

Supplementary Materials

Missing-Data Sensitivity Analyses

In the primary analyses, non-response on the physical-strain rating items and the CDC Healthy Days count items was coded 0. Because the analytic sample was restricted to participants who completed the survey, a non-response on these items was interpreted as denoting absence of the construct (no strain in that region, or no unhealthy days) rather than missing information. Tables S1–S5 repeat every coding-sensitive analysis using a complete-case coding, in which the same responses are instead treated as missing, so that readers can assess whether the substantive conclusions depend on this decision.

Table S1. Anatomical strain severity under the primary (0-coded) and sensitivity (complete-case) coding strategies.

Strain region (0–5)	KY/TN sample		Full sample	
	Primary	Sensitivity	Primary	Sensitivity
Back	1.95 ± 1.53	2.21 ± 1.44	1.93 ± 1.51	2.22 ± 1.40
Shoulder/neck	1.27 ± 1.36	1.65 ± 1.33	1.22 ± 1.31	1.63 ± 1.27
Knee/hip	0.83 ± 1.04	1.39 ± 1.02	0.78 ± 1.03	1.37 ± 1.03
Hand/wrist	0.66 ± 1.12	1.33 ± 1.29	0.59 ± 1.06	1.26 ± 1.26
Fatigue	1.38 ± 1.50	1.89 ± 1.46	1.34 ± 1.44	1.84 ± 1.39
Mean overall strain	1.22 ± 0.95	1.70 ± 1.02	1.17 ± 0.92	1.67 ± 1.00

Note. Values are mean ± standard deviation. Under the primary coding, non-engagement of a strain slider was treated as no strain (0), so denominators are the full respective samples (KY/TN N = 86; full N = 100). Under the sensitivity (complete-case) coding, the same responses are treated as missing, and denominators vary by item: KY/TN n = 76 (back), 66 (shoulder/neck), 51 (knee/hip), 43 (hand/wrist), 63 (fatigue), 80 (mean overall); full sample n = 87, 75, 57, 47, 73, and 94, respectively. The rank order of anatomical regions is identical under both codings, with back strain highest and hand/wrist strain lowest.

Table S2. Self-reported health status indicators under both coding strategies, with BRFSS reference values.

Measure	BRFSS		KY/TN sample		Full sample	
	KY	TN	Primary	Sensitivity	Primary	Sensitivity
Fair or poor general health (%)	23.9	23.4	7.0	7.0	8.0	8.0
Physically unhealthy days	4.8	4.4	2.2	5.3	2.6	5.9
Mentally unhealthy days	5.3	5.0	1.8	5.6	2.1	6.3
Activity-limitation days	3.0	2.7	1.4	5.9	1.9	6.8

Note. Day counts are means for the past 30 days. BRFSS values are population benchmarks (KY 2024; TN 2023), not a comparison group; no statistical comparisons were performed. Under the primary coding, a blank response on the day-count items was treated as zero days, so denominators are the full respective samples (KY/TN N = 86; full N = 100). Under the sensitivity coding these responses are treated as missing, and denominators fall to KY/TN n = 35, 28, and 21 and full-sample n = 44, 33, and 28 for physically unhealthy, mentally unhealthy, and activity-limitation days, respectively. Self-rated general health involves no zero-coding ambiguity and is therefore identical under both codings. The complete-case day counts approach or exceed the BRFSS population values for the same participants who report a fair/poor health rate roughly one-third of the BRFSS estimate, indicating that respondents reporting no unhealthy days characteristically left these items blank.

Table S3. Age-group comparison of mentally unhealthy days under both coding strategies.

Coding strategy	Sample	Under 55, mean (SD)	55+, mean (SD)	Effect estimate (r)	p
Primary (0-coded)	KY/TN	4.39 (6.85)	0.87 (2.12)	0.43	<0.001
	Full	4.74 (6.56)	1.11 (3.37)	0.46	<0.001
Sensitivity (complete-case)	KY/TN	7.21 (7.59)	3.93 (2.92)	0.34	0.126
	Full	7.53 (6.90)	5.06 (5.76)	0.37	0.072

Note. Values are mean number of mentally unhealthy days in the past 30 days. p-values are from Wilcoxon rank-sum tests, and the effect estimate is the rank-biserial correlation corresponding to that test, signed so that positive values indicate more unhealthy days among participants under 55. Under the primary coding, a blank response was treated as zero days, so denominators are the full respective samples (KY/TN: n = 23 under 55, n = 63 aged 55+; full: n = 27 and n = 73). Under the sensitivity coding these responses are treated as missing, reducing the comparison to KY/TN n = 14 and n = 14 and full-sample n = 17 and n = 16. The direction and magnitude of the effect are consistent across codings, but the comparison reaches conventional significance only under the primary coding. The complete-case subsample excludes participants reporting no unhealthy days, which removes most of the younger and older participants in good mental health and substantially reduces statistical power; the manuscript therefore reports the primary result while noting that it is not robust to the alternative coding. The accompanying comparison of sex composition by age group involves no zero-coding ambiguity and is identical under both codings (Fisher exact p = 0.027 in KY/TN and p = 0.166 in the full sample).

