	0.88	0.401	0.16
Instrumental Harm	OUS6	2.794	1.647	0.468	0.696	1.006	0.12	0.771	1.241	0.612	0.375
	OUS7	3.474	1.816	0.643	0.588	1.408	0.133	1.148	1.668	0.778	0.605
	OUS8	2.861	1.539	0.642	0.604	1.203	0.112	0.983	1.422	0.783	0.614
	OUS9	3.325	1.839	0.358	0.765	0.786	0.141	0.509	1.063	0.428	0.183

ITC: corrected item–total correlation (acceptable ≥ 0.30); IIDM: Cronbach's alpha if item deleted; LCI/UCI: lower/upper 95% confidence interval for the unstandardized loading estimate. For interpretation, α ≥ 0.70 is generally considered acceptable for group-level research; inter-item correlations of ~ 0.10, ~ 0.30, and ≥ 0.50 may be interpreted as small, moderate, and strong, respectively

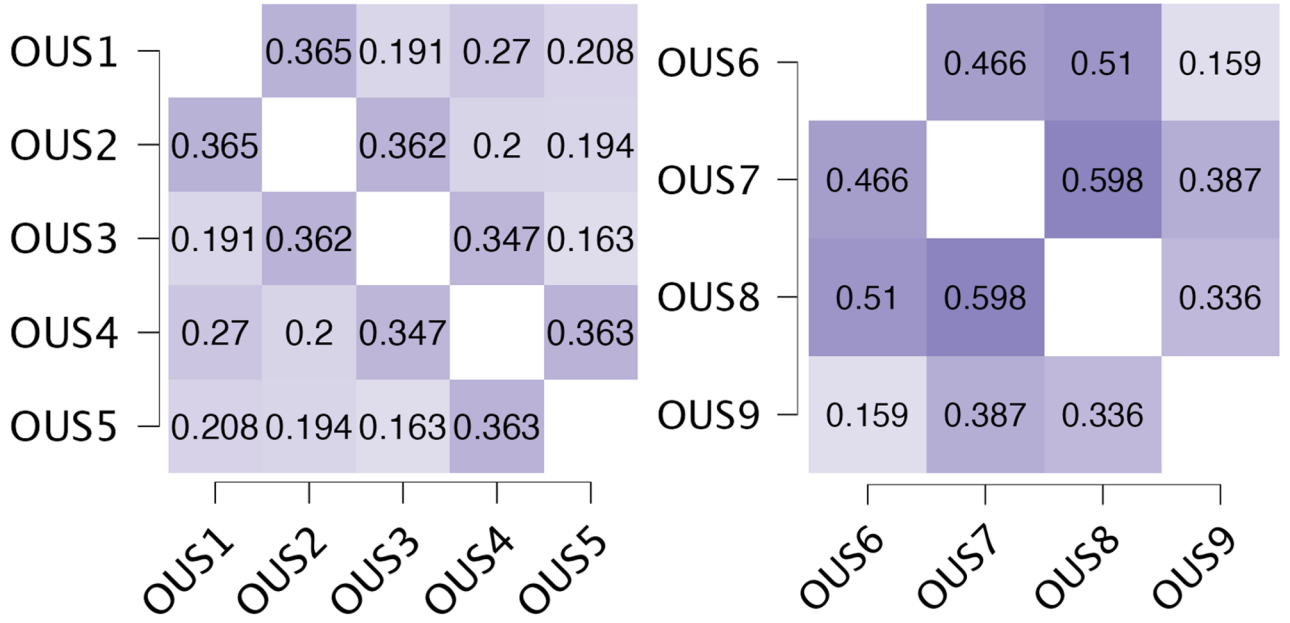

Fig. 2 Interitem correlation heatmap

Table 3 Measurement invariance of the OUS across male and female participants

			Baseline test			Difference test		
	AIC	BIC	χ^2	df	*p*	$\Delta\chi^2$	Δdf	*p*
Model 1	6473.738	6663.274	66.959	50	0 0.055			
Model 2	6471.242	6637.902	78.463	57	0.033	11.504	7	0.118
Model 3	6460.733	6597.983	85.954	66	0.050	7.491	9	0.586

Reliability analysis

Internal consistency

Cronbach's alpha coefficients were used to assess the internal consistency of the scale. The alpha coefficient for the overall scale was 0.74, and it was 0.75 for both the Impartial Beneficence subscale and the Instrumental Harm subscale. These results suggest that the scale has an acceptable level of internal consistency [21, 22].

