THE LANCET
Digital Health

Supplementary appendix
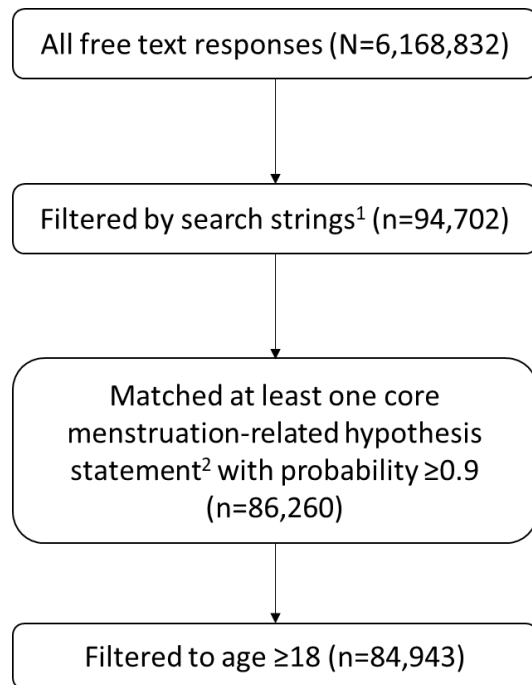
This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplementary material

Supplemental figure 1: Inclusion criteria

All free text responses (N=6,168,832)

↓

Filtered by search strings[1] (n=94,702)

↓

Matched at least one core menstruation-related hypothesis statement[2] with probability ≥0.9 (n=86,260)

↓

Filtered to age ≥18 (n=84,943)

[1] Matches pattern 1 or both patterns 2 and 3.

       Pattern 1: r'menses|menst|spotting|period(s\b|\b)|\bcycle|\bmiscarr|menorrh|metrorrh'

       Pattern 2: r'bleed|blood'

       Pattern 3: r'menop|\bpreg|\buter|hyster|placent|\bvag|breakthrough|(break through)|endomet|\bgest|term\b|trimester'
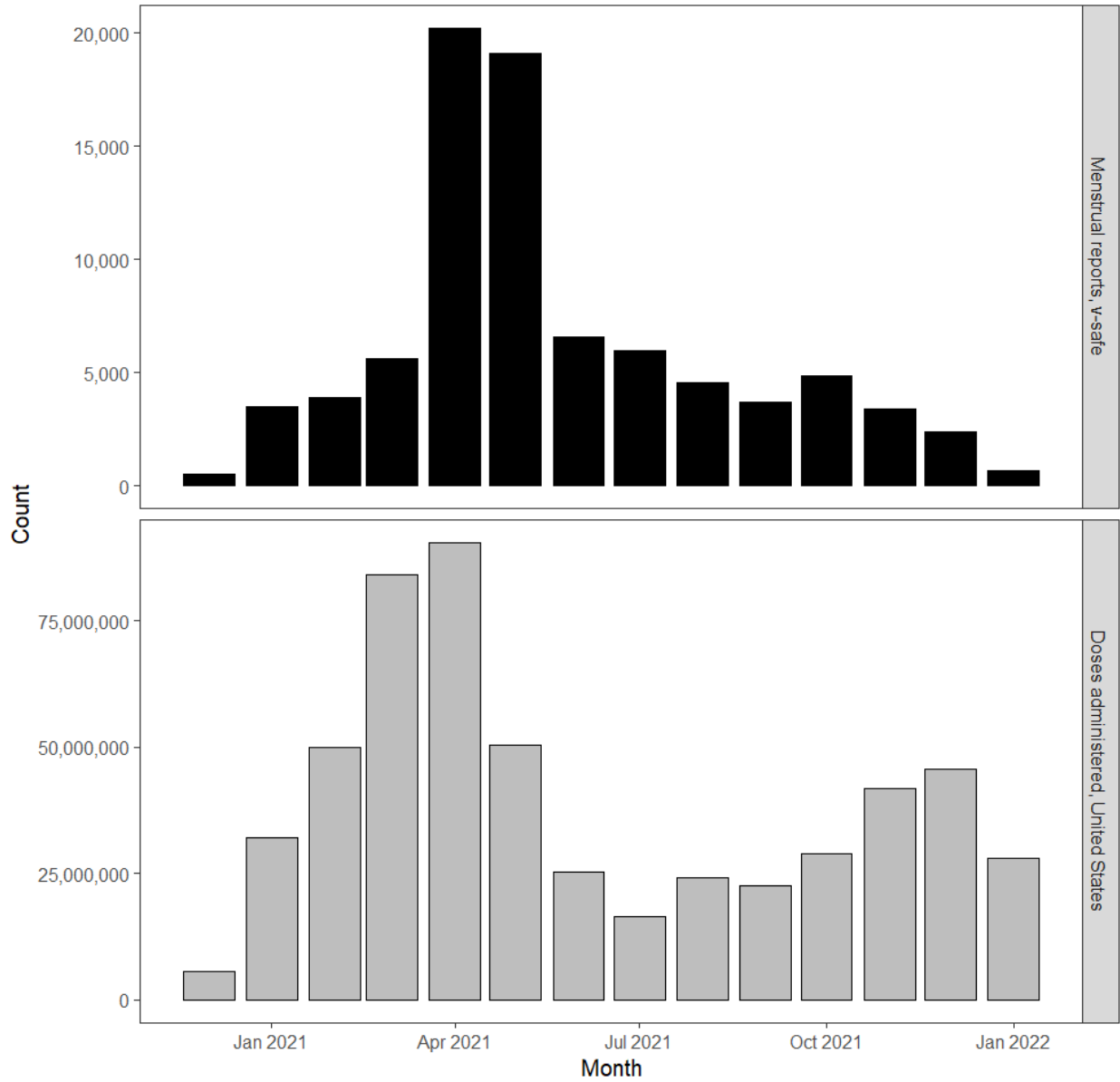
[2] Core menstruation-related hypothesis statements:

