Methods for cost-efficient, whole genome sequencing surveillance for enhanced detection of outbreaks in a hospital setting

1.1 Author names

Kady D. Waggle[1,2,4], *Marissa Pacey Griffith*[1,2], *Alecia B. Rokes*[5,6], *Vatsala Rangachar Srinivasa*[1,2,3], *Deena Ereifej*[1,2,3], *Rose Patrick*[1,2], *Hunter Coyle*[1,2], *Shurmin Chaudhary*[1,2], *Nathan J. Raabe*[1,2,3], *Alexander J. Sundermann*[1,2], *Vaughn S. Cooper*[5,6], *Lee H. Harrison*[1,2,3], *Lora Lee Pless*[1,2#]

ORCID iDs

Kady D. Waggle https://orcid.org/0009-0001-9237-9964

Marissa Pacey Griffith https://orcid.org/0000-0003-0677-8203

Alecia B. Rokes https://orcid.org/0000-0001-9496-0296

Vatsala Rangachar Srinivasa https://orcid.org/0000-0001-7376-8764

Deena Ereifej https://orcid.org/0009-0002-2984-3784

Rose Patrick https://orcid.org/0009-0003-3541-5513

Hunter Coyle https://orcid.org/0009-0006-0463-4168

Shurmin Chaudhary https://orcid.org/0009-0009-1640-7132

Nathan J. Raabe https://orcid.org/0000-0002-3258-9003

Alexander J. Sundermann https://orcid.org/0000-0002-2439-6287

Vaughn S. Cooper https://orcid.org/0000-0001-7726-0765

Lee H. Harrison https://orcid.org/0000-0002-3787-2705

Lora Lee Pless https://orcid.org/0000-0003-0163-3689

1.2 Affiliation(s)

[1]Microbial Genomic Epidemiology Laboratory, Center for Genomic Epidemiology, University of Pittsburgh, 3507 Victoria Street, BST-10 E1000-4A, Pittsburgh, Pennsylvania 15213, USA.

[2]Division of Infectious Diseases, University of Pittsburgh School of Medicine, 3550 Terrace Street, 818 Scaife Hall, Pittsburgh, Pennsylvania 15261, USA.

[3]Department of Epidemiology, School of Public Health, University of Pittsburgh, 130 De Soto Street, Pittsburgh, Pennsylvania 15261, USA

[4]Department of Infectious Diseases, School of Public Health, University of Pittsburgh, 130 De Soto Street, Pittsburgh, Pennsylvania 15261, USA

[5]Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[6]Center for Evolutionary Biology and Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.

1.3 Corresponding author and email address

#Lora Lee Pless

lora.pless@pitt.edu

University of Pittsburgh

Starzl Biomedical Science Tower

200 Lothrop St

Pittsburgh, PA 15213

1.4 Keywords

Healthcare-Associated Transmission; Bacteria, Antimicrobial Resistance, Whole Genome Sequencing; Laboratory Methods; Cost Analysis

2. Abstract

Introduction. Outbreaks of healthcare-associated infections (HAI) result in substantial patient morbidity and mortality; mitigation efforts by infection prevention teams have the potential to curb outbreaks and prevent transmission to additional patients. The incorporation of whole genome sequencing (WGS) surveillance of suspected high-risk pathogens often identifies outbreaks that are not detected by traditional infection prevention methods and provides evidence for transmission. Our approach to real-time WGS surveillance, the Enhanced Detection System for Healthcare-Associated Transmission (EDS-HAT), has 1) identified serious outbreaks that were otherwise undetected and 2) shown the potential to be cost saving because HAIs are expensive to treat and WGS has become relatively inexpensive.

Methods. We describe a cost-efficient method to perform WGS surveillance and data analysis of pathogens for hospitals that are interested in incorporating WGS surveillance. We provide an overview of the weekly workflow of EDS-HAT, discussing both the laboratory and bioinformatics methods utilized, as well as the costs associated with performing these methods.

Results. In an average week at our tertiary healthcare system, we sequenced 48 samples at a cost of less than $100 per sample, inclusive of laboratory reagents and staff salaries. The average turnaround time, from sample collection to data reporting to the infection prevention and control team, was ten days.

Conclusions. Our findings demonstrate that performing EDS-HAT in real-time can be both affordable and time-efficient. Providing such timely information to aid in outbreak investigations can identify transmission events sooner and thus increase patient safety.

3. Impact statement

Whole genome sequencing (WGS) surveillance to confirm or refute suspected outbreaks of potential healthcare-associated infections (HAI) is a highly effective approach for outbreak detection. Since November 2021, we have been conducting WGS surveillance in real-time through a program called the Enhanced Detection System for Hospital-Associated Transmission (EDS-HAT), to assist our hospital infection prevention and control (IP&C) team to identify and stop outbreaks. To our knowledge, our laboratory is the only group in the United States that has successfully implemented real-time WGS surveillance of multiple pathogens in the hospital setting. Our weekly workflow includes identifying HAI pathogens and performing WGS, followed by a variety of bioinformatic analyses that include species confirmation, determination of sequence type, and genetic relatedness comparisons. Based on this information, transmission clusters are identified, and the electronic health record is reviewed to determine probable transmission routes. Finally, IP&C implements appropriate interventions to mitigate the spread of infection. We detail the laboratory and analytical methods, along with the cost associated for laboratory materials and staff salary, for successful implementation of WGS surveillance in real-time establishing EDS-HAT as a unique and effective tool to detect HAI outbreaks.

