Supplementary Appendix for "Trends in U.S. Burden of *Clostridioides difficile* Infection and Outcomes"

Table of Contents

List of members of the Emerging Infections Program *C. difficile* Infection Working Group:

Lauren Korhonen, MPH[1], Brittany Martin, MPH[2], Geoffrey Brousseau, MPH[3], Wendy Bamberg, MD[3], Elizabeth Basiliere, AAS[3], James Meek, MPH[4], Rebecca Perlmutter, MPH[5], Maria Bye, MPH[6], Tory Whitten, MPH[6], Emily B. Hancock, MS[7], Rebecca Tsay, MPH, MLS[8], Deborah Nelson, MSN, RN[8], Trupti Hatwar, MPH[8], Valerie L.S. Ocampo, RN, BSN, MIPH[9], Miranda D. Smith, MPH[10], Ashley Paulick, BS[1], Michelle Adamczyk, BS[1]

[1]Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention, Atlanta, GA

[2]California Emerging Infections Program, Oakland, CA

[3]Colorado Department of Public Health and Environment, Denver, CO

[4]Connecticut Emerging Infections Program, Yale School of Public Health, New Haven, CT

[5]Maryland Department of Health, Baltimore, MD

[6]Minnesota Department of Health, St Paul, MN

[7]University of New Mexico, New Mexico Emerging Infections Program, Albuquerque, NM

[8]New York Emerging Infections Program and University of Rochester Medical Center, Rochester, NY

[9]Oregon Health Authority, Portland, OR

[10]Tennessee Department of Health, Nashville, TN

Funding source and author and working group member roles:

This work was supported through the Centers for Disease Control and Prevention's Emerging Infections Program Cooperative Agreement with funds from the CDC's Division of Healthcare Quality Promotion in the National Center for Emerging and Zoonotic Infectious Diseases. Author and working group member roles are as follows:

Project concept or design: Alice Y. Guh, Yi Mu, Lisa G. Winston, Helen Johnston, Wendy Bamberg, James Meek, Monica M. Farley, Lucy E. Wilson, MD, Stacy M. Holzbauer, Erin C. Phipps, Ghinwa K. Dumyati, Zintars G. Beldavs, Marion A. Kainer, Maria Karlsson, Dale N. Gerding, L. Clifford McDonald.

Data acquisition: Brittany Martin, Geoffrey Brousseau, Elizabeth Basiliere, Danyel Olson, Rebecca Perlmutter, Maria Bye, Tory Whitten, Emily B. Hancock, Rebecca Tsay, Deborah Nelson, Trupti Hatwar, Valerie L.S. Ocampo, Miranda D. Smith

Molecular characterization of isolates: Ashley Paulick, Michelle Adamczyk, Maria Karlsson

Data analysis: Alice Y. Guh, Yi Mu, L. Clifford McDonald

Alice Y. Guh wrote the first draft of the manuscript. All of the authors vouch for the completeness and accuracy of the data, and all authors decided to submit the manuscript for publication.

Methods

Epidemiologic classification

A case was classified as community-onset if the *C. difficile*-positive stool was collected as an outpatient or within 3 days of hospital admission. A case was classified as healthcare-facility onset (HCFO) if it was a hospital-onset CDI (positive stool was collected >3 days after hospital admission) or a long-term care facility (LTCF) onset CDI (positive stool collected in a LTCF or from a LTCF resident admitted to a hospital). Community-onset cases were further classified as community-associated if there was no documentation of admission to a healthcare facility in the preceding 12 weeks; all other

community-onset cases were considered community-onset healthcare-facility associated (CO-HCFA). Both the CO-HCFA cases and all HCFO cases were further classified as healthcare-associated CDI.

Isolate collection

A convenience sample of laboratories at each of the 10 EIP sites has been submitting stool samples for *C. difficile* culturing since 2011. Because of the increasing incidence of community-associated CDI, the initial focus of stool collection was on community-associated strains. Therefore, prior to 2016, each EIP site submitted up to 50% of their community-associated specimens, up to 5 CO-HCFA specimens, and up to 5 HCFO specimens per month. Starting in 2016, specimen submission process was revised to be more representative of the geographic distribution and epidemiologic classification of cases across all sites (i.e., the number of specimens submitted from each site was proportional to the number of CDI cases that were identified at each site and reflected each site's distribution of community-associated, CO-HCFA, and HCFO CDI).

