Assessment of Kentucky's Computerized Hospital Discharge Data System for use in the Surveillance of Injuries and Violence

Michael Singleton

Lei Yu

Kentucky Injury Prevention and Research Center

September 2006

Funded by the Centers for Disease Control and Prevention
Cooperative Agreement U17/CCU424761-01

Special thanks to:

Frank Blair, University of Kentucky Hospital
Travis Burgett, University of Kentucky Hospital
Rhonda Ransford, University of Kentucky Hospital
Paige Franklin, Kentucky Hospital Association

Summary of findings and recommendations

- Although there are some significant gaps (namely residents hospitalized out of state, and records with missing external cause of injury codes, or E-codes), the HD system seems to produce information complete and accurate enough to support high-level injury surveillance, provided these gaps are kept in mind.
- Primary data collection for high-priority issues may be necessary in order to obtain the information on circumstances and contributing factors needed to design interventions. Some possibilities include surveys, medical record abstraction, and designating certain types and causes of injury as reportable public health conditions.
- The state planning group should consider how resources should be divided between efforts to improve the existing system and development of supplemental, primary sources. The answer depends on a judgment of how much quality improvement of the existing system is realistically possible.
- E-coding is not bad, but can be improved. The reasons for missing and non-specific E-codes should be determined. If insufficient documentation on the medical record is found to be an issue, continuing education courses for physicians are one possible way of addressing it.
- Determine reasons for the E-code rate declining as age increases.
- Hospitals should be advised that E-codes should NOT be placed into diagnosis code fields when preparing computerized UB92 submissions. This practice can result in E-codes being truncated from the record, since the computerized file includes at most nine diagnosis code fields. E-codes should go in separate fields that are designated for that purpose.
- Hospital coding staff should be made aware/reminded of the supplemental injury codes that exist in ICD-9 and ICD-10 for indicating the place of injury and the perpetrator in abuse and maltreatment cases, and the importance of this information to injury prevention.
- Out-of-state hospitalizations impact our ability to identify geographic priorities, since the border counties are disproportionately affected. The Department for Public Health might pursue reciprocal reporting agreements with surrounding states who receive significant numbers of Kentucky-resident inpatients. Alternatively, it may be possible to impute, or otherwise estimate, out-of-state hospitalizations for geographic analyses.
- We estimate that between five and ten percent of episodes of care consist of multiple discharges. For injury surveillance purposes the impact of this is probably minimal. However, for research projects, it should be carefully considered whether combining length of stay, charges, etc. across such episodes may be appropriate.

INTRODUCTION

Computerized records generated by administrative state data systems are a major source of public health surveillance data. Administrative data has been defined as "information collected by government, for some administrative purpose (e.g. keeping track of the population eligible for certain benefits, paying doctors or hospitals), but not primarily for research or surveillance purposes (Spasoff, 1999). Such data systems have both strengths and weaknesses in terms of their use in public health surveillance and research.

The statewide computerized hospital discharge databases, generated by COMPdata on behalf of the Kentucky Hospital Association and the Kentucky Department for Public Health, are the main source of data on injury-related hospitalizations within the state. The primary purpose of this assessment was to identify both the benefits and drawbacks of using these computerized databases to conduct surveillance of fatal injuries.

The Centers for Disease Control and Prevention (CDC) have published guidelines for the evaluation of public health surveillance systems (CDC 2001). Those guidelines outline several desirable characteristics of such systems, including simplicity, flexibility, data quality, acceptability, sensitivity, predictive value positive, representativeness, timeliness and stability. Due to time and resource limitations, a thorough evaluation of all these facets was not possible. Instead, we focused on documenting the system's operation and assessing the quality of the data it produces.

Background

Chapter 216 of the Kentucky Revised Statutes (KRS 216.29) mandates the collection of data on the "cost, quality, and outcomes of health services provided by health facilities and health care providers in the Commonwealth." In 1996 collection of hospital discharge data began. Regulation 902 KAR 19:020 specifes the use of the national standard Uniform Bill (UB) reporting form, as well as other details of data collection and processing. As of this writing, the current version of the UB standard is UB-92. Computerized records are collected for all inpatients discharged from licensed, acute care hospitals in Kentucky. Reporting of selected outpatient procedures is also required. Hospitals in the Veterans Administration (VA) medical system are exempt from the reporting requirement.

Although hospital discharge data collection began in 1996, there were the usual startup challenges associated with the implementation of a major, statewide data collection initiative. Beginning with the calendar year 2000, several significant measures were instituted in order to upgrade the quality and completeness of reporting. For this reason, our assessment focuses on data from the years 2000 to 2004, which at the time we began data analysis were the most recent years available.

METHODS

Data sources

All data reported in this assessment were based on the 2000 to 2004 computerized hospital discharge files, unless otherwise specified.

Definition of injury-related discharge

In the broadest sense an injury-related discharge can be defined as one in which the principal diagnosis, or any of the reported secondary diagnoses, indicate an injury or poisoning (ICD-9 code between 800 and 999). However, the State and Territorial Injury Prevention Directors' Association recommends a much more restricted definition based on the principal diagnosis only, and omitting a number of codes in the range 800-999, such as late effects of injuries, adverse effects of drugs, surgical procedures, and medical procedures, and others (STIPDA 2003). We have used STIPDA's definition of an injury discharge in this report, unless otherwise indicated.

Our framework for classifying the external cause of injury by mechanism and manner was taken from the same source.

Interviews

Most of the information on the operation of the system – the collection of data, coding and computerization – was obtained through interviews with staff at the Kentucky Hospital Association (KHA) and the University of Kentucky Hospital (UKH). Given time and resource constraints it was not possible to survey all hospitals about their procedures.

RESULTS

System operation

In the case of injury mortality data, the National Vital Statistics System (NVSS) provides an overarching system and set of guidelines for all states and jurisdictions to follow. For the collection of hospital discharge data, there are some national standards that most states and hospitals follow, but there is not the same degree of uniformity in the underlying collection and coding processes. The following description of the operation of Kentucky's Hospital Discharge data collection system is based largely on a detailed examination of the process at UKH. We did not attempt to determine the degree to which UKH's process reflects the typical hospital scenario.

Data collection

From the time that a patient presents at the hospital until the time of discharge, data about all aspects of that patient's medical history, diagnosis, treatment, etc., are recorded. Unlike the death certificate, where one or two persons (funeral director and coroner/physician) are responsible for the data collection, the information comprising a patient's medical record is captured by a number of different persons located in a variety of departments throughout the hospital.