4. Introduction

Healthcare-associated infections (HAIs) are a growing concern in hospital settings and can be associated with substantial morbidity and mortality. HAIs also impose a significant economic burden on healthcare systems, costing hospitals an estimated USD$ 9.6 billion per year (1,2). Whole genome sequencing (WGS) for HAI organisms can provide insight on the transmission dynamics in hospital settings (3). Historically, determining the degree of genomic variation between organisms was accomplished using pulsed field gel

electrophoresis (PFGE; 4). Given recent declines in costs and its many advantages, WGS has emerged as the leading method for determining genetic relatedness between clinical isolates (5). Reactive WGS is currently the most commonly used method to confirm or refute the presence of a suspected outbreak. This approach can result in a failure to detect important outbreaks for a variety of reasons, including outbreaks caused by common organisms, those not clustering on a single nursing unit, those consisting of a small number of patients, or those caused by an unsuspected or complex transmission route (6). Reactive WGS could also falsely identify an outbreak supported by epidemiological methods by clustering genetically distinct isolates (7). Furthermore, using WGS to obtain information about the entire genome provides the required data for determining organism phylogeny, detecting the presence of antimicrobial resistance (AMR) genes and mobile genetic elements, and identifying rare or novel genetic variants (7,8). The Microbial Genomic Epidemiology Laboratory (MiGEL) at the University of Pittsburgh developed the Enhanced Detection System for Healthcare-Associated Transmission (EDS-HAT) to identify outbreaks of HAIs in real-time using WGS surveillance methods in partnership with the UPMC IP&C team and the UPMC Clinical Laboratories. EDS-HAT has been operational in real-time at our institution for over two years (7,9–12). The barriers for most hospital systems for implementing proactive WGS are cost, lack of technical guidance, and inadequate infrastructure.

In this paper, we describe our methods for WGS, the bioinformatics workflow, and provide a cost estimate of WGS surveillance, with the goal of providing guidance to hospitals who wish to implement WGS surveillance.

5. Theory and Implementation

Study Setting.

MiGEL is a non-Clinical Laboratory Improvement Amendments (CLIA) certified research laboratory located on the University of Pittsburgh main campus, in Pittsburgh PA, USA. EDS-HAT was developed and is currently implemented in real time at MiGEL in coordination with the University of Pittsburgh, UPMC, the UPMC Clinical Laboratory Building (CLB), the UPMC IP&C team, and Carnegie Melon University (CMU). UPMC Presbyterian is an adult tertiary acute care hospital with 758 total beds, 134 critical care beds, and over 400 annual solid organ transplants. The main campus is UPMC Presbyterian Hospital, also located in the Oakland neighborhood of Pittsburgh, PA, adjacent to the University of Pittsburgh main campus and CMU. The University of Pittsburgh Institutional Review Board provided ethics approval for EDS-HAT (Protocol: STUDY21040126).

Clinical Specimen Collection.

Isolate Inclusion Criteria. A list of select, high-concern bacterial pathogens was generated twice per week using Theradoc (5.4.0.HF1.102, Pittsburgh, PA; Fig 1A). Pathogens of interest include: extended-spectrum B-lactamase-producing (ESBL) *Escherichia coli,* ESBL *Enterobacter* species, *Acinetobacter* species, *Pseudomonas* species, *Klebsiella* species, *Stenotrophomonas* species, *Serratia* species, *Burkholderia* species, *Providencia* species, *Proteus* species, *Citrobacter* species, vancomycin-resistant *Enterococcus* (VRE), methicillin-resistant *Staphylococcus aureus* (MRSA), and *Clostridioides difficile*. EDS-HAT isolate inclusion criteria included patients who have been in the hospital for three or more days and/or had a previous hospital exposure during the 30-days prior to culture (7). For this study, we described the samples and methods utilized during a one-year period of time of performing real-time EDS-HAT (March 2022-March 2023).

Isolate Collection. Bacterial samples were collected by MiGEL twice per week at the UPMC CLB from pure cultures isolated from clinical specimens that were prompted by clinician suspicion of infection (Fig 1.B1). To ensure availability of the isolates for sequencing, CLB technologists subcultured all gram-negative isolates from aerobic bacterial cultures onto nutrient agar slants. We identified the gram-negative isolate slants of interest from the CLB, and then isolates of interest were subcultured to Trypticase Soy Agar with 5% sheep blood (BAP) plates (BD, Franklin Lakes, NJ), transported to MiGEL, and incubated at 37°C overnight in the presence of 5% CO_2. The gram-positive isolates from the CLB were transferred from one BAP to another and then transported and incubated at MiGEL following the same procedure. The next day, sample information was imported into the MiGEL database, and a de-identified specimen ID was generated for each sample.

***Clostridioides difficile* collection and culture.** In contrast to the methods described for the above organisms that were isolated as pure cultures, we collected and cultured clinical stool specimens that tested positive for *C. difficile* by culture-independent diagnostic testing (13). This organism is anaerobic; thus, we performed the following protocol to isolate this organism directly from the clinical stool specimens. In a biosafety cabinet, each stool sample was subcultured onto cycloserine-cefoxitin-mannitol-agar with taurocholate and lysozyme (CCMA-TAL) plates to select for *C. difficile* growth. Plates were transferred into a Coy anaerobic chamber (Coy Laboratory Products, Grass Lake, MI) and incubated at 37°C for 48 hours. Colonies of *C. difficile* were passaged to a second CCMA plate and incubated at 37°C in the anaerobic chamber for an additional 24-48 hours. Isolates were confirmed as *C. difficile* by testing for the production of L-Proline aminopeptidase using a PRO Disc test (Remel, San Diego, CA; Fig1.B2).

