

Supplemental Figure 1 (a – b). Schematics of visits to 4 different fluid milk processing facilities for (a) sampling visit 1, where 2 ½-pint (473-mL) single-serve fluid milk cartons were collected from each of the 4 lanes and the 8 cartons from all lanes were commingled into 1 ½-gallon (1.9-L) container to represent a single-sample. This was repeated in triplicate. Each of 4 milk types were collected (white skim, white 1%, chocolate skim, and chocolate 1%). Samples of each milk type were collected for each of 3 time points defined as the beginning (first 1,000 cartons), middle (approximate middle of production of milk type), and end of production (last 1,000 cartons). For (b) sampling visit 2, 16 ½-pint (473-mL) single-serve fluid milk cartons were collected from each of the 4 lanes and all 16 cartons from a single lane were commingled into 1 1-gallon (3.8-L) container to represent a single-sample. Only a single replicate was collected. Each of 4 milk types were collected (white skim, white 1%, chocolate skim, and chocolate 1%). Samples of each milk type were collected for each of 1 time points defined as the beginning (first 1,000 cartons) and end of production (last 1,000 cartons).

Supplemental Figure 2. Picture and indication of the mandrel and mandrel hubs of an N-8 single-serve fluid milk filler

Supplemental Figure 3. Root cause analysis (RCA) fishbone diagram used for determining potential causes of gram-negative contamination in single-serve milk. This diagram was used during the pre-intervention visits and presented during a discussion with all quality management staff from all 4 facilities.

Supplemental Figure 4 (a – d). Day 14 gram-negative spoilage frequency [defined as a sample having > 20,000 cfu/mL standard plate count (SPC) and typical growth (i.e., ≥1 red colony) on Crystal Violet Tetrazolium Agar (CVTA)] for commingled samples collected over the initial two sampling visits, separated by (a) processing facility; (b) milk type and fat level (WS = white skim, W1 = white 1%, CS = chocolate skim, C1 = chocolate 1%); (c) sampling time point (B = beginning, M = middle, E = end), (data represent only sampling visit 1); and (d) filler lane (data represent only sampling visit 2). For figures (b) through (d) data are separated by panels by facilities 1 to 4 (indicated on top of each panel)

Supplemental Figure 5 (a – b). Association between gram-negative spoilage frequency (%) and number of observed downtimes due to (a) any reason and (b) filler and non-filler related reasons, separated by facility (color) and sampling visit (shape). Estimated average production times for the initial 2 sampling visits were 15 hours for facility 1, 11 hours for facility 2, 17 hours for facility 3, and 13 hours for facility 4. As it was usually not feasible to observe the entire production run, number of occurrences per h of production time could not be reported

Supplemental Figure 6 (a – b). Non-metric multidimensional (NMDS) scaling plot for assessing genus level classification based on 16S rRNA gene sequencing data between (a) lanes (shape) and processing facility (color), and (b) processing facility (color) and sample type (shape)