Item–total correlation

Item–total correlations, a common measure of internal consistency, were also calculated. A corrected item–total correlation above 0.30 is considered acceptable [23, 24]. For the Impartial Beneficence subscale, item–total correlations ranged from 0.350 to 0.451; for the Instrumental Harm subscale, they ranged from 0.358 to 0.643. The specific values for each item are presented in Table 2.

Measurement invariance

To examine the measurement invariance of the OUS across male and female participants, a multigroup CFA was performed. The analysis began with configural invariance, which tests whether the same factor structure is applied across genders. The results of this analysis are presented in Table 3. Model 1 served as the baseline configural model with no cross-group constraints. The fit indices — CFI = 0.950, TLI = 0.928, and RMSEA = 0.059 — indicated a generally good fit, suggesting that the model adequately represented the data for both groups independently.

Next, Model 2 (Metric Invariance), which constrained the factor loadings to be equal across genders, was tested. The slight decrease in CFI (from 0.950 to 0.936) and TLI (from 0.928 to 0.920), along with a minimal increase in RMSEA (from 0.059 to 0.062), indicated a minor reduction in model fit. However, these values remained within acceptable ranges and supported the assumption of metric invariance. The change in chi-square ($\Delta\chi^2$ = 11.504, Δdf = 7, p = .118) was nonsignificant, which also supported metric invariance. Finally, Model 3 (Scalar invariance) was examined by constraining both factor loadings and item intercepts across genders. The CFI and TLI improved slightly (to 0.941 and 0.935, respectively), and the RMSEA remained stable at 0.056, indicating an adequate model fit. The change in chi-square from Model 2 to Model 3 ($\Delta\chi^2$ = 7.491, Δdf = 9, p = .586) was again nonsignificant. This provides strong support for scalar invariance, suggesting that item intercepts can be considered equivalent across genders without substantially reducing the model fit. Taken together, these findings support the

metric and scalar invariance of the OUS across gender groups. This implies that the scale measures comparable constructs for both men and women, enabling valid comparisons of latent means in subsequent analyses.

Discussion

This study aimed to evaluate the psychometric properties of the Turkish version of the Oxford utilitarianism scale (OUS) among a sample of obstetricians and gynecologists and to assess utilitarian tendencies in their ethical decision-making processes. The construct validity of the OUS was assessed via confirmatory factor analysis (CFA), and the resulting factor structure was compared with that of the original scale [6] and previous Turkish adaptations [12, 13].

The CFA results confirmed the two-factor structure—impartial beneficience and instrumental harm—which is consistent with the original model by Kahane et al. [6]. While Erzi's [13] adaptation excluded OUS item 7 because the factor loading was less than 0.4, Kumova's study [12] reported a factor loading of 0.59 for this item. In the present study, all factor loadings were statistically significant, and the item structure of the original scale was preserved without modification.

The reliability of the scale was evaluated via Cronbach's alpha internal consistency coefficients. The alpha coefficient was found to be 0.74 for the total scale and 0.75 for both subscales. While these values are slightly lower than those reported in the original scale development study (where the total scale and subscales showed alpha values of 0.87, 0.81, and 0.77, respectively), they remain above commonly accepted thresholds for psychometric adequacy. The lower reliability coefficients in the present study may be attributed to the smaller sample size. The corrected item–total correlations were also analysed to further assess internal consistency, ranging from 0.350 to 0.451 for Impartial Beneficence and 0.358–0.643 for Instrumental Harm. In the original scale, item factor loadings ranged from 0.55 to 0.88 [6]. Turkish adaptations reported similar ranges, with corrected item-total correlations between 0.44 and 0.70. The presence of positive interitem correlations throughout the scale supports the conclusion that the OUS demonstrates a consistent and reliable internal structure. These findings are consistent with the results of other cultural validations, including the French adaptation [11], the Polish study [7], and the Turkish version by Kumova et al. [12], all of which confirmed the two-dimensional structure of the OUS and its internal consistency, suggesting its cross-cultural applicability.

