

## **Appendix C: Data Sources and Linking Process**

### **Arrest**

*Data Sources.* Washington State Patrol (WSP) data is a statewide database of all arrests that occur in WA state. WSP data was sent to study staff and linked to HiFi patients. Administrative Office of the Courts (AOC) data is an electronic statewide database that includes information from cases files in most state courts. This is a comprehensive statewide database on all civil and criminal convictions in Washington State. The Seattle Municipal Court does not report infraction data to AOC, although they do share information with WSP.

*Linkage.* We were provided with all WSP arrest records for any individual in Washington State from 1974 through June 30, 2019. Limiting data to 2011 to capture arrests in the 5 years prior to HiFi enrollment, we used Link King to identify arrest records among HiFi patients matching on first name, last name, middle name, date of birth, sex and race. Link King software was developed to support linking of administrative datasets in Washington State and first uses deterministic matching followed by probabilistic matching. Using the combination of deterministic and probabilistic linkage, Link King categorizes links by level of certainty where Level 1 is ‘highest possible’, Level 2 is ‘very high’, Level 3 is ‘high’, Level 4 is ‘moderate to high’, Level 5 is ‘probabilistic probable twins’ and Level 6 is ‘probabilistic maybe.’ We accepted any match with a Level 1-3 classification and reviewed by hand any matches made with Level 4-6 certainty. Typically, 90% or more of matches are classified at certainty Levels 1 or 2.<sup>44</sup> For AOC data, we provided AOC data managers with HiFi patient names, date of birth (DOB), sex, address and derived driver’s license number for a manual matching process. Following linkage, data for HiFi patients was sent to study staff where two trained coders conducted a manual review to exclude non-criminal traffic violations, juvenile offenses and probable cause arrests that resulted in a subsequent arrest to prevent double-counting arrest charges. Remaining arrests and probable cause arrests were included. AOC and WSP data were merged to identify any arrest for HiFi patients prior to enrollment and in the 12 months following enrollment.

### **Subsequent Injury**

*Data Sources.* To identify subsequent injuries resulting in hospitalization we used the Comprehensive Hospital Abstract Reporting System (CHARS), which is a database covers discharges from all acute care hospitals in Washington State. It provides ICD coded data on diagnoses and procedures, as well as demographic information, treatment information, and hospital disposition. For injuries resulting in death, we used the Washington State vital statistics files which include name, sex, race, date of birth, date of death, and cause of death.

As there is not yet a comprehensive dataset with all ED visits that occur in Washington State, we combined a series of four data sources to ascertain subsequent injuries requiring ED treatment. Our first data source for subsequent injury was the Emergency Department Information Exchange (EDIE), a program run by Collective Medical Technologies, Inc. which allows hospitals to voluntarily participate in EDIE by reporting ED visits. In turn, when a patient presents to an ED at participating hospitals, physicians can see prior ED visits in the past 6 months with short visit summaries, although all recorded ED visits are retained by Collective Medical Technologies. Hospitals began participating in 2009 in Washington State, and as of 2013, EDIE included all major EDs in the state. EDIE allows the creation of patient groups within their interface to facilitate patient follow up. As patients were enrolled in HiFi, a Research Assistant would add them to the HiFi patient group in EDIE. Collective Medical Technologies built a custom report which queried any ED visit for all HiFi patients in Washington without any date restrictions.

