Unbiased Identification of Patients with Disorders of Sex Development

David A. Hanauer[1], **Melissa Gardner**[2], **David E. Sandberg**[2]*

[1] University of Michigan, Department of Pediatrics & Communicable Diseases, Ann Arbor, Michigan, United States of America, [2] University of Michigan, Pediatrics & Communicable Diseases and Child Health Evaluation & Research (CHEAR) Unit, Ann Arbor, Michigan, United States of America

Abstract

Disorders of sex development (DSD) represent a collection of rare diseases that generate substantial controversy regarding best practices for diagnosis and treatment. A significant barrier preventing a better understanding of how patients with these conditions should be evaluated and treated, especially from a psychological standpoint, is the lack of systematic and standardized approaches to identify cases for study inclusion. Common approaches include "hand-picked" subjects already known to the practice, which could introduce bias. We implemented an informatics-based approach to identify patients with DSD from electronic health records (EHRs) at three large, academic children's hospitals. The informatics approach involved comprehensively searching EHRs at each hospital using a combination of structured billing codes as an initial filtering strategy followed by keywords applied to the free text clinical documentation. The informatics approach was implemented to replicate the functionality of an EHR search engine (EMERSE) available at one of the hospitals. At the two hospitals that did not have EMERSE, we compared case ascertainment using the informatics method to traditional approaches employed for identifying subjects. Potential cases identified using all approaches were manually reviewed by experts in DSD to verify eligibility criteria. At the two institutions where both the informatics and traditional approaches were applied, the informatics approach identified substantially higher numbers of potential study subjects. The traditional approaches yielded 14 and 28 patients with DSD, respectively; the informatics approach yielded 226 and 77 patients, respectively. The informatics approach missed only a few cases that the traditional approaches identified, largely because those cases were known to the study team, but patient data were not in the particular children's hospital EHR. The use of informatics approaches to search electronic documentation can result in substantially larger numbers of subjects identified for studies of rare diseases such as DSD, and these approaches can be applied across hospitals.

Citation: Hanauer DA, Gardner M, Sandberg DE (2014) Unbiased Identification of Patients with Disorders of Sex Development. PLoS ONE 9(9): e108702. doi:10.1371/journal.pone.0108702

Editor: Yu-Kang Tu, National Taiwan University, Taiwan

Received March 28, 2014; **Accepted** August 29, 2014; **Published** September 30, 2014

Data Availability: The authors confirm that, for approved reasons, some access restrictions apply to the data underlying the findings. Relevant data and tables are included within the Supporting Information files. The original data included private patient clinical notes derived from each hospital's electronic health record systems. Our team was given permission to review the records through each hospital's institutional review board, but we cannot make those records publicly available, which is consistent with other studies that report on findings conducted through 'chart reviews'. These data are protected by US Laws and regulations such as the Privacy Rules of the Health Insurance Portability and Accountability Act (HIPAA) which "requires appropriate safeguards to protect the privacy of personal health information, and sets limits and conditions on the uses and disclosures that may be made of such information without patient authorization." Because of this we can only share the summary data of our findings without including details that could potentially re-identify each patient. We are willing to provide additional summary data upon request. Readers interested should contact the corresponding author, David Sandberg, PhD.

Funding: This project was supported by award nos. 3R01HD053637-02S1 and R01HD068138 (DSD–Translational Research Network) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health. This study utilized the Electronic Medical Record Search Engine (EMERSE). Support for EMERSE has come from the University of Michigan Institute for Clinical and Health Research through the Clinical Translational Science Award (UL1RR024986), as well as the University of Michigan Comprehensive Cancer Center Informatics Core with support from the National Institutes of Health (CA46592). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: dsandber@med.umich.edu

Background

Disorders of sex development (DSD) represent a prototype for rare diseases research. Patients with DSD have congenital conditions in which development of chromosomal, gonadal, or anatomic sex is atypical [1]. The various conditions subsumed under the umbrella term DSD are individually rare but, in the aggregate, have an estimated incidence of 0.1 to 0.5% of live births [2]. Substantial controversy exists regarding "best practices" in DSD; debate surrounds principles guiding gender assignment decisions, genital or gonadal surgery and their timing, hormone replacement protocols, and strategies toward educating patients and others about details of the medical condition [3–5].