The higher scores for impartial benefit (OUS-IB) compared to instrumental harm (OUS-IH) identified in our study show a trend consistent with previous validity studies in the literature. This profile, expected within a clinician sample, is inherently linked to the fundamental philosophy of medical ethics; indeed, modern clinical ethics is a delicate balancing mechanism built upon the principles of beneficence, non-maleficence, autonomy, and justice [25]. However, obstetric practice entails unique ethical challenges due to the management of 'dual patients' (maternal and fetal), where these universal principles frequently conflict. This distinctive context requires the physician to uphold general ethical tenets while simultaneously establishing a balance between the fetus and the mother, thereby rendering the tendency toward impartial beneficence—focused on collective well-being—more pronounced in clinical reasoning.

In critical processes that center the physician's moral judgment, such as prenatal screening, termination of pregnancy for severe fetal anomalies, and decisions regarding periviable prematurity, obstetricians operate within a professional ethical framework that emphasizes compassion, non-maleficence, and patient-centered beneficence. The prenatal screening process can function in clinical practice not only as a diagnostic step but also as an ethical mechanism of "selection" and "preparation" for possible outcomes. When interpreted through the subdimensions of the Oxford Utilitarianism Scale (OUS), it has been reported that as the tendency toward instrumental harm increases, interventions aimed at preventing genetic defects (e.g., termination) may be evaluated in a more rational and goal-oriented manner; and the screening test itself may be conceptualized as an instrumental step that "serves to prevent greater unhappiness that may arise in the future [26]. Similarly, the Impartial Beneficence orientation is seen as a guide that supports the effective use of resources (economy, time, care burden) at the family and community levels; and may frame the principle of "what is best for everyone" as a moral responsibility to protect overall welfare by enabling the earlier detection of a pregnancy that could have serious consequences in terms of quality of life.

Consequently, it is observed that physicians, who prioritize impartial concern for overall welfare rather than endorsing harm to achieve a particular end, lean toward an 'altruistic' utilitarianism focused on individual well-being and perspective-taking. Obstetricians are trained in a professional ethics of care that naturally aligns with promoting impartial concern for welfare rather than instrumental harm. Ultimately, this demonstrates that while obstetricians and gynecologists prioritize individual well-being and altruistic utilitarianism, they remain highly sensitive to actions involving potential harm. Obstetricians are trained within a professional care ethic that prioritizes impartial concern for well-being over instrumental harm. This ethical orientation demonstrates that they prioritize individual well-being and altruistic

benefit, while maintaining a high level of sensitivity to potentially harmful interventions.

Another notable finding of this study is the measurement invariance of OUS across genders. This finding is consistent with the original scale development study, which reported no significant gender difference in impartial beneficience or instrumental harm tendencies. In contrast, in the Spanish adaptation, gender had statistically significant main effects on both dimensions, with women scoring higher on Impartial Benficience and men scoring higher on Instrumental Harm [11]. This has been interpreted as reflecting that women are more prone to emotional and empathetic responses, while men are more prone to rational decision-making. However, the French adaptation [10] attributed gender differences to sampling imbalance, emphasizing that additional analyses with gender-balanced subsamples yielded similar findings. Based on these results, the authors concluded that gender may influence individuals' responses to moral dilemmas and should therefore be considered in future studies on moral psychology and utilitarian decision-making.

From an applied perspective, mapping IB and IH separately may help educators and policy makers distinguish between (i) clinicians who strongly endorse impartial welfare promotion (useful for discussions about equity, global health, and fair resource allocation) and (ii) clinicians who are more accepting of instrumental harm (relevant for high-stakes trade-offs such as emergency obstetrics, fetal–maternal conflicts, and public health constraints). Accordingly, the OUS may be used as a structured reflection tool in medical ethics curricula and as an outcome measure to evaluate educational interventions.

Limitations and future research

First, the study relied on a voluntary online sample of obstetricians and gynaecologists, which may introduce self-selection and coverage bias and therefore limits generalisability beyond this professional group and to physicians with lower digital engagement. In addition, although the total sample was adequate for the planned CFA and invariance framework, power for finer-grained subgroup comparisons may have been limited in this professional sample [27]. Second, the cross-sectional design precludes examination of temporal stability; thus, test–retest reliability could not be evaluated. Third, although the OUS is designed as a self-report instrument, administering it within a professional ethics context may elicit social desirability or profession-normative responding, which could attenuate variability in scores and contribute to conservative reliability estimates.