Sample Preparation and DNA Extraction

To begin sample preparation for WGS, microcentrifuge tubes containing 750 µL phosphate buffered saline (PBS) were inoculated with a quarter-portion of a 10 µL loop of bacteria (a half-portion was used for *C. difficile*) from the BAP or CCMA plate. The tubes were centrifuged at 6.0 × *g* for 10 minutes to generate a pellet, and the supernatant was removed using a P1000 pipette (Fig 1C). For samples not proceeding immediately to extractions, the pellets were stored at –20°C. Isolate stocks for long-term storage for all bacterial isolates (including *C. difficile*) were prepared by inoculating a 10 µL loop of bacteria into cryovials containing 1 mL of nutrient broth mixed with 20% glycerol and then stored at –80°C.

The bacterial pellets were re-suspended in 500 µL PBS prior to extraction. DNA was extracted using the MagMAX DNA Multi-Sample Ultra 2.0 extraction kit on the King Fisher Apex (Thermo Fisher Scientific, Waltham, MA) per manufacturer's instructions (Fig 1D). Briefly, this procedure isolates and purifies nucleic acids using magnetic bead-based technology. DNA was eluted in 100 µL of elution buffer supplied by the kit and then quantified using a Qubit broad range dsDNA kit (Life Technologies, Carlsbad, CA). Samples with a concentration ≥3.5ng/µL were considered for WGS. For samples that did not meet this criterion, DNA was extracted again.

WGS Library Preparation

DNA libraries were prepared on an epMotion 5075t (Eppendorf, Hamburg, Germany) liquid handler using a DNA Prep (M) Tagmentation kit (Illumina, San Diego, CA), utilizing half-volume reactions for BLT/TB1 and EPM reagents (Fig 1E). A unique 10-mer index adapter sequence was ligated to each sample (IDT, Coralville, IA). Briefly, the DNA Prep protocol uses bead-linked transposomes to tagment and amplify the adapter-tagged DNA segments. Eight individual libraries were pooled together by combining 5 µL per library into a single tube. Pooled libraries were quantified using a Qubit high sensitivity dsDNA kit. The library

pool was normalized to 4 nM with resuspension buffer (RSB). Additional pools were combined using equimolar concentration into a single pool. The distribution of the fragment sizes for the sequencing pool was assessed using an Agilent Tapestation D5000 screen tape and reagents per manufacturer's protocol (Agilent Technologies, Santa Clara, CA).

Whole Genome Sequencing

DNA libraries were sequenced weekly using an Illumina MiSeq (≤32 samples on a v3, 600 cycle kit) or NextSeq550 (>32 samples on a v2.5, 300 cycle kit) platform (Fig 1F). The DNA library was denatured using 0.2N NaOH and spiked with 1% PhiX to increase diversity on the flow cell. The DNA library was diluted, using the average library length, to the final loading concentration of 16 pM for the MiSeq or 1.5-1.6 pM for the NextSeq550. A commercial lab was used for sequencing in rare cases where personnel were unavailable for in-house sequencing. For these occasions, DNA was extracted and sent for same-day delivery using a local medical courier service, followed by library preparation and sequencing at the commercial lab. For sequencing using any of the options described, DNA extraction and library preparation were performed using automated methods; however, it was possible to perform all steps manually.

Bioinformatics and Data Analysis

Sequencing Data Quality Control (QC). We have developed a real-time bioinformatics pipeline that is executed once per week as a single command written in the programming language Python. This customized pipeline is one of four commands that are executed on the new samples, as well as previously sequenced genomes. These commands include: 1) data download from the BaseSpace Sequence Hub v7.18.0 (Illumina); 2) sample demultiplexing; 3) file transfer into individual directories; and 4) real-time bioinformatics pipeline execution. Specifically, we begin by converting and demultiplexing the base call files using Illumina bcl2fastq (v2.20) software. WGS reads were assembled using Unicycler

v0.5.0 and then annotated using Prokka v1.14 (14). Multilocus sequence types (STs) were assigned using PubMLST typing schemes for all organisms with the exception of *Serratia spp.* and *Providencia spp.*, which do not have ST schemes (mlst v2.11; 15). Reads were mapped using Kraken2 with the Kraken standard database to determine the most prevalent species (16). Isolates passed QC if 1) the most prevalent species by Kraken2 was the expected organism, 2) the assembly length was within 20% of the expected genome length, 3) the assembly was ≤ 350 contigs, and 4) there was at least 35× depth (Fig 1G).

Determining Infection Clusters and Downstream Applications. Pairwise single nucleotide polymorphisms (SNPs) between all real-time EDS-HAT isolates of the same species were determined using one of two programs (Fig 1H). i) Pairwise core genome SNPs (cgSNPs) were determined using Snippy v4.3.0, a reference-based method, for isolates with the same ST (17). SNP distances were calculated from the core alignment using 'snp-dists' (18). ii) SKA v1.0, a reference-free method, was used to calculate SNP distances using the 'ska distance' command for isolates of the same species (19). We selected the minimum SNP distance for each pairwise comparison quantified by Snippy or SKA to determine clusters of genetically similar isolates. These genetically similar clusters were defined using hierarchical clustering with average linkage and a cutoff of ≤15 SNPs for all species except *C. difficile*, for which a cutoff of ≤2 SNPs was used (Fig 1H; https://scipy.org/). The electronic health records for patients with genetically similar isolates were reviewed to determine potential epidemiological links. This information was then communicated to the hospital IP&C team, which implemented targeted mitigation measures when possible. See Supplementary Figure 1 for real-time bioinformatics pipeline.

Cost Analysis

A cost estimate for EDS-HAT real-time genomic surveillance methods was determined in 2023 US dollars and included the cost of personnel, reagents, and supplies, and was

analyzed comparatively for each sequencing platform used by MiGEL (Supplementary Table 1). Non-fringe personnel costs (salary) were determined using the average pay scale of Laboratory Technician III (90% effort) and Bioinformatics Research Analyst II (50% effort) positions at UPMC, Pittsburgh, PA in 2023. Reagent and supply costs were determined using manufacturer pricing (data accessed: December 1, 2023).

