# Extracting sources of admission and discharge destinations from discharge summaries

**Dina Demner-Fushman, MD, PhD, Swapna Abhyankar, MD,**
**Fiona M. Callaghan, PhD, Clement J. McDonald, MD**
**National Library of Medicine, Bethesda, MD**

**Abstract**

*Sources of admission and discharge destinations are of interest in the analysis of repeat hospitalizations and survival. When coded information is not available, locations can be extracted from discharge summaries. We used a dependency parser and a Medicare-based dictionary of admission and discharge locations to extract and code the source and destination locations. The extraction results were deemed appropriate for inclusion in our ongoing survival analysis.*

**Introduction**

Sources of admission and discharge destinations are of interest in the analysis of repeat hospitalizations and survival after ICU and hospital discharge. Patients with chronic obstructive pulmonary disease discharged to nursing homes were found less likely to be readmitted within 30 days after discharge than patients discharged to personal homes[1]. Complete medical records include coded fields for discharge destination (patient status discharge code) and the point of origin (source of admission codes) within the Medicare claims data. This coded information is not present in the publicly available rich source of clinical data, MIMIC II database[2], but can be extracted from discharge summaries.

**Methods**

We analyzed randomly selected discharge summaries to derive surface representations of the Medicare codes. For example, *Discharge Disposition: Home with Service* was mapped to status code 06. The automatic extraction of the codes proceeded as follows: If none of the section titles and noun phrases containing discharge destinations were found, the text was searched for discharge verbs (discharged, sent, transferred, etc.) and locations (home, hospice, rehab, etc.). If both term types were found in a sentence, the sentence was processed using Stanford parser. If a *TO* dependency was established between the verb and location, the discharge code was assigned. For the admission codes, the same method was applied using admission verbs (transfer, arrive, present, etc.) and locations. If a *FROM* dependency was established, the appropriate admission code was assigned. If the algorithms failed to extract locations, admission code 09 (information not available) and discharge code 07 (discontinued care) were assigned. To gauge the extraction accuracy, the extracted discharge code 20 (*expired*) was compared to coded information in the *date_of_death* and *expired_flag* fields. We inspected 25 additional randomly selected (but covering all remaining assigned codes) documents to see if the extraction quality is comparable to that of code 20 and suitable for our exploratory analysis of the data.

**Results**

The coded and extracted information for discharge code 20 differed for two patients. In one case, the discharge summary clearly indicated the patient has expired. In the other case, the discharge summary contained conflicting information. Other codes assigned to 25 documents, such as discharge code 50 (hospice) and admission code 04 (transfer from hospital) were correct. We found one error in the 09 admission code assignment for complicated text spanning several sentences and requiring co-reference resolution and inference. The extracted variables were included in the survival analysis.

**References**

1. Camberg LC, Smith NE, Beaudet M, Daley J, Cagan M, Thibault G. Discharge destination and repeat hospitalizations. Med Care. 1997 Aug; 35(8):756-67.
2. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman LW, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database. Crit Care Med. 2011 May;39(5):952-960.