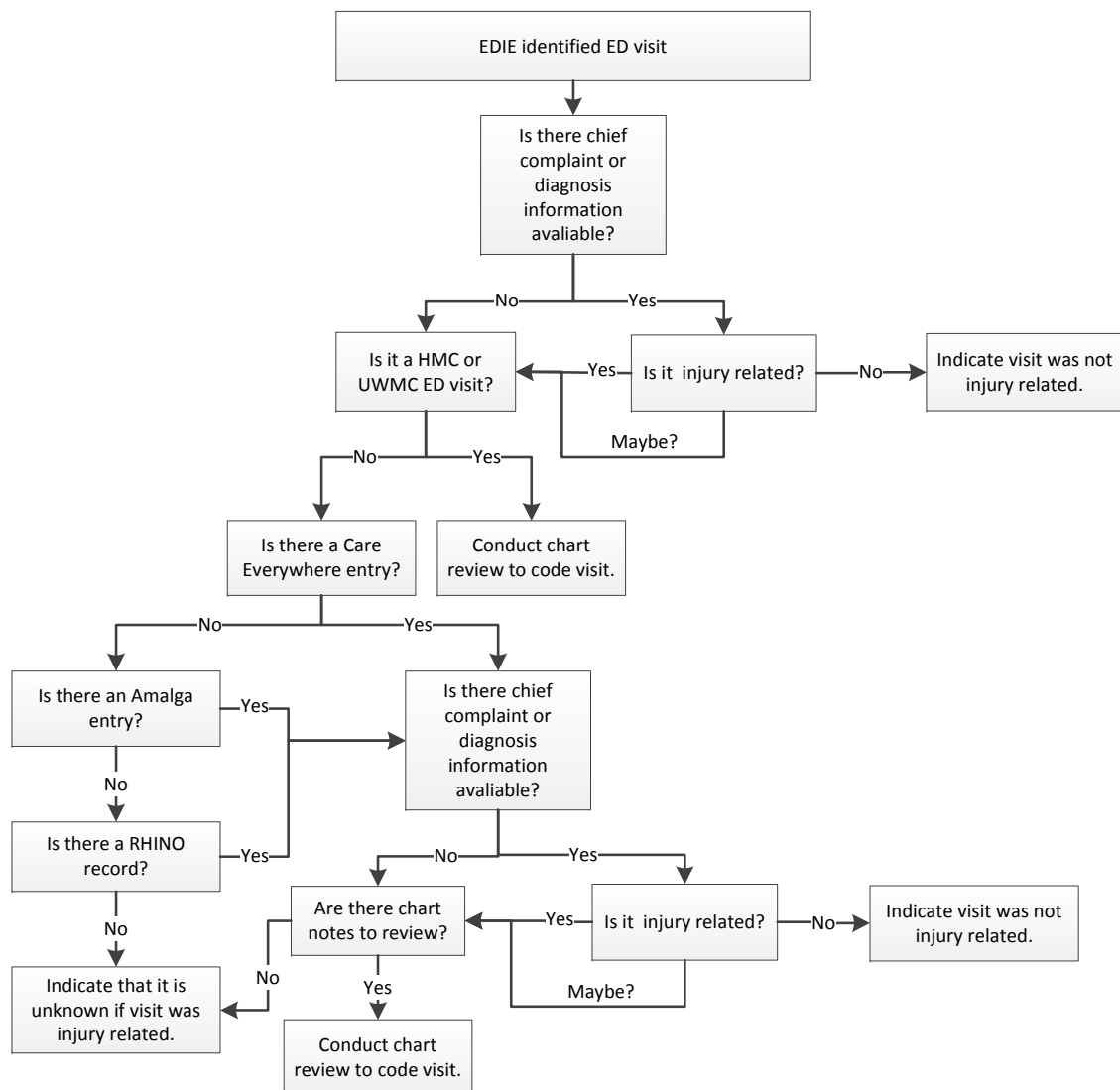


Figure 1: Process to determine if ED visit was injury-related

While EDIE was reliable in the reporting of specific dates of patient visits to EDs and clinics in the state, many visits were missing chief complaint information. Second, for any ED visits that occurred at the Harborview Medical Center (HMC) or affiliated hospitals, we conducted chart reviews of prior ED visits and any resulting hospitalizations using HMC electronic medical records (EMR) using EPIC to ascertain whether or not the visit was injury-related, an incident injury, mechanism used, body area injured and reason for a return visit (if not an incident injury). We used Care Everywhere to ascertain ED utilization as our third data source. Care Everywhere is a data sharing platform integrated into the EPIC electronic health record system that was developed to create a more comprehensive health record and improve the speed of care, especially when seeing multiple providers in different hospitals and clinics. For all medical centers using EPIC, a physician can open Care Everywhere and see a log of all patient visits made to other clinics that use EPIC, including the date and location of visit, provider seen and (often) a short chief complaint description. The provider can then query chart notes for previous visits in different hospitals, and can have almost immediate access. Any charts that have been queried by a HMC or affiliated physician as part of care for a patient are then imported into their HMC medical record.

As our fourth ED visit data source, we used the Rapid Health Information Network (RHINO) was developed for syndromic surveillance and is a Washington State Department of Health run real-time ED data. However, it is not yet fully online and initially EDs did not need to report patient name and date of birth, although some did. Patient sex, race and age were reported. We received records from the start of RHINO through June 30, 2019

for any ED visit reported to RHINO. Using patient sex, race and age with the ED visit time and date obtained from EDIE, we further supplemented data on any remaining ED visits without known chief complaint.

*Linking.* For CHARS data we conducted a two-step linking process where protected health information (PHI) for all individuals with a hospital visit in Washington State from 2016 through 2018 were sent from the Department of Health (DOH) to study staff where they were linked with HiFi patient PHI with Link King using first name, last name, middle name, date of birth, sex and race. Following the linkage, a crosswalk file with unique identifiers for HiFi patients in CHARS was sent to the DOH where it was linked with hospitalization records, which were then sent back to HIPRC for final linking.

*Building the state-wide ED visit dataset.* For all recorded ED visits, we conducted a structured tiered approach to determine if the visit was injury related beginning with the EDIE generated report on ED visits for HiFi patients (Figure 1). For each ED visit, we coded whether or not the visit was injury related, an incident injury, mechanism, body area, links between multiple ED visits for the same underlying injury and any treatment received for return visits. If the first data source had sufficient information to determine whether or not the visit was injury related, and if so, could complete all coding then no additional data sources were used. There were 1,639 total ED records identified in EDIE, of which 61% had some chief complaint information, we did not assess level of missingness in other EDIE fields. We conducted chart review for 569 visits with HMC EMR and 268 visits with CareEverywhere records. We found chief complaint for 4 patient visits in RHINO. Following adjudication, there were 219 ED visits that were unknown if they were injury related or not, and 4 injury-related visits which we were unable to determine if they were incident injuries. Following creation of this ED visit dataset, we merged subsequent ED injury data with CHARS for hospitalization data and Vital statistics for death records.