Upon presenting at UKH, the patient is registered. Demographic and insurance data are entered into the Patient Management (PM) system, and insurance status is verified. At this time also the patient is assigned a medical record number (MRN). If the patient has been admitted to the hospital previously, the existing MRN is used.

As the patient is treated, records of treatment provided by all departments are accumulated into the Patient Accounting (PA) system. Also, narrative data on the patient's medical history, reason for admission, diagnoses, etc., are recorded in various departmental charts and hard-copy documents.

Medical record coding

After the patient is discharged, the data collection phase ends. Tuesday through Friday, medical records staff go to their assigned floors and collect departmental charts on patients discharged on the previous day. (On Monday charts are collected for discharges that occurred on Friday, Saturday, and Sunday). These charts are then assembled in a standard order into the final, hard-copy medical record.

UKH uses 3M encoding software to derive diagnostic, procedure, and external-cause-of-injury codes based on narrative information entered from the hard-copy medical record. This software incorporates all standard coding classification rules, regulations, and recommendations. Also, based on all supplied information it identifies complications and comorbities among the diagnostic codes. *In the resulting computerized record, complications and comorbidities are the first codes listed after the principal diagnosis.* These could include conditions such as pneumonia, urinary tract infections, sepsis, and many others. Any additional diagnostic codes follow the complication and comorbidity codes. There is essentially no limit to the number of diagnostic codes that can be entered in the computerized medical record.

Submission of computerized UB92 records to CompData

At this point the computerized patient medical record has been created. The Patient Accounting department will generate a hard-copy UB92 billing form to be sent to the primary expected payor (Appendix A contains a copy of the UB-92 form). The hard-copy UB-92 can include up to 25 diagnostic codes. There are explicit boxes ("Form Locators") for the Principal Diagnosis Code (Form Locator 67) and up to eight additional codes (FL-68 through FL-75). At UKH, any diagnostic codes beyond these nine will be printed in the "Remarks" box (FL-84).

However, the hard-copy UB92 form is not the record that is ultimately reported via CompData. UKH's Decision Support Systems (DSS) department is responsible for generating the computerized UB92 records and submitting them quarterly to CompData. The computerized UB92 records include only the Principal Diagnosis Code and the first (up to) 8 additional codes after the Principal. This means that on some records the diagnostic information is incomplete. From 2000 to 2004, about 26% of all (not only UKH's) computerized HD records had all nine diagnostic code fields populated. This was more common for patients 65 and older. These cases could have had codes truncated from the computerized record.

Processing of UB92 records by COMPdata

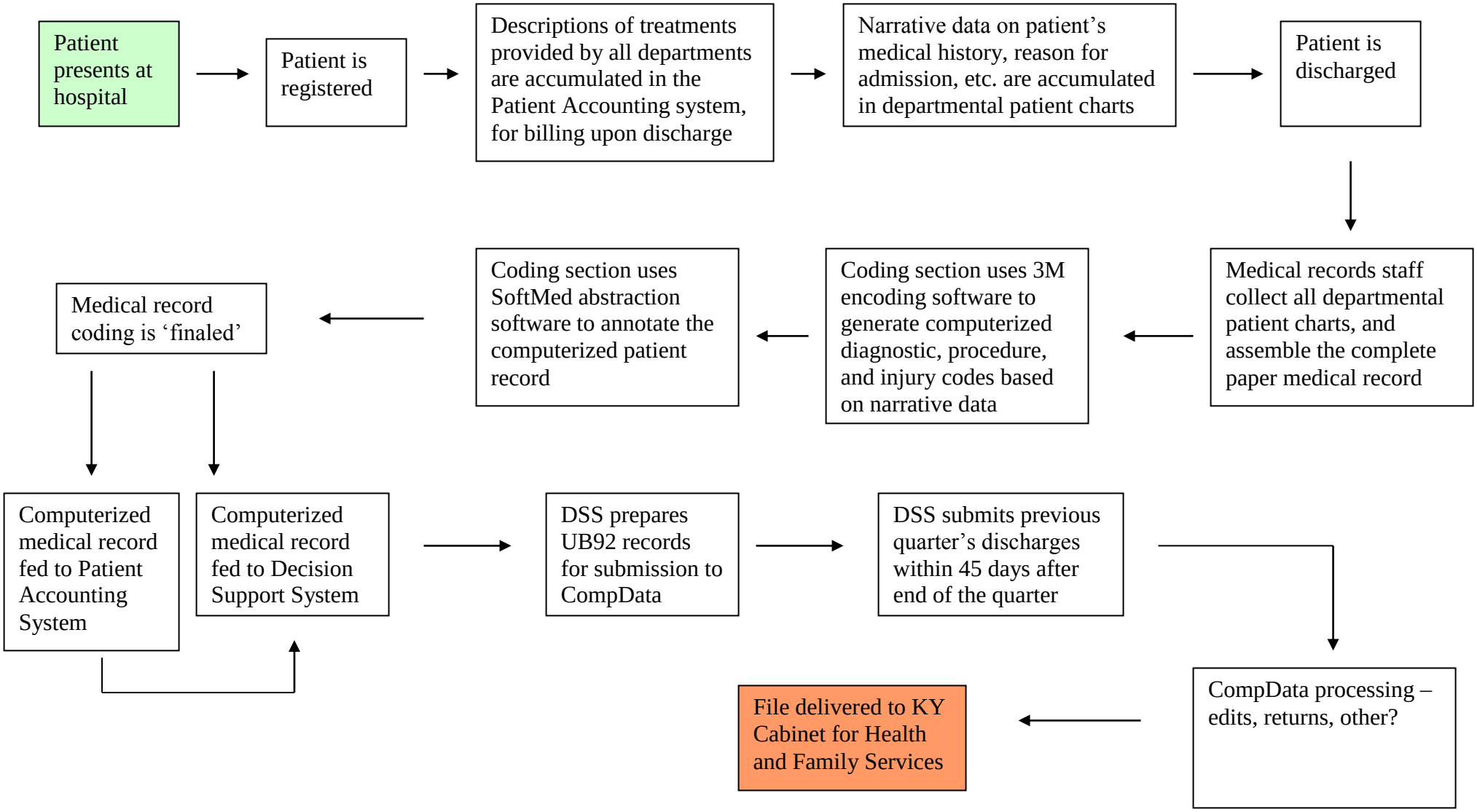
COMPdata accepts only computerized records. (As of this writing, only one hospital was unable to submit computerized records, and that facility was scheduled to begin doing so during 2006). According to KHA, at least 500 edits are applied to each record received. Some are simple checks for valid values for individual fields, others are more complicated, cross edits of one field against another – for example, checking reported gender against reported procedure codes. If a record fails any edit it is rejected and returned to the submitting hospital for correction and resubmission. Appendix C contains a list of error codes, which gives an idea of the kind of edits that are applied to the records.