A major barrier to an improved understanding of the relationships between clinical practice and outcomes include small, incomplete, or selected (i.e., "hand-picked") study samples. Indeed, the ability to generalize findings from studies of DSD patients has been limited by the lack of systematic and standardized approaches to identify patients for inclusion, and selection bias has been described as one of six "General problems of outcome studies in Disorders of Sex Development" [6]. Typical cohort ascertainment approaches include (1) clinician nomination (e.g., physician or other health care provider) or use of informal registries (e.g., patient lists) maintained by individual clinicians [7–9], (2) invitations to members of DSD peer support organizations [9,10], and (3) reviewing clinic schedules during the recruitment phase for recognizable patient names or relevant diagnoses [11]. Many published studies do not consider how these approaches might impact interpretation of the results [12,13].

Perhaps nowhere are these shortcomings more relevant than in the study of psychosocial and psychosexual outcomes in DSD-affected persons, where many of the factors that might lead a research team to select a patient for study inclusion (e.g., predicted willingness to participate, relative ease to locate, etc.) are likely correlated with psychological characteristics and thereby result in a potentially biased sample [14]. More robust and unbiased approaches are, therefore, necessary to reduce the lack of representativeness that often results from less rigorous cohort ascertainment protocols.

Electronic health records (EHRs) have the potential for supporting improved methods of identifying eligible study subjects for rare diseases, but significant challenges remain. For many rare diseases, easily extractable structured data elements (e.g., ICD-9 codes) do not provide the discrimination necessary for accurate cohort identification. For example, there is no specific ICD-9 code for DSD, and inferences must often be made based on a variety of candidate codes and other clinical attributes found in the free text narrative documents. Additionally, assignment of ICD-9 codes for a variety of disorders has been shown to be inaccurate [15–19]. This is not surprising because for many disorders, including those classified as DSD, the diagnoses are often made over time, yet the initial ICD-9 codes were assigned when diagnostic uncertainty was high. For example, it is not uncommon to find a patient with a coded diagnosis based on the initial presentation of "hypospadias," whereas the final and more accurate diagnosis of "partial androgen insensitivity syndrome" is only mentioned in the narrative clinical notes that are created subsequent to the initial encounter.

Substantial amounts of clinical data can only be found in the free text narrative portions of clinical documents [20,21]. Such documents are often created by clinicians to record the salient details of a clinical encounter and contain many details that are difficult to express using more structured data entry methods. In the case of DSD, these details can include visual descriptions of the genitalia that are easiest to express in an unstructured, free text, narrative format. They are often created via typing directly into an EHR system or through dictation and subsequent transcription [22]. These narrative documents remain the primary means of communication between clinical providers [23]. Because of their central role in communication and capture of important clinical details, these free text clinical documents must often be read to identify features necessary to make an accurate assessment about study eligibility. Accordingly, simply having the data available in electronic format does not necessarily lead to an improved capacity for cohort identification because many EHRs provide inadequate tools for improving efficiency by searching. This has resulted in an incongruity wherein the data may be available to yield a more complete sample for research, but the ability for research teams to make use of the data remains limited [24].

Computational approaches for information extraction and cohort identification can involve the use of natural language processing (NLP) [25], but the broader application of NLP is impeded by the technical expertise required for implementing the systems and the complexity of applying the algorithms across multiple institutions [26–28]. Other informatics approaches that do not involve traditional NLP may be easier to implement and can still help research teams access the data "locked" in the medical record [29–31] as well as provide reliable methods for cohort discovery in electronic clinical documentation across diverse clinical environments.

Here we describe one such approach for rare disease cohort discovery that was tested at three large academic medical centers for the purpose of identifying complete cohorts of DSD-affected patients. We show that the method yields substantially larger numbers of patients compared to traditional cohort identification approaches, and that the approach is applicable across multiple institutions using different EHRs. By doing so, we demonstrate that similar approaches are potentially within reach of any medical institution with an EHR interested in identifying representative cohorts for rare diseases research.