Statistical analysis

The CDI datasets of each surveillance year that were used for our analyses were generated on different dates: the 2011 CDI data were generated on May 22, 2013; the 2012 data on March 26, 2014; the 2013 data on January 5, 2015; the 2014 data on June 7, 2018; the 2015 data also on June 7, 2018; the 2016 data on July 23, 2018, and the 2017 data on March 27, 2019. Since surveillance data can change, any annual dataset generated after these dates may produce slightly different results from our analyses. For generating the annual CDI burden estimates, we included data from all counties that participated in surveillance for the entire year (only one county was excluded from the 2012 burden estimate because it started surveillance mid-year). For the trends analysis, we included data from all counties that consistently participated since 2011.

We calculated the sampling weights (W1) for generating the national CDI burden estimates by dividing the estimated number of CDI cases by the corresponding observed number of CDI cases, stratified by age, sex, race and epidemiologic class. The sampling weights (W2) for the census were calculated by dividing the US census population by the EIP population, stratified by age, sex, race, and EIP sites. We excluded infants under the age of 1 year from the census population since they were not included in our analysis. The product of the cases' sampling weights and census weights (W1*W2) were the final weights used for generating the national community-associated and healthcare-associated CDI burden estimates and corresponding 95% confidence intervals. We then calculated the total national burden of CDI by adding the national estimates of community-associated and healthcare-associated CDI and 95% confidence intervals. Using the same methodology as above, we replaced the number of CDI cases with each of the CDI-associated outcomes to calculate their respective sampling weights for the national estimates of first CDI recurrence, hospitalizations, and in-hospital deaths and corresponding 95% confidence intervals. We used the U.S. Census data to calculate the national incidence rates of CDI and associated outcomes.

We used weighted multilevel models for the trend analyses, with observation level weights (W1) applied to either cases or outcomes (where applicable) and cluster (site) level weights (W2) applied separately to EIP sites. We used multiple imputed data for all analyses and combined the results to account for imputation errors. SAS 9.4 was used for all statistical analyses, and the variances and confidence intervals of the burden estimates were produced through Taylor series method (PROC SURVEYMEANS).

Results

Of the estimated national burden of 235,700 healthcare-associated CDI cases in 2017, we estimated that 87,000 (95% CI, 81,800 to 92,200) were hospital-onset cases, 56,600 (95% CI, 53,200 to

59,900) were LTCF-onset cases, and 92,400 (95% CI, 86,900 to 97,900) were CO-HFCA cases. Note that the estimates of healthcare-associated CDI subgroups do not add up to the total estimate of healthcare-associated CDI due to rounding.

Discussion

Our analyses have the following additional limitations. We did not have information regarding the indication for repeat *C. difficile* testing, and it is possible that some recurrent *C. difficile* cases might represent tests for cure rather than a true recurrent episode. We did not adjust for the average number of inpatient-days per hospital in each surveillance site when estimating the 2011–2017 national burden of healthcare-associated CDI, as was previously done for the 2011 published CDI burden estimate.[1] This step was omitted because the annual inpatient-days data from the U.S. Health Resources and Services Administration's Area Health Resource File were not available for the entire period of interest. As a result, we did not use a model to generate the national CDI burden estimates as was done previously but, instead, used the actual EIP data extrapolated to the entire country, after accounting for the age, sex, and race of the U.S. population. However, our modified approach produced similar results compared to the previous method; using the updated 2011 NAAT usage rate of 55%, the previous method would have generated a 2011 national burden estimate of 456,400 (95% CI, 399,000-513,800) cases, which overlaps substantially with the confidence interval of our current estimate of 476,431 (95% CI, 419,934 to 532,929) cases. Although the EIP sites are geographically diverse with similar demographic characteristics as the U.S. population, they may not be fully representative of the entire country. In addition, the NAAT usage rate of EIP sites, which was used to estimate the national CDI incidence and burden, may not be representative of the national NAAT usage rate. However, we performed a sensitivity analysis to show how the 2017 national burden estimate changes according to NAAT use. Lastly, only a convenience sample of *C. difficile* isolates from each surveillance site were available for

strain typing; therefore, the results may not be representative of the strain distribution of all *C. difficile* isolates.

References

1. Lessa FC, Mu Y, Bamberg WM, et al. Burden of Clostridium difficile infection in the United States. N Engl J Med. 2015;372:825-34.