For the majority of hospitals, the number of initially rejected records in any given batch is less than 2% of those submitted. KHA receives quarterly reports from COMPdata showing the percentage of submitted records successfully loaded for each hospital. For facilities that had very low percentages for a given quarter, KHA requests that they resubmit the rejected records. According to KHA, serious reject problems are usually the result of a computer system change at the hospital or a programming error.

Per statute, hospitals have 45 days after the end of the quarter to submit their records, and 30 days to correct and resubmit rejected records. The percentage of rejected records that are never resubmitted is not known, but certainly this is a source of missing discharge cases. In effect, a higher standard of quality for accepted records is enforced, at the cost of missing those cases that are not corrected and resubmitted.

Each year, typically in June or July, a copy of the final database is delivered on CD to the Kentucky Cabinet for Health and Family Services.

Figure 1. Flowchart of hospital discharge data collection and processing



System quality control/quality assurance

We are not aware of any systematic studies of the accuracy of the Kentucky HD injury data. In this assessment we focused on documenting the existence (or lack) of measures designed to promote data quality, such as computerized edit checks, consistent use of standards, and auditing practices. But we did not attempt to assess the extent to which those measures are successful.

Do the computerized data accurately represent what actually occurred?

This a very difficult question to answer, since we generally don't know what "really happened"; at best we know what was reported in the hard-copy medical record. On the surface, it would seem that in the data collection phase, hospitals might have a more vested interest or motivation than funeral directors and coroners in being thorough and accurate in their collection of information. Data entry software usually includes edit checks of some kind. These ensure, at a minimum, that a valid entry was made for an individual data element. However, they can't ensure that the *correct* entry was made. (For example, a male misclassified as a female would pass an edit check on Gender.) More sophisticated "cross-edits" can check whether entries are logically consistent with one another (e.g. men generally don't have hysterectomies). Also, standards and auditing process can promote uniformity in coding across hospitals.

However, it's not only a matter of diligence on the part of the hospital staff. Some information is self-reported by the patient. In one case we examined in connection with a Crash Outcome Data Evaluation System study, a man presented to UKH in the early morning hours, with facial injuries which he claimed resulted from a fall from a porch. The record indicated he was intoxicated. He later changed his story to say that he'd been involved in a motor vehicle crash. Presumably he initially reported a fall due to fear of legal consequences for DUI. Assuming that his second story was the truth, had he not revealed this information his cause of injury would have been misclassified as a fall.

Finally, it needs to be remembered that the computerized discharge record is a truncated version of the complete medical record, most obviously with respect to diagnostic coding.

Are the captured data accurately transferred to the computerized record?

A more tractable question is whether the computerized data accurately reflect the information captured in the medical record. This can be studied through medical record review, though there have been few such systematic studies of the Kentucky HD injury data that we are aware of. The Intimate Partner Violence Surveillance (IPVS) project at KIPRC did analyze the reporting of perpetrator

codes in cases of hospitalization resulting from abuse and maltreatment. A number of discrepancies were discovered between the perpetrator as specified in the E-code and as recorded in the medical record. As a result, the investigators recommended that the fourth and fifth digits of E-codes and diagnostic codes be interpreted with caution.

Completeness of computerized data for injury discharges

Completeness is a bit easier to assess than accuracy, since there are some convenient benchmarks against which to compare the reported data.

Discharges that are not captured

While we know that the system does not capture 100% of discharges of Kentucky residents, comparisons with U.S. estimates from the National Hospital Discharge Survey (NHDS) indicate that discharge rates for Kentucky residents (based on our HD data) are comparable to national discharge rates (Table 1, first and second rows). This suggests that the great majority of discharges are being captured.

Previous research (Singleton, Xiao 2005) has suggested that the number of Kentucky residents discharged from hospitals in surrounding states is similar to the number of out-of-state residents discharged from Kentucky hospitals. If we include non-Kentucky residents we find that the rate of discharges reported by Kentucky hospitals is nearly identical to the national reported rate, and in fact is slightly higher Table 1, third row).

Table 1. Rate* of discharges from non-federal, acute care hospitals (2000-2004)

Source	**2000**	**2001**	**2002**	**2003**	**2004**
NHDS (U.S.)	887	921	939	978	973
Kentucky HDD – Kentuckians only	765	853	887	897	898
Kentucky HDD – all reported cases	837	937	972	983	986

* Per 100,000 population

Nevertheless, we attempted in several ways to locate areas where discharges might be missed by the system.

Reporting by in-state hospitals: Each year, along with the copy of the HD database we receive from KCHFS, we get a list of Kentucky hospitals that should be reporting. Since Veterans Administration, rehabilitation, and psychiatric hospitals are not required by the statute to report discharges, we omitted them from this analysis. Considering inpatient hospitals only, in every year from 2000

to 2004 there were only one or two listed hospitals that failed to report any discharges, The percentage of reporting inpatient hospitals in each year was more than 98%.

In addition we checked for periodic lapses in reporting by looking at the number of discharges reported by month for each inpatient hospital. We found only a very few cases of inconsistent month-to-month reporting.

Finally, we checked for hospitals that might be consistently underreporting. First, we looked at discharge rates per 100,000 residents by county, reasoning that an outlying (on the low side) discharge rate in a particular county would suggest an underreporting hospital in the vicinity, though not necessarily in the same county. (The exception is that in border counties, a low rate might also be the result of frequent out-of-state hospitalizations. Therefore we applied this reasoning only to non-border counties). Indeed, we found one case involving a county in the interior of the state, with an unrealistically low rate of injury discharges compared to the state as a whole. Looking at the nearby hospitals, we found one (in the same county, in this case) that had clearly been underreporting, as shown by a substantial increase in reported discharges in 2003 and 2004 compared with 2000 through 2002. The problem appears to have been corrected.

We also calculated the rate of discharges per acute care bed for each hospital. The rates ranged from a high of more than 13 to a low of less than 1, with an average of just under 6. It was not clear to us whether this is a meaningful result, and we did not pursue it further.