6. Results

Weekly Sequencing Runs

From March 2022 to March 2023, MiGEL collected and sequenced 2,070 bacterial isolates (with an average of 48 isolates per week) as part of real-time EDS-HAT. The most commonly sequenced organism was *Pseudomonas aeruginosa* (617 genomes) and the least sequenced was *Burkholderia sp.* (11 genomes; Table 1). To determine which platform was best suited for weekly sequencing, we considered the count of organisms and the average genome size. The weekly average genome size was 4.85 Mbp, roughly equating to a maximum of 37 or 98 samples on the MiSeq or NextSeq flow cells, respectively, to achieve a minimum target of 80× coverage. When sequencing pools of organisms with smaller average genome sizes, a greater number of isolates could be appropriately accommodated per flow cell without compromising run quality or per organism coverage data (Figure 2). Based on the MiGEL average genome size and to maximize cost efficiency, runs containing a range of 32-40 samples were sequenced on the MiSeq platform, and runs containing > 40 samples were sequenced on the NextSeq550 platform. During this study, 17 runs were performed on the MiSeq platform, and 28 runs were performed on the NextSeq550 platform. For an average run of 48 samples on the NextSeq550, MiGEL observed a maximum output of 52 Gb of data and an average of 100 million reads (Supplementary Table 2). The average turnaround time to complete the EDS-HAT workflow from sample collection by MiGEL to bioinformatic analysis using either platform for sequencing was approximately 10 days, with

an average WGS instrument run time of 25 hours (Supplementary Table 2). The turnaround time for using a commercial lab was approximately two weeks or less.

Cost Analysis

The cost to run real-time EDS-HAT weekly was categorized into sample processing, DNA extraction and quantification, library preparation, and flow cell cost (Table 2). The lowest cost per sample ($48) was achieved when the maximum number of samples (n=96) were sequenced using the NextSeq550 platform. Costs ranged from $48 to $83 per sample, dependent on platform and sample counts. There was an inverse relationship between the number of samples sequenced and flow cell cost as per sample costs significantly decreased when a greater number of samples were multiplexed on the appropriate flow cell. The gray dashed line in Figure 3 shows the cost to sequence 40 samples using all sequencing options, with the MiSeq having the lowest cost and commercial lab having the highest. The estimated weekly cost of personnel, based on the pre-tax salaries for one lab technician and one bioinformatician based on percent efforts, totaled $1,077. When all costs were considered, the cost to run EDS-HAT on an average week totaled $4,293 (min $3,626 – max $4,758) or $223,236 per year (min $188,552 – max $247,416).

7. Discussion

In this study, we detailed an efficient laboratory workflow, our approach for bioinformatics analyses, and estimated the cost associated with implementing real-time WGS surveillance for pathogenic bacteria in a hospital system that was designed to detect otherwise unrecognized hospital outbreaks. EDS-HAT began in 2016 as a retrospective study (7) and, once we demonstrated the superiority of the system over traditional approaches, transitioned in November 2021 to a real-time workflow, subsequent bioinformatic analyses, and reporting of results to the hospital IP&C team. To our knowledge, UPMC is the only hospital system in

the US that is actively performing prospective WGS surveillance methods for multiple pathogens in real time. By doing so, our hospital system has dramatically changed the way outbreaks are being detected.

We provide details about our methods for a one-year timeframe, after our initial optimization period, beginning in March 2022. We determined that the per sample cost for WGS ranged from $48 to $83, with an average of $65. Furthermore, with the addition of staff salaries, the mean weekly cost for an average week of real-time sequencing was $4,293 (*N*=48 samples). This cost of real-time WGS is lower compared to prior studies (20,21). Our lower cost was achieved, in part, by increasing sample counts per flow cell while utilizing the appropriate instrument, using half-volumes of reagents for some stages of library preparation, and an overall decline in sequencing costs.

With our quick turnaround time from the day the sample is collected by MiGEL, we have identified ongoing outbreaks that serve as a guide for the IP&C team to implement infection prevention interventions. We previously showed that there was an estimated cost savings of $96,204–$346,266 per year by implementing a real-time WGS surveillance system, which was based on an average cost of $86 for sample preparation and sequencing (adjusted for inflation to 2023 USD; 22). We optimized the average per isolate cost of sample preparation and sequencing from $74 on the MiSeq platform (SD, $3.30) to $60 on the NextSeq platform (SD, $6.40), achieving even greater cost savings per year. 2/16/2024 1:00:00 PMMore importantly, stopping transmission events quickly at the first sign of an outbreak cluster has the potential to reduce further spread of the infection and thus reduce patient morbidity and mortality.

The foremost concern of hospital systems with implementing programs like EDS-HAT is cost, with the vast majority of interested parties assuming that there is a large expense associated with real-time sequencing surveillance. While this was true years ago, the cost of

sequencing has decreased over time (23). In addition, the laboratory and bioinformatics methods have become more streamlined, automatable, and efficient. Furthermore, the cost of treating preventable hospital infections is high, and, in fact, EDS-HAT has been shown to be cost saving. Taken together, these facts and the evidence that this approach can identify important, otherwise-undetected outbreaks, suggest that WGS surveillance should eventually become standard practice in hospitals.