Methods

Study Context

As part of a broader health-related quality of life study, we sought to identify a complete cohort of patients with DSD from three large, tertiary, academic children's hospitals, referred to here as Hospitals A, B, and C. Hospital characteristics are described in Table 1. Each hospital had in place an EHR with free text clinical documents containing details about the patients that were unlikely to appear in other coded, administrative datasets. At the time the study was conducted, Hospital A was using a homegrown EHR called CareWeb which had been in use since 1998. Hospital B was using the Cerner PowerChart system, and Hospital C had clinical documents from both an Epic EHR as well as their older McKesson system which Epic had replaced. This study was approved by each hospital's local institutional review board. Waivers of informed consent and HIPAA waivers were granted by each institution's review board to allow medical records review without written informed consent. Written informed consent was sought from those eligible to participate in the broader quality of life study.

Probands in the study were determined to be eligible (or ruled ineligible) using a complex set of criteria based on information that is not routinely captured using coded or administrative data. Rather, the salient details are often captured only in the free text (i.e., dictated or typed) notes. For example, the study included patients with proximal hypospadias and either unilateral or bilateral undescended testes. To find such patients using administrative data, both hypospadias and undescended testes would have to be coded independently (which often does not occur), whereas the narrative descriptions more likely capture these details. Additionally, patients had to be between 0 and 81 months of age, without signs of significant developmental delay, and come from an English-speaking family. All institutions identified cases born during a seven-year time period. During each step of the patient identification process, study teams at each institution recorded the number of cases identified so that we could compare the capture rate of two distinct approaches for case ascertainment, described below.

Case Ascertainment – Traditional Approach

The study was initiated at Hospital A. Collaborators at Hospitals B and C were provided with a detailed guide and instructions for identifying eligible patients along with a case-ascertainment log to record various patient characteristics for further review. Both institutions were asked to identify eligible patients for the study using their traditional approaches for cohort identification.

Table 1. Hospital characteristics and results of two approaches for DSD case ascertainment at three children's hospitals.

	Hospital A	Hospital B	Hospital C
Hospital Characteristics			
Hospital beds	300	254	245
Inpatient admissions	9,000	14,000	11,000
Surgical procedures	10,000	15,000	17,000
Traditional Approach			
Cases initially identified	NP	30	16
Final chart review			
Uncertain eligibility	NP	2	1
Not eligible	NP	0	1
Confirmed eligible	NP	28	14
Informatics Approach*			
ICD-9 cases	737	2,868	3,557
ICD-9 cases + text matches	728	2,742	2,550
Initial chart review	81	153	260
Final chart review			
Uncertain eligibility	6	13	18
Not eligible	16	63	16
Confirmed eligible	59	77	226

NP = not performed.
*At Hospital A, all cases were reviewed using the search engine EMERSE and it was not necessary to run a database query for keyword matching.
doi:10.1371/journal.pone.0108702.t001

At Hospital B, a Master's level genetic counselor, and an active member of the multidisciplinary DSD clinical team, gathered names of patients referred to their clinic. At Hospital C, a Master's level research nurse coordinator, similarly a member of the local DSD clinical team, reviewed a database separate from the EHR containing a list of patients and other clinical data maintained by one subspecialty involved in the clinical management of patients with DSD. Cases in this database were identified using ICD-9 codes and then the charts were read for details about inclusion and exclusion criteria. Hospital A did not use the traditional approach because the research team recognized the limitations of these strategies and because the study team had access to specialized software for searching the clinical documents, discussed in the following section.