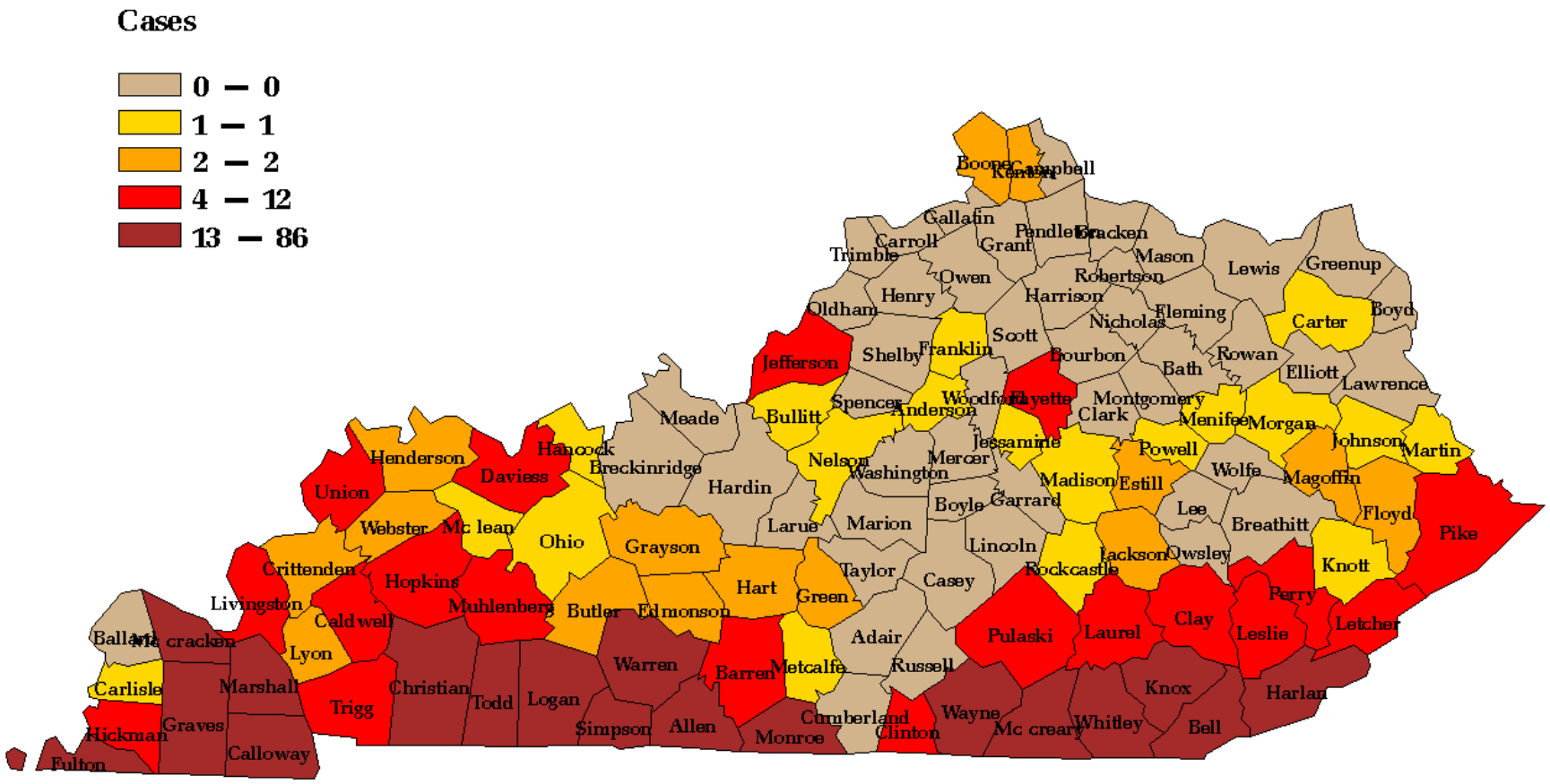
Rejected submissions: As noted previously, COMPdata rejects all records that fail any edit check and returns them to the hospital for correction and resubmission. It is a certainty that not all rejected records are resubmitted, but the number of such occurrences is not known.

Out-of-state hospitalizations: Kentucky is bordered by seven states: Illinois, Indiana, Ohio, West Virginia, Virginia, Tennessee, and Missouri. Of these, we know that hospitals in Tennessee, Ohio, and West Virginia receive non-trivial numbers of Kentucky-resident inpatients. For example, Tennessee hospital discharge data for 2002 and 2003 show 806 and 850 cases, respectively, of Kentucky residents with principal diagnoses meeting the STIPDA definition for an injury discharge. Figure 2 shows the distribution of the Tennessee injury cases for 2003, by the patient's county of residence.

If we assume, as noted above, that the number of Kentuckians hospitalized in other states is similar to the number of non-Kentucky residents hospitalized in Kentucky, then out-of-state hospitalizations comprise about 9% of discharges of Kentucky residents per year. The fact that these cases are more common in border counties complicates geographic analyses. For this reason it might be worthwhile to develop a method of imputing the number of out-of-state

discharges by county, if we are unable to obtain the actual cases in a timely fashion.

Figure 2. Number of Kentucky residents discharged from Tennessee hospitals in 2003, by county of residence



Incompletely reported data elements on computerized records

Missing values: Since 2000, there has been very little problem with missing values on the reported cases. Presumably this is because records with missing values are rejected by CompData, and returned for completion and resubmission. (However, as noted above, this creates the issue that rejected records may end up becoming missing cases if a corrected record is not resubmitted.)

External-cause-of-injury reporting (E-coding): The completeness of E-codes deserves special attention, since along with the principal diagnosis it is the most fundamental piece of injury-related information on the HD file. If the E-code is missing on an injury record (as indicated by the principal diagnosis), we cannot determine the cause of the injury, nor whether it was unintentional or intentional. Further, the ICD-9 and ICD-10 schemes provide the capacity to code a significant amount of relevant information about the circumstances of injury, but unfortunately this capacity is underutilized.