To accompany our methods, we computed the cost per sample, which accounts for staff salaries, to be $91 on average (range $62-$119), and is specific for the greater Pittsburgh region in Pennsylvania, USA and is likely to be different at other locations. This fact is summarized by Price and colleagues, who find the cost to perform WGS varies by country and city (20). For example, Price (20) converted the cost per sample from prior studies to 2023 USD and showed the per sample cost of sequencing ranged from approximately $72-$470 for the US and Italy, respectively. In this study, we determined our average per sample cost (without considering staff salaries, for comparison) was $65 per sample. The primary factor in determining this cost estimate was sample count per run and average organism genome size. For reference, we provide the maximum number of samples that can be sequenced on either MiSeq or NextSeq platforms by organism, considering genome size, along with the average genome size sequenced over one year by MiGEL (Figure 2). In addition, we show in Figure 3 that a sample count of 40 is an appropriate cutoff to decide which machine to use for sample sequencing, while maintaining sufficient genome coverage. Generally, we find sequencing more samples at a time reduced the cost of sequencing per sample, with the exception of utilizing a commercial lab. While the commercial lab offered a discounted price once the sample count reached 48, we find the fixed price was overall more costly than performing in-house sequencing. Furthermore, we find a decrease in sequencing costs over time. MiGEL estimated a $72 average per sample cost in 2021, which we show is

lower in cost by $7 during our study period (March 8, 2022 to March 9, 2023; average cost is $65).

We note limitations with this study. First, the costs for reagents and supplies presented in this manuscript represent discounted pricing provided to our university from some manufacturers. Other institutions may have different discounted pricing or pay manufacturers rates, which will alter the costs described in our methods. Second, MiGEL benefits from the use of robotic instruments for nucleic acid extractions and library preparation, which can help save time and decrease pipetting errors on the bench. Some institutions may not have such instruments available and will need to accommodate the laboratory methods we described accordingly; however, we do not think this represents a significant detriment to the process. Third, we only considered Illumina-based technology for this study. Other short-read sequencing technologies or long-read sequencing were not assessed. Fourth, we have demonstrated the cost-efficiency at an academic, tertiary hospital system. These estimates are likely not reflective of a healthcare system located at a smaller locale.

In conclusion, we have shown that a real-time WGS surveillance program is both feasible and affordable. Healthcare institutions wishing to do the same could potentially discover outbreaks that would otherwise be missed. Further adoption of this approach has the potential to significantly enhance patient safety.

8. Tables

Table 1.

EDS-HAT Bacterial genome sizes and number of each organism that has been sequenced by MiGEL (March 8, 2022 to March 9, 2023).

Organism	Genome Size (Mb)	Count Sequenced by MiGEL
Acinetobacter species	3.9	58
Burkholderia species	7.2	11
Clostridioides difficile	4.2	67
Citrobacter species	4.7	46
Enterobacter cloacae	5.3	51
Escherichia coli	5.3	152
Klebsiella oxytoca	5.9	28
Klebsiella pneumoniae	5.3	115
MRSA	2.9	306
Providencia species	4.5	28
Pseudomonas species	6.6	643
Serratia species	5.2	154
Stenotrophomonas species	4.8	112
VRE	2.9	127
Proteus species	4.06	269

Table 2. Cost Estimates

Method	MiSeq cost per sample	NextSeq cost per sample	Commercial Sequencing Cost

	(32 samples)	(48 samples)	(96 samples)	(any # samples)
Sample Processing	$2.06	$2.06	$2.06	$2.06
DNA Extraction & Quantification	$7.15	$7.15	$7.15	$7.15
Library Preparation	$25.90	$24.96	$24.02	-
Sequencing Kit (Flow Cell)	$49.16	$37.29	$18.65	$75
TOTAL COST PER SAMPLE	$84.27	$71.46	$51.88	$84.21

9. Author statements

9.1 Conflicts of interest

The authors declare that there are no conflicts of interest, including financial interests, activities, relationships, and affiliations.

9.2 Funding information

This work was supported by the National Institutes of Health (grant numbers R01AI127472 and R21AI109459).

9.3 Ethical approval

The University of Pittsburgh institutional review board provided ethics approval for this study.

9.4 Acknowledgements

The authors would like to thank SeqCenter for their assistance with WGS. We thank the leaders and staff of the UPMC Clinical Laboratories, especially Tung Phan, MD, PhD, D(ABMM), and Hannah Creager PhD, D(ABMM) and all members of the UPMC

Presbyterian/Shadyside Infection Prevention & Control Team, especially Graham Snyder, MD, Ashley Ayres, MBA, CIC for their continued support. This publication made use of the PubMLST website (https://pubmlst.org/) developed by Keith Jolley (Jolley & Maiden 2010, BMC Bioinformatics, 11:595) and sited at the University of Oxford. The development of that website was funded by the Wellcome Trust.

10. References

1. Scott RD. The Direct medical costs of healthcare-associated infections in U.S. hospitals and the benefits of prevention [Internet]. 2009. 16 p. Available from: https://stacks.cdc.gov/view/cdc/11550

2. Alioth Finance. $6,500,000,000 in 2007 → 2023 | Inflation Calculator [Internet]. 2023. Available from: https://www.officialdata.org/us/inflation/2007?amount=6500000000

3. Mustapha MM, Srinivasa VR, Griffith MP, Cho ST, Evans DR, Waggle K, et al. Genomic Diversity of Hospital-Acquired Infections Revealed through Prospective Whole-Genome Sequencing-Based Surveillance. mSystems. 2022 Jun 28;7(3):e0138421.

4. Neoh HM, Tan XE, Sapri HF, Tan TL. Pulsed-field gel electrophoresis (PFGE): A review of the “gold standard” for bacteria typing and current alternatives. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2019 Oct;74:103935.

5. Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, et al. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. Clin Microbiol Rev. 2017 Oct;30(4):1015–63.

6. Sundermann AJ, Babiker A, Marsh JW, Shutt KA, Mustapha MM, Pasculle AW, et al. Outbreak of Vancomycin-resistant Enterococcus faecium in Interventional Radiology: Detection Through Whole-genome Sequencing-based Surveillance. Clin Infect Dis Off Publ Infect Dis Soc Am. 2020 May 23;70(11):2336–43.