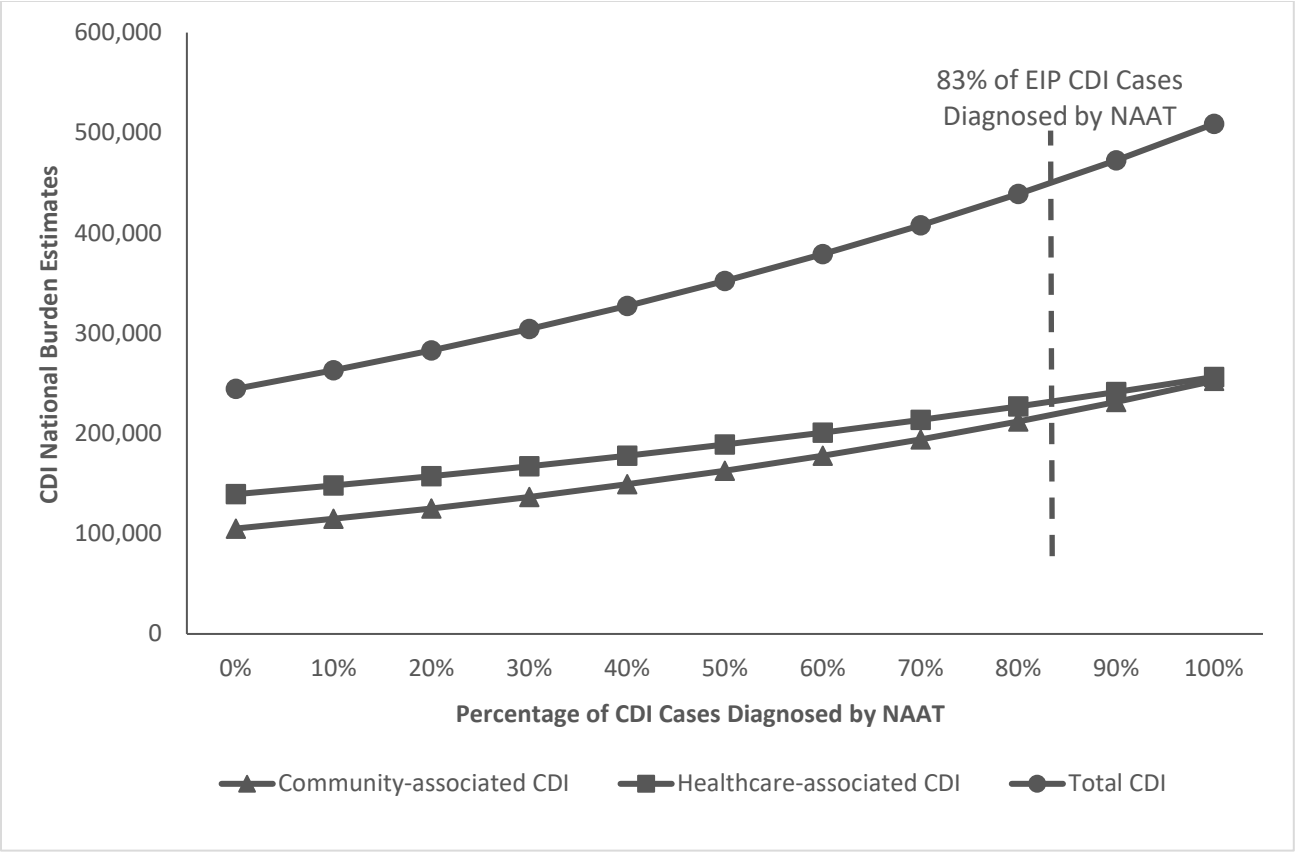

Figure S1. Sensitivity Analysis: *Clostridioides difficile* Infection National Burden Estimates by the Percentage of Cases Diagnosed by NAAT, 2017. Abbreviations: CDI, *Clostridioides difficile* infection; NAAT, nucleic acid amplification test.

Table S1. Distribution of Ribotypes Among Community-associated and Healthcare-associated *Clostridioides difficile* Isolates, 2012-2017.