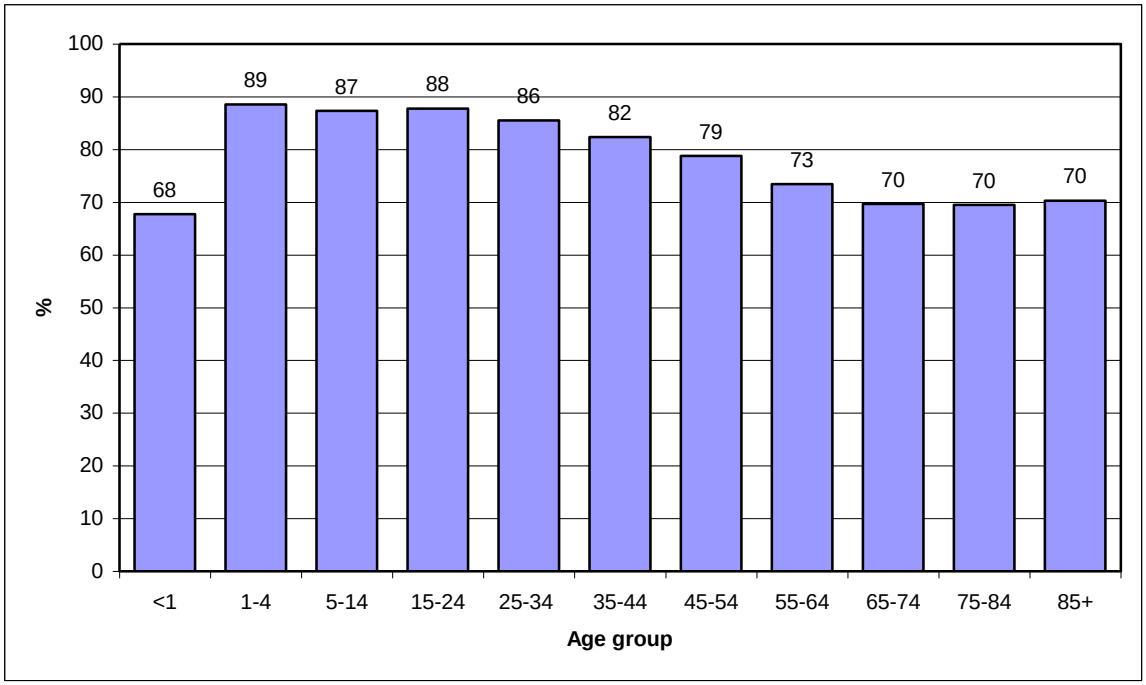
Overall, if we define an injury according to the STIPDA case definition, then the percentage of injury records having at least one valid E-code (not including 'Location of Injury' codes) is nearly 80%. Reporting for ages 1 to 34 approaches 90%, whereas reporting for ages 65 and older is about 70%. The reasons for the lower rate among older patients are not known. Older patients tend to have more diagnosis codes, on average, than younger patients due to poorer overall health resulting in more codes for complications and contributing factors. It follows that they are more likely to have diagnosis codes truncated when the records are computerized, due to the limited number of fields (9) available on the computerized record. E-codes are sometimes mistakenly reported in diagnosis code fields, so truncation of diagnosis codes might also eliminate some E-codes. But this doesn't seem likely to explain the full difference.

Table 2. E-coding of injury and poisoning cases

	Year				
	2000	**2001**	**2002**	**2003**	**2004**
Number of discharges having a principal diagnosis of 'Injury and Poisoning' (800-999)	24,171	26,294	26,955	27,234	27,515
Number of such cases having at least one valid e-code*	18,034	20,256	21,217	21,352	21,648
Percent of such cases having at least one valid e-code*	75%	77%	79%	78%	79%

*Excluding place of injury codes (E849)

Figure 3. Percentage of computerized injury and poisoning discharge records having at least one valid E-code, by age group (2000-2004)



Lack of specificity on reported E-codes: Another issue with E-codes is that even when reported, their potential is often underutilized. For example, from 2000 to 2004 there were 39,458 discharges related to falls. Of these, 51% were reported as an "Unspecified Fall" (E888), when more specific codes describing the nature of the fall were available.

Examples include "Fall from stairs or steps," "fall from ladder or scaffolding," "fall from slipping, tripping, or stumbling," etc. Presumably this is due to lack of information reported in the medical record. However, in the American Journal of Public Health, Langlois et. Al (1995) reported that the completeness and specificity of E-codes in Rhode Island's HD records could be substantially improved by "making better use of existing documentation in medical records." Whether this is true for Kentucky has not been determined.

Underutilization of 'place of injury' codes: ICD-9 and ICD-10 both provide supplemental codes that allow for reporting of the place where an injury occurred, and ICD-10 also provides a method of classifying the kind of activity in which the person was engaged when the injury took place. These supplemental codes are discussed briefly below. The ICD-10 codes are included here because at some point in the future, hospitals will be required to transition from ICD-9 to

ICD-10 for coding hospital discharge records. Actually it is an understatement to call these codes 'underutilized.' In fact they are very rarely used.

ICD-9: The code E849 can be used to indicate the place where an injury occurred. The allowable codes are as follows.

849.0 Home
849.1 Farm
849.2 Mine and quarry
849.3 Industrial place and premises
849.4 Place for recreation and sport
849.5 Street and highway
849.6 Public building
849.7 Residential institution
849.8 Other specified places
849.9 Unspecified place

From 2000 to 2004, less than 1% of injury records specified a location of injury E-code.

ICD-10: For non-transportation E-codes (W00-Y34, excluding Y06 and Y07), a fourth digit may be specified indicating the place where the injury occurred. The codes are as follows.

.0 Home
.1 Residential institution
.2 School, other institution, and public administration area
.3 Sports and athletics area
.4 Street and highway
.5 Trade and service area
.6 Industrial and construction area
.7 Farm
.8 Other specified places
.9 Unspecified place

Also, an "Activity Code" is provided in ICD-10. Although it is not entirely clear from the ICD-10 manual, it appears to be intended for use as a separate data element, rather than appended to the E-code as with the place of occurrence codes. The codes are as follows.

0 While engaged in sports activity
1 While engaged in leisure activity
2 While working for income
3 While engaged in other types of work
4 While resting, sleeping, eating, or engaging in other vital activities
8 While engaged in other specified activities
9 During unspecified activity

Duplication

The existence of duplicate computerized records representing the same patient and episode of care seems to be quite rare. From 2000 to 2004 the number of cases having true duplicates was less than a tenth of one percent. Most the these duplicate records are nearly identical to the original, differing only in the expected payment source and hospital charges.

Multiple-discharge cases

A more common issue is the existence of multiple-discharge cases. Whereas there is only (except in rare cases) one death certificate record for a given person and injury event, there can be multiple hospital inpatient records related to the same person and injury event. We estimate that in 5% to 10% of injury cases, the full record of acute treatment is spread across multiple discharge records.