7. Sundermann AJ, Chen J, Kumar P, Ayres AM, Cho ST, Ezeonwuka C, et al. Whole-Genome Sequencing Surveillance and Machine Learning of the Electronic Health Record for Enhanced Healthcare Outbreak Detection. Clin Infect Dis. 2022 Aug 31;75(3):476–82.

8. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet. 2012 Sep;13(9):601–12.

9. Sundermann AJ, Rangachar Srinivasa V, Mills EG, Griffith MP, Waggle KD, Ayres AM, et al. Two Artificial Tears Outbreak-Associated Cases of Extensively Drug-Resistant Pseudomonas aeruginosa Detected Through Whole Genome Sequencing-Based Surveillance. J Infect Dis. 2023 Sep 13;jiad318.

10. Raabe NJ, Valek AL, Griffith MP, Mills E, Waggle K, Srinivasa VR, et al. Genomic Epidemiologic Investigation of a Multispecies Hospital Outbreak of NDM-5-Producing Enterobacterales Infections. MedRxiv Prepr Serv Health Sci. 2023 Sep 1;2023.08.31.23294545.

11. Sundermann AJ, Griffith M, Rangachar Srinivasa V, Ereifej D, Waggle K, Van Tyne D, et al. Environmental contamination of postmortem blood cultures detected by whole-genome sequencing surveillance. Infect Control Hosp Epidemiol. 2023 Aug 24;1–2.

12. Sundermann AJ, Chen J, Miller JK, Saul MI, Shutt KA, Griffith MP, et al. Outbreak of Pseudomonas aeruginosa Infections from a Contaminated Gastroscope Detected by Whole Genome Sequencing Surveillance. Clin Infect Dis Off Publ Infect Dis Soc Am. 2021 Aug 2;73(3):e638–42.

13. Crobach MJT, Planche T, Eckert C, Barbut F, Terveer EM, Dekkers OM, et al. European Society of Clinical Microbiology and Infectious Diseases: update of the diagnostic guidance document for Clostridium difficile infection. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. 2016 Aug;22 Suppl 4:S63-81.

14. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl. 2014 Jul 15;30(14):2068–9.

15. Seemann T. mlst [Internet]. Available from: https://github.com/tseemann/mlst

16. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019 Nov 28;20(1):257.

17. Seemann T. Snippy [Internet]. 2023. Available from: https://github.com/tseemann/snippy

18. Seemann T. snp-dists [Internet]. 2023. Available from: https://github.com/tseemann/snp-dists

19. Harris SR. SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology [Internet]. Genomics; 2018 Oct [cited 2023 Oct 4]. Available from: http://biorxiv.org/lookup/doi/10.1101/453142

20. Price V, Ngwira LG, Lewis JM, Baker KS, Peacock SJ, Jauneikaite E, et al. A systematic review of economic evaluations of whole-genome sequencing for the surveillance of bacterial pathogens. Microb Genomics. 2023 Feb;9(2):mgen000947.

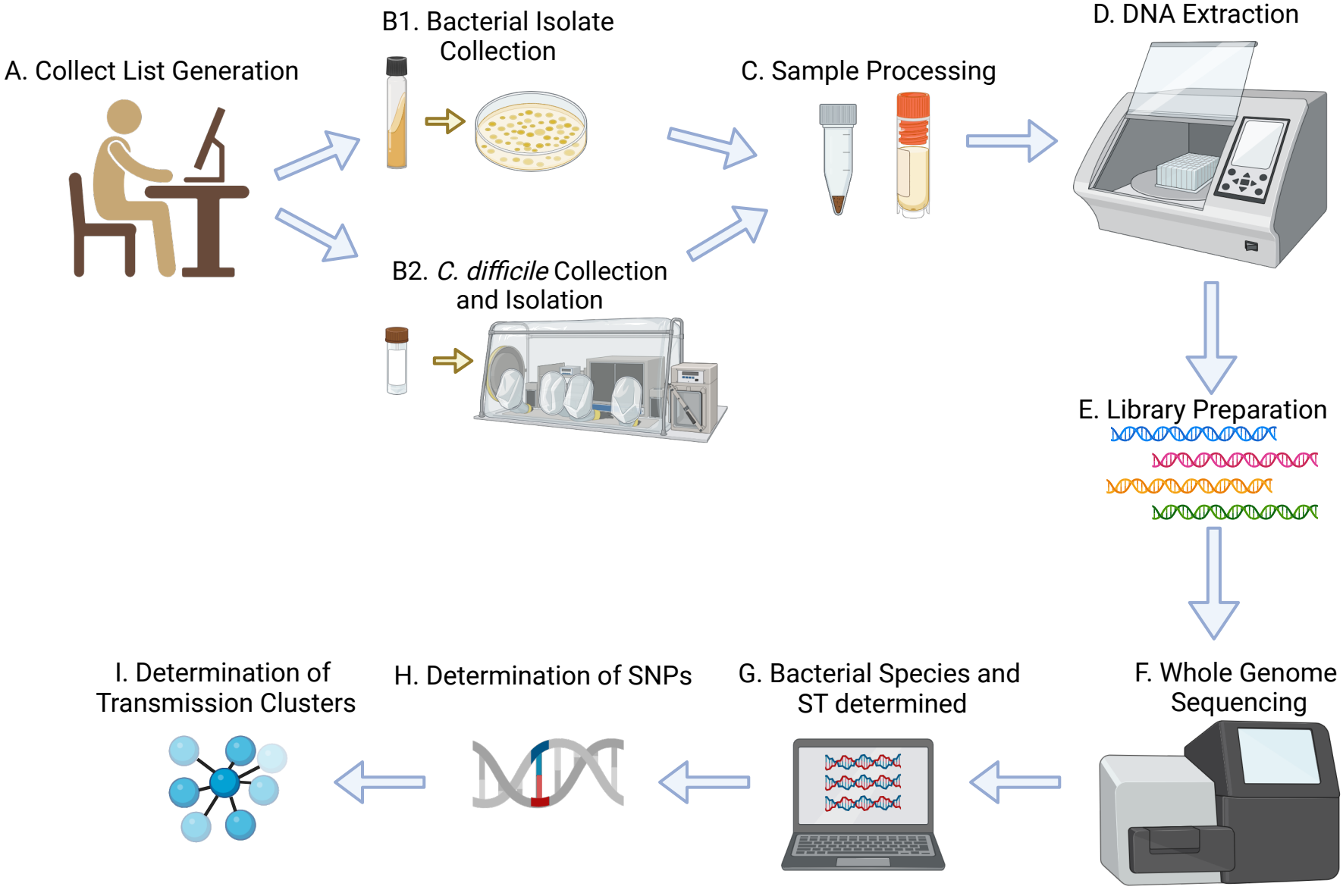
21. Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, et al. World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. PLoS Med. 2015 Dec;12(12):e1001923.