Table S4. Subgroup comparisons of mean strain under both coding strategies.

Comparison	Sample	Group 1, mean (SD)	Group 2, mean (SD)	Effect estimate (r)	p
<i>Primary coding (0-coded)</i>					
Operation type	KY/TN	1.03 (1.01)	1.32 (0.85)	0.24	0.058
Operation type	Full	1.02 (0.94)	1.26 (0.84)	0.19	0.100
Sex	KY/TN	1.11 (1.06)	1.29 (0.88)	0.16	0.204
Sex	Full	1.11 (1.01)	1.22 (0.86)	0.12	0.314
<i>Sensitivity coding (complete-case)</i>					
Operation type	KY/TN	1.38 (1.04)	1.91 (0.92)	0.34	0.011
Operation type	Full	1.43 (1.03)	1.83 (0.92)	0.26	0.029
Sex	KY/TN	1.52 (1.04)	1.82 (0.99)	0.20	0.141
Sex	Full	1.56 (1.00)	1.73 (1.00)	0.12	0.334

Note. For operation-type comparisons, Group 1 = hobbyist and Group 2 = sideline/commercial; for sex comparisons, Group 1 = female and Group 2 = male. All p-values are from Wilcoxon rank-sum tests, and the effect estimate is the rank-biserial correlation corresponding to that test, signed so that positive values indicate higher strain in Group 2. Sex comparisons are non-significant under both codings. The operation-type comparison is the only analysis in which statistical significance differs between codings: sideline/commercial beekeepers report higher mean strain under both codings and in both samples, but the difference reaches conventional significance only under the complete-case coding. Because the complete-case coding retains only participants who engaged the strain sliders – a subgroup that by definition reported some strain – the primary analysis is the more conservative estimate, and the difference is reported in the manuscript as a non-significant trend.

Table S5. Proportional-odds regression predicting back-strain severity under both coding strategies.

Coding strategy	Predictor	OR (95% CI)	p	N
Primary (0-coded)	Colony count	1.73 (1.19, 2.51)	0.004	100
	Lifting-aid use	1.03 (0.74, 1.43)	0.866	100
Sensitivity (complete-case)	Colony count	1.78 (1.20, 2.66)	0.005	87
	Lifting-aid use	0.91 (0.64, 1.30)	0.603	87

Note. Odds ratios are per one-category increase on the ordinal predictor. Colony count was modeled across five categories (1–5 to more than 200) and lifting-aid use across five categories (never to always). The association between colony count and back-strain severity is essentially unchanged across coding strategies, indicating that the principal multivariable finding of the study is not sensitive to the treatment of non-response. Logistic regression predicting injury and Poisson regression predicting tool ownership do not involve the strain or Healthy Days items and are therefore unaffected by the coding decision; they are not repeated here.

Summary of sensitivity findings

Across Tables S1–S5, the substantive conclusions of the study are stable under both coding strategies. Strain severity is uniformly higher under complete-case coding, as expected when participants who did not engage the sliders are excluded, but the rank order of anatomical regions is unchanged (Table S1). The principal multivariable association between colony count and back-strain severity is essentially identical (OR = 1.73 versus 1.78; Table S5). One comparison differs in statistical significance: the difference in mean strain between hobbyist and sideliner/commercial beekeepers is significant under complete-case coding but not under the primary coding (Table S4), and it is therefore reported in the manuscript as a non-significant trend. Conversely, the greater number of mentally unhealthy days reported by participants under 55 is significant under the primary coding but not under complete-case coding (Table S3), where the comparison retains only 28 to 33 participants; that finding is reported in the manuscript with an explicit note of this limitation. The Healthy Days comparison in Table S2 demonstrates why the primary coding was selected for those items: complete-case estimates imply that beekeepers experience as many or more unhealthy days than the general population, while the same participants report a rate of fair/poor general health roughly one-third of the state BRFSS estimate.