2012				2013				2014			
CA (N=801)		HA (N=642)		CA (N=688)		HA (N=540)		CA (N=618)		HA (N=504)	
RT	No. (%)	RT	No. (%)	RT	No. (%)	RT	No. (%)	RT	No. (%)	RT	No. (%)
027	137 (17)	027	136 (21)	027	82 (12)	027	128 (24)	106	70 (11)	027	70 (14)
106	74 (9)	106	55 (9)	106	65 (9)	106	43 (8)	002	49 (8)	106	61 (12)
002	70 (9)	002	36 (6)	020	54 (8)	014	37 (7)	020	48 (8)	002	46 (9)
020	52 (6)	020	34 (5)	002	52 (8)	002	29 (5)	027	44 (7)	014	35 (7)
014	42 (5)	014	29 (5)	014	33 (5)	020	29 (5)	014	32 (5)	020	31 (6)
056	36 (4)	054	21 (3)	078	23 (3)	017	15 (3)	054	25 (4)	056	18 (4)
001_072	26 (3)	017	19 (3)	015	21 (3)	001_072	14 (3)	015	20 (3)	001_072	15 (3)
078	24 (3)	078	19 (3)	076	21 (3)	005	14 (3)	076	20 (3)	103	11 (2)
015	20 (3)	053	18 (3)	001_072	19 (3)	056	14 (3)	078	18 (3)	017	10 (2)
019	17 (2)	A12	17 (3)	056	18 (3)	015	11 (2)	005	17 (3)	054	10 (2)
054	16 (2)	046	15 (2)	A12	18 (3)	054	11 (2)	001_072	16 (3)	078	9 (2)
076	15 (2)	009	12 (2)	054	17 (2)	078	11 (2)	017	16 (3)	A05	9 (2)
046	14 (2)	015	12 (2)	010	14 (2)	A12	11 (2)	103	14 (2)	009	8 (2)
103	13 (2)	056	12 (2)	017	14 (2)	053	10 (2)	019	11 (2)	010	8 (2)
012	12 (2)	A05	12 (2)	005	13 (2)	012	9 (2)	A12	11 (2)	012	8 (2)
Others	233 (29)	Others	195 (30)	Others	224 (33)	Others	154 (29)	Others	207 (33)	Others	155 (31)
2015				2016				2017			
CA (N=614)		HA (N=538)		CA (N=460)		HA (N=509)		CA (N=495)		HA (N=555)	
RT	No. (%)	RT	No. (%)	RT	No. (%)	RT	No. (%)	RT	No. (%)	RT	No. (%)
106	58 (9)	027	102 (19)	106	62 (13)	027	79 (16)	106	60 (12)	027	81 (15)
027	52 (8)	106	48 (9)	027	42 (9)	106	56 (11)	002	48 (10)	106	54 (10)
014	46 (7)	002	40 (7)	002	33 (7)	014	35 (7)	020	32 (6)	002	38 (7)
020	40 (7)	014	36 (7)	014	27 (6)	002	30 (6)	027	28 (6)	014	37 (7)
002	35 (6)	020	30 (6)	020	20 (4)	020	30 (6)	014	26 (5)	076	26 (5)
015	21 (3)	015	18 (3)	056	18 (4)	015	18 (4)	054	16 (3)	020	24 (4)
054	20 (3)	001_072	17 (3)	019	16 (3)	056	14 (3)	076	15 (3)	054	20 (4)
005	19 (3)	056	17 (3)	054	16 (3)	054	13 (3)	019	13 (3)	015	19 (3)
056	18 (3)	017	15 (3)	015	14 (3)	078	12 (2)	015	12 (2)	056	18 (3)
046	17 (3)	005	14 (3)	087	12 (3)	017	11 (2)	017	11 (2)	078	15 (3)
017	16 (3)	A12	14 (3)	001_072	10 (2)	019	11 (2)	078	11 (2)	001_072	14 (3)
076	16 (3)	153_251	13 (2)	A12	9 (2)	076	11 (2)	001_072	10 (2)	046	13 (2)
001_072	14 (2)	054	12 (2)	005	8 (2)	005	10 (2)	005	10 (2)	A12	12 (2)
019	14 (2)	078	12 (2)	076	8 (2)	046	10 (2)	056	10 (2)	005	10 (2)
087	12 (2)	076	10 (2)	078	8 (2)	009	9 (2)	A05	9 (2)	A05	10 (2)
Others	216 (35)	Others	140 (26)	Others	157 (34)	Others	160 (31)	Others	184 (37)	Others	164 (30)

Abbreviations: CA, community-associated; HA, healthcare-associated; RT, ribotype. Isolates that could not be identified with the available reference library were assigned an internal CDC nomenclature beginning with the letter “A” (A12, A05).