22. Kumar P, Sundermann AJ, Martin EM, Snyder GM, Marsh JW, Harrison LH, et al. Method for Economic Evaluation of Bacterial Whole Genome Sequencing Surveillance

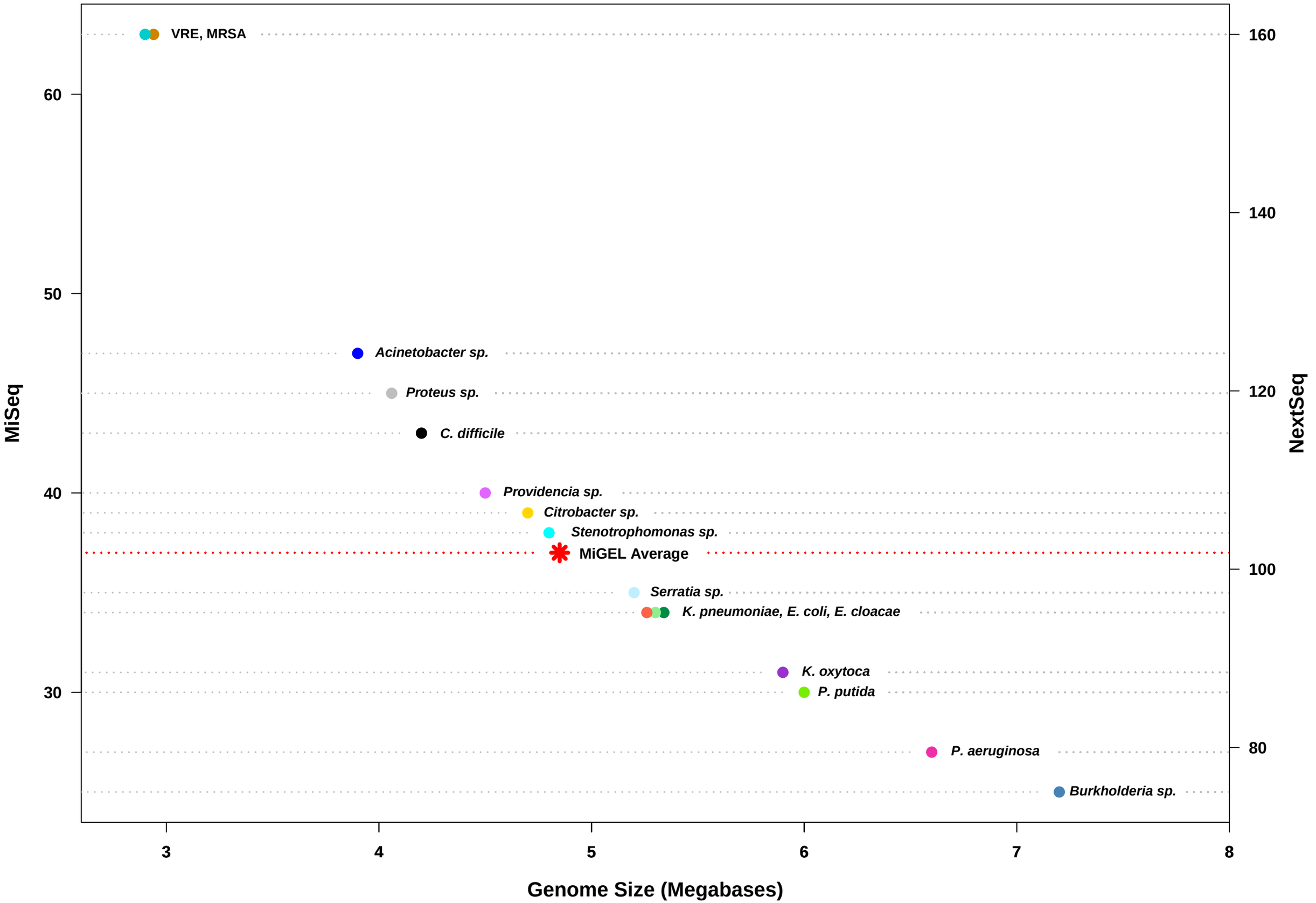
451 Compared to Standard of Care in Detecting Hospital Outbreaks. Clin Infect Dis. 2021 Jul
452 1;73(1):e9–18.

453 23. Wetterstrand K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing
454 Program (GSP) Available at: www.genome.gov/sequencingcostsdata.