Figure 4 flowcharts the logic that we used to identify clusters of discharges that pertained to the same patient and injury event. We identified four types of multiple-discharge cases, which are described below. (Note: this work represents a low-end estimate, because our conservative method of identifying multiple-discharge cases relied on exact, rather than probabilistic, matching of patient characteristics. Also, we somewhat arbitrarily identified seven days as the cutoff beyond which we considered discharges for the same individual to be unrelated. Clearly it is possible for discharges separated by several weeks, or even months, to be related to the same incident or underlying condition. But for longer intervals between discharges, it becomes difficult to make that judgement based on the information on the computerized record, particularly for older patients.)

Consecutive discharges from the same hospital: These are chains of records representing distinct discharges of the same patient from the same hospital, where the discharge date for one record was equal, or one day prior, to the admit date for the next. These chains are the easiest type to identify, since the discharges are close together in space and time. Roughly 2% to 3% of the discharges in a typical year are part of this type or chain (a more sophisticated search might find a higher percentage). The majority of these cases, typically between 60% and 70%, involved persons aged 65 and older. The most common scenario involved an emergent admission for a fall, followed by an elective admission. The elective admission was frequently for rehabilitation, but in many cases the reason was not entirely clear based on the information contained in the computerized record. Among younger patients, another common scenario was an emergent admission for poisoning (either unintentional or intentional), followed by an elective admission for treatment of related conditions such as depression or anxiety.

Consecutive discharges from different hospitals: These are chains of records representing distinct discharges of the same patient from different hospitals, where the discharge date for one record was equal, or one day prior, to the admit date for the next. The kinds of scenarios in this category were similar to the previous type, but self-inflicted poisonings were more common in this category. This might be due to the fact that the hospital that provided the initial treatment for the poisoning lacked capacity for subsequent treatment of the underlying psychological condition (depression, anxiety, etc.). Also in this category would fall cases where a patient was transferred to a trauma center for advanced treatment that the initial hospital couldn't provide.

About 3% of 4% of the discharges from 2000 to 2004 belonged to a chain of this type.

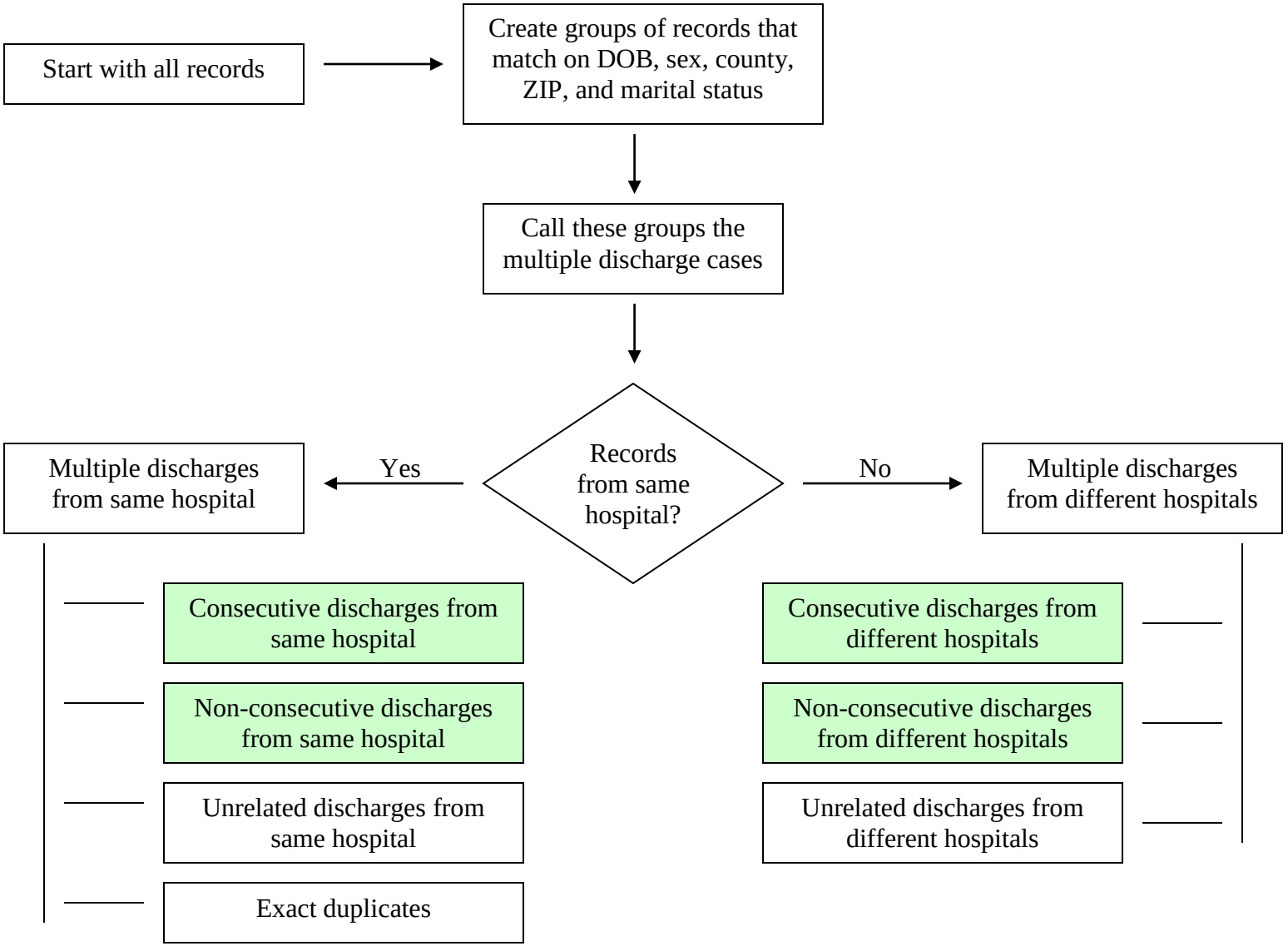
Non-consecutive discharges from the same hospital: These are groups of records representing distinct discharges of the same patient from the same hospital, where the discharge dates were separated by at least two, and no more than seven, days. In this category, the most common reasons for the multiple discharges were infections and other complications, adverse effects of drugs or medical/surgical procedures, and late effects of injury, following an initial trauma from a fall or motor vehicle crash. About 3% of the discharges from 2000 to 2004 belonged to a group of this type.

Non-consecutive discharges from different hospitals: These are groups of records representing distinct discharges of the same patient from different hospitals, where the discharge dates were separated by at least two, and no more than seven, days. In the absence of personally identifying information, these cases are difficult to identify with a high degree of certainty, due to the separation of the events in both time and space. Therefore we did not attempt an analysis of this type of multiple-discharge scenario.