Future research should further evaluate the psychometric robustness of the Oxford Utilitarianism Scale (OUS) using larger samples, longitudinal designs, and cross-cultural comparisons, particularly across different medical specialties. In parallel, studies should explicitly consider how features of moral dilemma paradigms—such as personal versus impersonal contexts (e.g., direct harm vs. track-switching scenarios), the relational value of potential beneficiaries (e.g., saving a relative vs. a stranger), and decisions made under group membership or majority–minority dynamics—systematically shape moral judgments [2]. Given that individuals often endorse self-sacrifice for others' welfare while simultaneously exhibiting self-preservation tendencies that may be amplified when identifying with fih

the majority, examining how dilemma-based choices map onto the OUS's two-dimensional structure could yield a more nuanced understanding of the ethical determinants of medical decision-making. Moreover, integrating behavioral indices within dilemma paradigms would help clarify the correspondence between self-reported utilitarian orientations and actual decision patterns in ecologically relevant contexts. Future studies should combine online and offline recruitment, include explicit data-quality checks for careless responding, and evaluate temporal stability with test–retest designs.

Conclusion

Given that obstetricians and gynecologists frequently encounter ethical dilemmas in their clinical decision-making processes, the Oxford Utilitarianism Scale has been demonstrated to be applicable within this professional group. Using the OUS in studies aiming to understand the ethical dimensions of physicians' professional judgments and, more broadly, in ethics education may contribute to a deeper awareness of the roles played by empathy, impartial beneficence, and instrumental harm in moral reasoning.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12910-026-01457-9.

Supplementary Material 1.

Figure 1

Acknowledgements

Information about this part can be found on the title page.

Authors' contributions

Conceptualization, data collection, data analysis, and original draft writing: AFT; Statistical analysis: UD; Manuscript review and editing: NOB. All the authors read and approved the final manuscript.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to ethical and privacy considerations. De-identified data may be made available from the corresponding author upon reasonable request, subject to review of the proposed use and provided that the request is consistent with the participants' consent and institutional ethics approval.

Declarations

Ethics approval and consent to participate

This study was reviewed and approved by the Hacettepe University Non-Interventional Clinical Research Ethics Committee on 19 November 2019 (decision no. 2019/27 – 07). All procedures involving human participants were performed in accordance with relevant guidelines and regulations, including the ethical standards of the institutional research committee and the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Informed consent was obtained from all participants prior to participation.

Consent for publication

Not applicable. This manuscript does not contain any individual person's data in any form.

Competing interests

The authors declare no competing interests.