455


A. Collect List Generation
B1. Bacterial Isolate Collection
B2. C. difficile Collection and Isolation
C. Sample Processing
D. DNA Extraction
E. Library Preparation
F. Whole Genome Sequencing
G. Bacterial Species and ST determined
H. Determination of SNPs
I. Determination of Transmission Clusters

Maximum Sample Counts



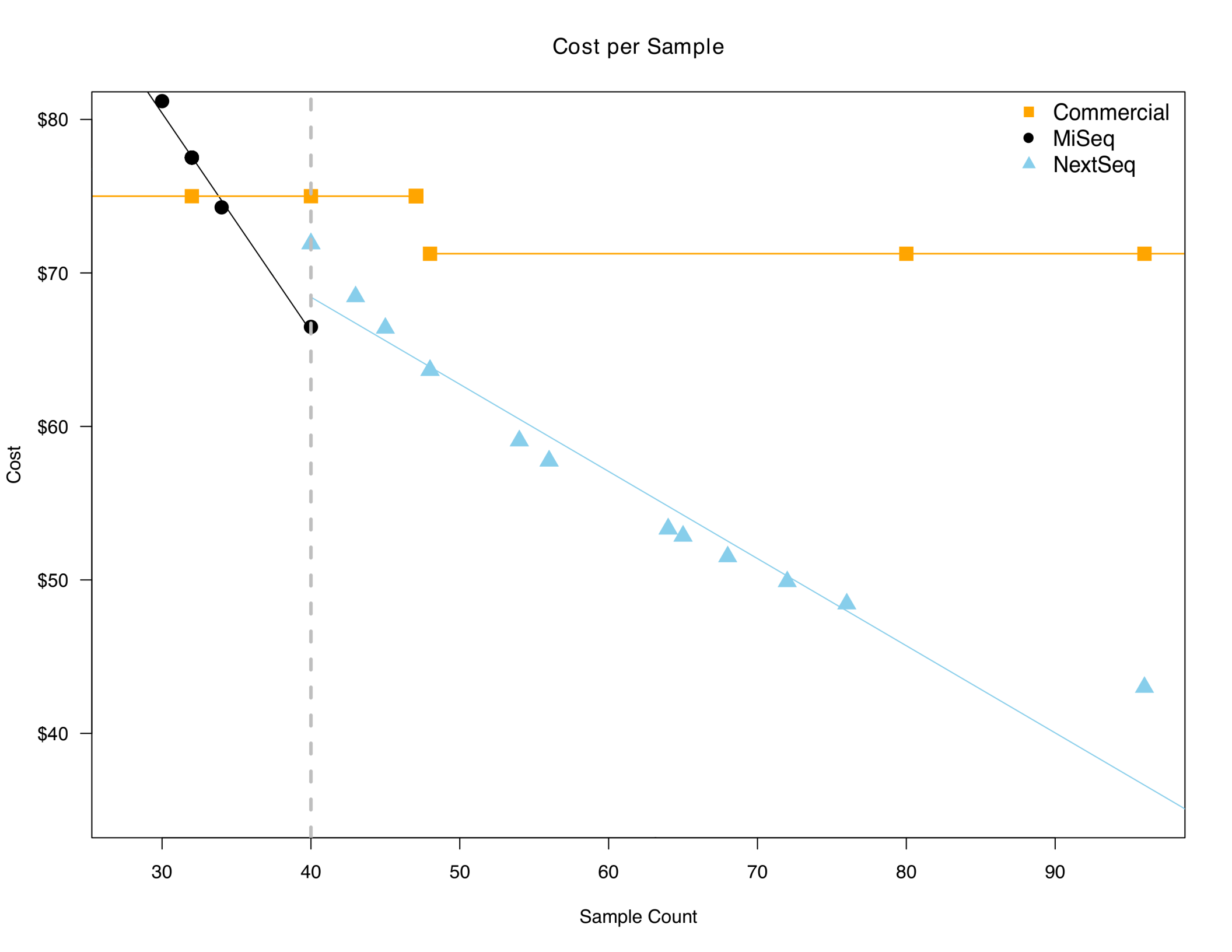

Cost per Sample
Commercial
MiSeq
NextSeq
$80
$70
$60
$50
$40
Cost
30
40
50
60
70
80
90
Sample Count

Figure 1. EDS-HAT real-time genomic surveillance methods. A) Biweekly collect list generation for all EDS-HAT organisms of interest obtained from patients admitted to the hospital for ≥ 3 days or had a previous hospital exposure in the prior 30 days, B1) Bacterial isolate collection from the UPMC Clinical Microbiology Laboratory, B2) Clinical stool specimen collection and *C. difficile* isolation using a Coy anaerobic chamber, C) Processing samples into pellets and glycerol stocks, D) DNA extraction, E) Library preparation using Eppendorf epMotion 5075, F) WGS using MiSeq or NextSeq550, G) Bioinformatic analysis to determine bacterial species and sequence type (ST), H) Determination of SNPs between isolates, and I) Determination of transmission clusters. The average turnaround time from MiGEL sample collection to determination of transmission clusters was 10 days.

Figure 2. Maximum number of genomes that can be sequenced on the MiSeq v3 600 cycle flow cell and NextSeq550 v2.5 300 cycle flow cell based on size (Mb). Sample counts were calculated using Illumina coverage calculator based on 80× coverage criteria. Genome size of an average run by MiGEL is shown in comparison to individual organism sizes (red star).

Figure 3. Whole genome sequencing cost per sample comparison between MiSeq, NextSeq550, and commercial laboratory. Throughput cutoff between platforms is shown at *N*=40 samples (gray dashed line). The commercial lab used by MiGEL offers a discount of 5% for orders ≥ 48 samples (as of December 2023). Data points represent instances of cost by sample count and lines of best fit are shown for each sequencing method.

Supplementary Figure 1. EDS-HAT bioinformatics pipeline.
Note, ST = sequence type; SNP = single nucleotide polymorphism; EHR = electronic health record).