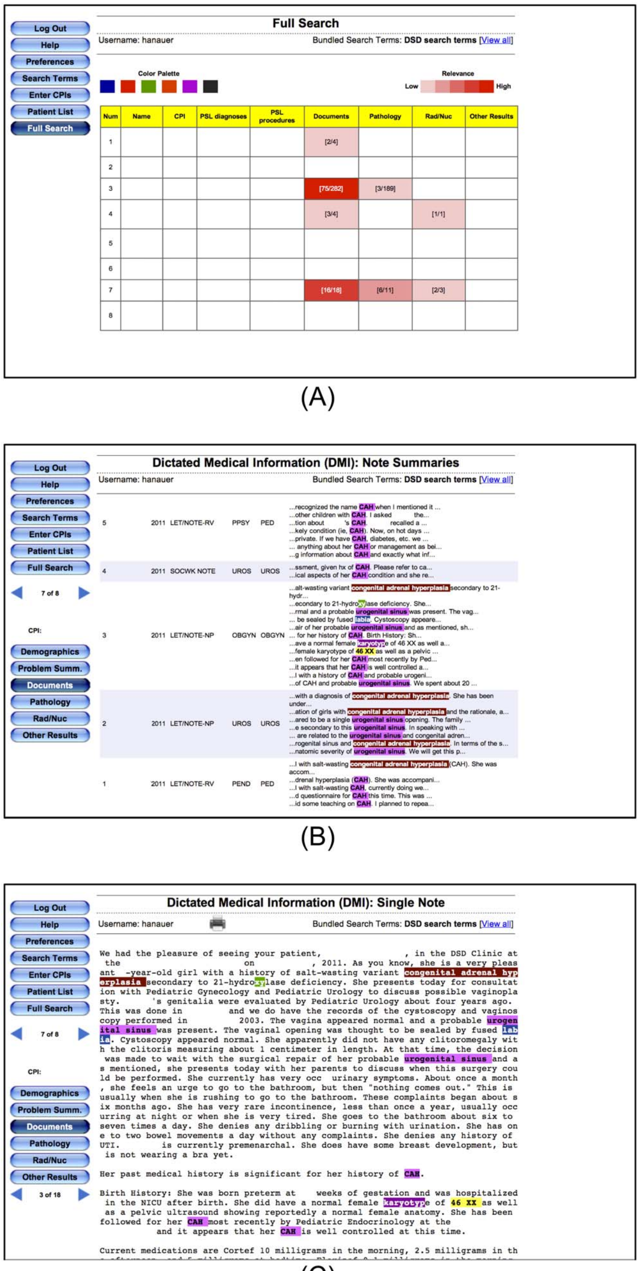
Case Ascertainment – Informatics Approach

The goal of the informatics approach was to replicate the search process supported by the search engine software (EMERSE) at Hospital A [32–35]. EMERSE, the Electronic Medical Record Search Engine, accepts a set of patient medical record numbers and a set of search terms as input. It then scans each patient's documents, and returns the results in a format ideal for chart review. The display includes highlighting all relevant terms in the documents. A typical workflow for a study utilizing EMERSE is to first identify a potential patient cohort using structured data sources including ICD-9 codes, procedure codes, clinical schedules, or existing registries. Cohort identification tools for structured data, such as the i2b2 Workbench, are already in widespread use [36]. The use of ICD-9 codes can often generate lists much larger than those obtained through traditional approaches by casting a wider net to capture patients that might otherwise be missed. The main downside is that the codes are often very non-specific and inaccurately assigned. However, these larger lists can easily be reviewed with tools such as EMERSE.

Using the patient list identified through structured data sources, a user can then enter any number of keywords or phrases into the system, and EMERSE will then search for those phrases in the free text clinical documentation. With EMERSE, the complexity of the searching is hidden from end users, who do not need to know how to use advanced computational tools to complete their work.

EMERSE has been used for a wide variety of studies that have been published in peer-reviewed journals including clinical [33,37–41] as well as translational research [42,43]. The use of EMERSE has allowed investigators at our institution to carry out many studies (both for case ascertainment and data abstraction) that were not practical prior to its implementation. One study that leveraged EMERSE to automate the identification of the postoperative surgical complications myocardial infarction and pulmonary embolus was able to achieve a sensitivity of 100% and 93%, respectively with specificities of 93% and 96%, respectively [44]. Another study that used EMERSE for eligibility determination in a depression study found significant time savings while maintaining clinical accuracy [33]. Figure 1 displays three screen shots from the EMERSE search engine as it was used for the current DSD case ascertainment study.