Received: 9 June 2025 / Accepted: 6 April 2026
Published online: 11 April 2026

References

1. Kahane G, Everett JA, Earp BD, Farias M, Savulescu J. Utilitarian'judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. Cognition Epub 2014 Nov 13 Jan. 2015;134:193–209. https://doi.org/10.1016/j.cognition.2014.10.005.
2. Park J, Shin Y, Kim S, et al. Effects of perspective switching and utilitarian thinking on moral judgments in a sacrificial dilemma among healthcare and nonhealthcare students. Curr Psychol. 2024;43:984–96. https://doi.org/10.1007/s12144-023-04380-z.
3. Conway P, Goldstein-Greenwood J, Polacek D, Greene JD. Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. Cognition. 2018;179:241–65. https://doi.org/10.1016/j.cognition.2018.04.018.
4. Everett JAC, Kahane G. Switching Tracks? Towards a Multidimensional Model of Utilitarian Psychology. Trends Cogn Sci. 2020;24(2):124–134. https://doi.org/10.1016/j.tics.2019.11.012. Epub 2020 Jan 3. PMID: 31911126.
5. Kahane G, Everett JAC, Earp BD, Caviola L, Faber NS, Crockett MJ, Savulescu J. Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. Psychol Rev. Epub. 2017 Dec 21. 2018;125(2):131–164. https://doi.org/10.1037/rev0000093. Erratum in: Psychol Rev. 2018;125(2):164. https://doi.org/10.1037/rev0000112.
6. Paruzel-Czachura M, Kocur DD, Meditation. Mindfulness, and Self-Compassion Impact Utilitarian Moral Judgment? *Mindfulness* 2023 June 14, 1650–1661. https://doi.org/10.1007/s12671-023-02161-0
7. Bester JC, Beneficence. Interests, and Wellbeing in Medicine: What It Means to Provide Benefit to Patients. Am J Bioeth. 2020;20(3):53–62. https://doi.org/10.1080/15265161.2020.1714793 PMID: 32105204.
8. FIGO Ethics and Professionalism Guidelines for Obstetrics and Gynecology. Edited by Frank A. Chervenak, M.D. and Laurence B. McCullough, Ph.D. https://www.figo.org/sites/default/files/2021-11/FIGO-Ethics-Guidelines-onlinePDF.pdf
9. Martins VSM, Santos CMNC, Bataglia PUR, Duarte IMRF. The Teaching of Ethics and the Moral Competence of Medical and Nursing Students. Health Care Anal Epub 2020 Sep. 2021;17(2):113–26. https://doi.org/10.1007/s10728-020-00401-1.
10. Carron R, Blanc N, Anders R, Brigaud E. The Oxford Utilitarianism Scale: Psychometric properties of a French adaptation (OUS-Fr). Behav Res Methods Epub 2023 Oct. 2024;4(5):5116–27. https://doi.org/10.3758/s13428-023-02250-x.
11. Pascual-Soler M, Berrios-Riquelme J, Gomez-Frias R, Caamaño-Rocha L, Frias-Navarro D, Pascual-Soler M, Berrios-Riquelme J, Gomez-Frias R, Caamaño-Rocha L, Frias-Navarro D. Oxford Utilitarianism Scale: Psychometric Properties of a Spanish Adaptation (OUS-Spa). SAGE Open. 2025;15(1). https://doi.org/10.1177/21582440251325342.
12. Kumova F, Bahçekapılı HG. Oxford Faydacılık Ölçeği'nin Türkçe Uyarlaması. Stud Psychol. 2021;41(2):491–520. https://doi.org/10.26650/SP2020-0055.
13. Erzi S. Psychometric Properties of Adaptation of The Oxford Utilitarianism Scale to Turkish. HUMANITAS- Uluslararası Sosyal Bilimler Dergisi. 2019;7(13):132–47.
14. Chervenak FA, McCullough LB, Brent RL. The professional responsibility model of obstetrical ethics: avoiding the perils of clashing rights. Am J Obstet Gynecol Epub 2011 Jun. 2011;12(4):315..e1-5. https://doi.org/10.1016/j.ajog.2011.06.006.
15. Kline RB. Principles and Practice of Structural Equation Modeling. 4th ed. Guilford Press; 2016.
16. Brown TA. Confirmatory Factor Analysis for Applied Research. 2nd ed. Guilford Press; 2015.
17. JASP Team. JASP [Computer software]. Version 0.18.3. 2024. Available from: https://jasp-stats.org/
18. R Core Team. R: A Language and Environment for Statistical Computing [Computer software]. Viyana: R Foundation for Statistical Computing. 2021. Available from: https://www.R-project.org/
19. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures [Internet]. Vols. 8, Methods of Psychological Research. Department of Psychology- University of Koblenz-Landau; 2003. pp. 23–74. Available from: https://psycharchives.org/en/item/1a8dea48-0285-4dac-a612-9dc0ff2532f6
20. Keith TZ. Multiple regression and beyond: An introduction to multiple regression and structural equation modeling. 2019 (3rd ed.). Routledge. https://doi.org/10.4324/9780429318290.
21. Brunner M, Süβ HM. Analyzing the reliability of multidimensional measures: An example from intelligence research. Educ Psychol Meas. 2005;65(2):227–40.
22. Flora DB. Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates. Adv Methods Practices Psychol Sci. 2020;3(4):484–501. https://doi.org/10.1177/2515245920951747.
23. Santoso A. Utilizing Fisher's-Z Transformation for Item Selection. ANIMA Indonesian Psychol. J. 2018 Apr.25;33(3). Available from: https://journal.ubaya.ac.id/index.php/jpa/article/view/1694
24. Green SB, Yang Y. Evaluation of Dimensionality in the Assessment of Internal Consistency Reliability: Coefficient Alpha and Omega Coefficients. Educational Measurement: Issues Pract. 2015;34(4):14–20. https://doi.org/10.1111/emip.12100.
25. Varkey B. Principles of Clinical Ethics and Their Application to Practice. Med Princ Pract. 2021;30(1):17–28. https://doi.org/10.1159/000509119 Epub 2020 Jun 4. PMID: 32498071; PMCID: PMC7923912.
26. Maftei A, Dănilă O. The good, the bad, and the utilitarian: attitudes towards genetic testing and implications for disability. Curr Psychol 2022 Jan 17:1–22. https://doi.org/10.1007/s12144-021-02568-9. Epub ahead of print. PMID: 35068904; PMCID: PMC8761521.
27. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. Educ Psychol Meas. 2013;73(6):913–34.

Publisher's Note