Bottom line: If we are only interested only in counting numbers of cases, the existence of multiple discharges for a single patient and injury event probably does not make a significant difference. In fact, STIPDA's official guidelines recommend not removing readmissions, transfers, etc. for the purposes of injury surveillance.

However, if we are interested in testing research hypotheses, it is necessary to consider whether the results will be significantly affected by how we treat these multiple-discharge cases. For example, suppose we have supplemental data on seat belt usage for persons hospitalized as a result of motor vehicle crashes. A common approach is to compare injury severity among the belted and non-belted inpatients using the length of stay and/or total charges. If a researcher does not identify multiple-discharge cases, and combine the length of stay and/or total charges where appropriate, he or she may reach a misleading conclusion.

Figure 4. Flowchart of process for identifying duplicates and multiple-discharge cases



DISCUSSION

Two-tiered approach to surveillance

The strength of administrative data sets – such as those based on hospital discharge records – for surveillance purposes lies in high-level overviews based on a select number of relatively complete and reliable variables (e.g. age, sex, county of residence). Efforts to drill deeper into these datasets on specific topics generally yield disappointing results in terms of information for prevention purposes. This is understandable, since by definition 'administrative data sets' were not created for the purpose of supporting surveillance and prevention research.

A two-tiered system would use the computerized hospital discharge records primarily for overall prioritization and monitoring of demographic and geographic trends at a macro level, and secondarily as a sampling frame for in-depth, primary data collection. The latter would focus on the circumstances of injury, presence of risk and protective factors, etc., and would likely need to be targeted at a limited number of focus areas identified by the macro level analysis. Statewide inferences could then be made based on the sample. Of course priorities may change over time. The Centers for Disease Control and Prevention's guidelines for central nervous injury surveillance provide a working model of such a system (CDC 2002).

In general, the planning group should consider how much effort to put into attempts to improve UB92 data collection through training and education, compared to efforts at developing primary data collection methods (surveys, abstraction, designating reportable conditions, etc.)

Dissemination strategy

Unlike mortality data, we can publish data on injury morbidity data without a great deal of concern about consistency, because there is little or no officially published data on the subject from other sources. Nevertheless, it is important to consider strategically, in coordination with the injury planning group, who are the key audiences for our messages, and what are the most effective formats for those messages.

Managing system changes

Comparability of surveillance data over time depends on well-defined and stable data collection and processing operations. It needs to be clearly recognized that changes to the underlying processes have implications for the surveillance systems that are built on them.

Internal changes: Permanent changes in, or short-term deviations from, usual data collection, data entry, coding, or other operations will affect the results produced by surveillance systems that are fed by the hospital discharge data. At a minimum, persons responsible for those surveillance systems should be notified when changes or deviations occur. Ideally they should be informed before they occur, and offered an opportunity to provide feedback and as to how the changes will impact surveillance.

External changes: Changes at the federal level also occur periodically and are not under our control. Examples include the upcoming transition from the UB92 to the UB04 billing form, and the eventual transition from ICD-9 to ICD-10 for diagnostic, procedural, and cause of injury coding. Such changes need to be carefully managed, so as to minimize their impact on surveillance systems.

REFERENCES

DeFrances CJ, Podgornik MN. *2004 National Hospital Discharge Survey.* Advance data from vital and health statistics; no 371. Hyattsville, MD: National Center for Health Statistics. 2006.

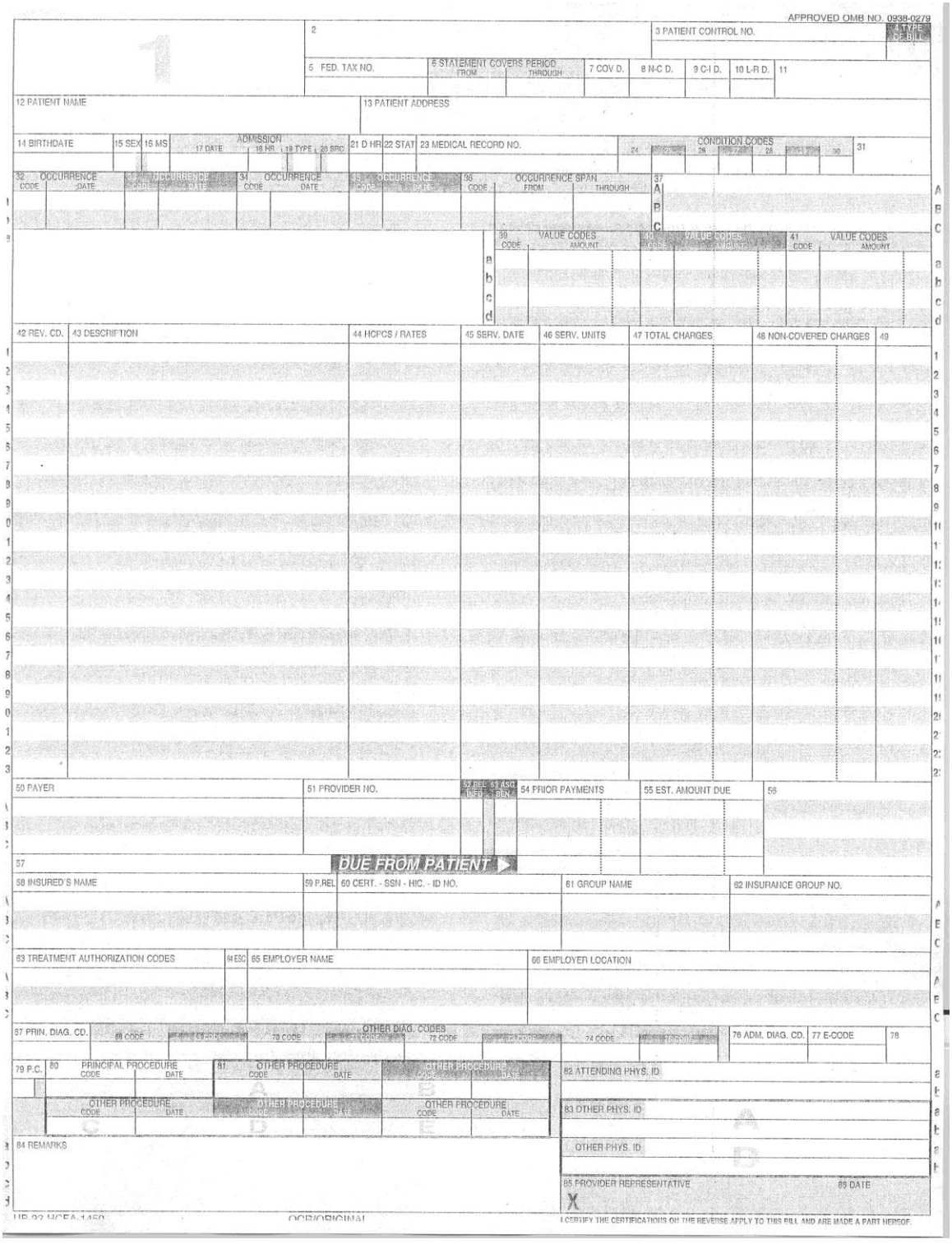
Langlois JA, Buechner JS, O'Connor EA, Nacar EQ, Smith GS. *Improving the E Coding of Hospitalizations for Injury: Do Hospital Records Contain Adequate Documentation?* American Journal of Public Health; September 1995, Vol. 85, No. 9.