- 'This mentions menstruation'
- 'This mentions spotting'
- 'This mentions vaginal bleeding'
- 'This mentions uterine bleeding'
- 'This specifically mentions missing or skipping a period'

Supplemental table 1: Criteria for classifying responses into themes

Theme	Criteria
Timing	Entailment probability ≥0·9 with any of the following statements: • 'This mentions a period that came early', • 'This mentions a period that came late', • 'This mentions spotting', • 'This mentions an irregular period specifically', • 'This specifically mentions missing or skipping a period'
Severity	Entailment probability ≥0·9 with any of the following statements: • 'This mentions a heavy menstrual period', • 'This mentions a painful menstrual period', • 'This mentions prolonged bleeding specifically'
Menopausal	Response contains "menop" or respondent is aged ≥59 years
Resumption of menses	Entailment probability ≥0·9 with the following statement: • 'This mentions not having a period for years'
Other	Does not meet criteria for any of the four other themes

Supplemental figure 2: Number of v-safe reports related to menstrual irregularities and vaginal bleeding and number of COVID-19 vaccine doses administered, United States, by month

Supplemental table 2: Evaluation of classification performance against manually classified samples

Classification objective	Sample data	TP	FP	TN	FN	Sensitivity (95% CI)	Specificity (95% CI)	F1 (95% CI)
Identify menstruation-related responses from all responses using search strings	2000 random responses	17	1	1982	0	1 (0.771, 1)	0.999 (0.997, 1)	0.971 (0.778, 1.0)
Identify core menstruation-related responses using ZSL from search string-filtered responses	1000 random responses matching search strings	912	1	71	16	0.983 (0.972, 0.990)	0.986 (0.915, 0.999)	0.991 (0.985, 0.994)
Identify timing-related responses using ZSL	928 random menstruation-related responses[a]	690	83	125	30	0.958 (0.940, 0.971)	0.601 (0.531, 0.667)	0.924 (0.908, 0.937)
Identify severity-related responses using ZSL	928 random menstruation-related responses[a]	510	88	281	49	0.912 (0.885, 0.934)	0.762 (0.714, 0.803)	0.882 (0.859, 0.899)
Identify responses related to resumption of menses using ZSL	928 random menstruation-related responses[a]	17	6	898	7	0.708 (0.488, 0.866)	0.993 (0.985, 0.997)	0.723 (0.538, 0.847)
Identify "other" menstruation-related responses[b]	928 random menstruation-related responses[a]	15	25	873	15	0.5 (0.332, 0.668)	0.972 (0.959, 0.982)	0.429 (0.275, 0.581)

ZSL: Zero-shot learning; TP: true positives; TN: true negatives; FP: false positives; FN: false negatives; CI: confidence interval

[a] Responses matching search strings deemed menstruation-related after manual review (1000 excluding 72 unrelated responses)

[b] Responses were categorized as "other" if they did not meet criteria for timing, severity, or resumption of menses using zero-shot classification, or for menopausal bleeding (age- and search string-based criteria). This was compared with manual classification into an "other" category.

Supplemental table 3: V-safe female respondents aged ≥18 years by age group and vaccine received

	18–49 years n=3,095,361	50–64 years n=1,698,520	65–74 years n=887,481	75–84 years n=249,567	85+ years n=44,434
BNT162b2	1622348 (52.4)	831395 (48.9)	391331 (44.1)	110824 (44.4)	19540 (44)
mRNA-1273	1352143 (43.7)	796965 (46.9)	483701 (54.5)	136861 (54.8)	24344 (54.8)
Ad26.COV2.S	166818 (5.4)	94905 (5.6)	20054 (2.3)	3241 (1.3)	757 (1.7)
Other/unknown	820 (<0.1)	399 (<0.1)	313 (<0.1)	118 (<0.1)	12 (<0.1)

Data are n (%).