Singleton M, Xiao Q, Agent K. *Economic Costs of Low Safety Belt Usage in Motor Vehicle Crashes in Kentucky*. University of Kentucky College of Engineering, Kentucky Transportation Center; 2005.

Marr A, Coronado V, editors. *Central Nervous System Injury Surveillance Data Submission Standards – 2002*. Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2004.

State and Territorial Injury Prevention Directors Association. *Consensus Recommendations for Using Hospital Discharge Data for Injury Surveillance.* Atlanta (GA): STIPDA; 2003.

APPENDIX A: UB-92 billing form



APPENDIX B: Error messages generated by COMPdata software edits

ADMISSION DATE
- ADMISSION DATE CANNOT PRECEDE BIRTHDATE
- ADMISSION DATE BLANK
- ADMISSION DATE > CURRENT DATE
- ADMISSION DATE AFTER DISCH. DATE
- ADMISSION DATE INVALID

ADMISSION SOURCE
- ADMISSION SOURCE CONFLICTS WITH ADMISSION TYPE
- ADMISSION SOURCE BLANK
- ADMISSION SOURCE INVALID (Not between 1-9 and A)

ADMISSION TYPE "ADMISSION TYPE INVALID (Not 1,2,3,4 or 9)"
- "ADMISSION TYPE ""4"", AGE MUST BE ""0"""
- ADMISSION TYPE BLANK

BILL TYPE TYPE OF BILL BLANK
- TYPE OF BILL INVALID
- UPDATE ERROR
- BILL TYPE DOESN'T MATCH BATCH TYPE OF BILL
- DELETE ERROR

DIAG./PROC. ERRORS
- DIAGNOSIS - AGE NOT NEWBORN
- DIAGNOSIS - AGE NOT CHILD
- DIAGNOSIS - AGE NOT MATERNITY
- DIAGNOSIS - SEX NOT MALE
- DIAGNOSIS - SEX NOT FEMALE
- UNACCEPTABLE PRINCIPAL DIAGNOSIS (V-CODE)
- PRINCIPAL DIAGNOSIS REQUIRES SECONDARY DIAGNOSIS
- UNACCEPTABLE PRINCIPAL DIAGNOSIS (MANIFESTATION)
- PROCEDURE - SEX NOT MALE
- PROCEDURE - SEX NOT FEMALE
- DIAGNOSIS - AGE NOT ADULT

DIAGNOSIS
- PRINCIPAL DIAGNOSIS BLANK
- SECONDARY DIAGNOSIS INVALID
- PRINCIPAL DIAGNOSIS INVALID
- DIAGNOSIS NOT VALID FOR PATIENT STATUS
- PRINCIPAL DIAGNOSIS CANNOT BE E-CODE

DISCHARGE DATE DISCHARGE DATE MUST BE >= TO ADMISSION DATE
- DISCHARGE DATE INVALID
- DISCHARGE DATE IS BLANK
- DISCHARGE DATE MUST BE <= TO CURRENT DATE

DUPLICATES
- EXACT DUPLICATE RECORD
- DUPLICATE RECORD

FEDERAL TAX ID
- FEDERAL TAX NUMBER BLANK

HOSPITAL ID
- HOSPITAL ID BLANK
- HOSPITAL ID ON PHYSICAL RECORD DOES NOT MATCH ID ON HEADER
- HOSPITAL ID INVALID

LENGTH OF STAY
- LENGTH OF STAY ERROR

CLOSED QUARTER
- DATA SUBMITTED FOR CLOSED QUARTER

RECORD ERROR
- RECORD ERROR - REVIEW RECORD

PHYSICIANUPIN NOT ON FILE
- STATE LICENSE NUMBER NOT ON FILE

MULTIPLE PAGE RECORD
- PATIENT RECORD CANNOT EXCEED 45 MULTIPLE PAGES

PATIENT BIRTH DATE
- BIRTH DATE INVALID
- BIRTH DATE OVER 124 YEARS OLD
- BIRTH DATE BLANK
- " BIRTH DATE, MUST BE <= CURRENT DATE AND DISCHARGE DATE"

PATIENT CONTROL NO.
- PATIENT CONTROL NUMBER BLANK
- PATIENT CONTROL NUMBER IS NOT CONSISTENT ON ALL CASE RECORDS

PATIENT STATUS
- PATIENT STATUS VALUE INVALID
- PATIENT STATUS BLANK

PAYER ERROR
- RINCIPAL PAYER BLANK
- SECONDARY PAYER INVALID
- PRINCIPAL PAYER INVALID
- THIRD PAYER INVALID

PHYSICIAN ID
- ATTENDING PHYSICIAN BLANK
- OUTPATIENT - PRINCIPAL PROCEDURE PHYSICIAN BLANK
- PHYSICIAN ID INVALID FORMAT

PROCEDURE CODE METH.
- CODING METHOD INVALID
- CODING METHOD BLANK

PROCEDURE DATE

- PRINCIPAL PROCEDURE DATE BLANK
- SECOND PROCEDURE DATE BLANK
- THIRD PROCEDURE DATE BLANK
- FOURTH PROCEDURE DATE BLANK
- FIFTH PROCEDURE DATE BLANK
- SIXTH PROCEDURE DATE BLANK
- PRINCIPAL PROCEDURE DATE FOUND - NO PROCEDURE CODE
- SECOND PROCEDURE DATE FOUND - NO PROCEDURE CODE