In addition to using their traditional approaches for case ascertainment, the two institutions that did not use EMERSE (Hospitals B and C), were asked to identify an initial cohort of potential patients using ten ICD-9 codes (Table 2). This list was intentionally broad, meant to initially favor sensitivity over specificity. Because both institutions also had clinical notes in electronic, free text format, we also provided a list of 26 search terms (Table 3) that could be used to narrow the list of eligible patients initially identified using the ICD-9 codes. That is, if any of the terms appeared in a patient's note, that patient was a potential candidate. Furthermore, these terms, when highlighted within the



Full Search

Username: hanauer Bundled Search Terms: **DSD search terms** [View all]

Num	Name	CPI	PSL diagnoses	PSL procedures	Documents	Pathology	Rad/Nuc	Other Results
1					[2/4]			
2								
3					[75/282]	[3/189]		
4					[3/4]		[1/1]	
5								
6								
7					[16/18]	[6/11]	[2/3]	
8								

Dictated Medical Information (DMI): Note Summaries

Username: hanauer Bundled Search Terms: **DSD search terms** [View all]

5	2011	LET/NOTE-RV	PPSY	PED	...recognized the name CAH when I mentioned it ... ...other children with CAH. I asked the... ...tion about 's CAH. recalled a ... ...kely condition (ie, CAH). Now, on hot days ... ...private. If we have CAH, diabetes, etc. we ... ... anything about her CAH or management as bei... ...g information about CAH and exactly what inf...
4	2011	SOCWK NOTE	UROS	UROS	...ssment, given hx of CAH. Please refer to ca... ...ical aspects of her CAH condition and she re...
3	2011	LET/NOTE-NP	OBGYN	OBGYN	...alt-wasting variant congenital adrenal hyperplasia secondary to 21-hydr... ...econdary to 21-hydroxylase deficiency. She... ...rmal and a probable urogenital sinus was present. The vag... ... be sealed by fused labia. Cystoscopy appeare... ...air of her probable urogenital sinus and as mentioned, sh... ... for her history of CAH. Birth History: Sh... ...ave a normal female karyotype of 46 XX as well a... ...female karyotype of 46 XX as well as a pelvic ... ...en followed for her CAH most recently by Ped... ...it appears that her CAH is well controlled a... ...l with a history of CAH and probable urogeni... ...of CAH and probable urogenital sinus. We spent about 20 ...
2	2011	LET/NOTE-NP	UROS	UROS	...with a diagnosis of congenital adrenal hyperplasia. She has been under... ...ation of girls with congenital adrenal hyperplasia and the rationale, a... ...ared to be a single urogenital sinus opening. The family ... ...e secondary to this urogenital sinus. In speaking with ... ... are related to the urogenital sinus and congenital adren... ...rogenital sinus and congenital adrenal hyperplasia. In terms of the s... ...natomic severity of urogenital sinus. We will get this p...
1	2011	LET/NOTE-RV	PEND	PED	...l with salt-wasting congenital adrenal hyperplasia (CAH). She was accom... ...drenal hyperplasia (CAH). She was accompani... ...l with salt-wasting CAH, currently doing we... ...d questionnaire for CAH this time. This was ... ...id some teaching on CAH. I planned to repea...

Dictated Medical Information (DMI): Single Note

Username: hanauer Bundled Search Terms: **DSD search terms** [View all]

```
We had the pleasure of seeing your patient,                , in the DSD Clinic at
 the                        on            , 2011. As you know, she is a very pleas
ant  -year-old girl with a history of salt-wasting variant congenital adrenal hyp
erplasia secondary to 21-hydroxylase deficiency. She presents today for consultat
ion with Pediatric Gynecology and Pediatric Urology to discuss possible vaginopla
sty.     's genitalia were evaluated by Pediatric Urology about four years ago.
This was done in         and we do have the records of the cystoscopy and vaginos
copy performed in          2003. The vagina appeared normal and a probable urogen
ital sinus was present. The vaginal opening was thought to be sealed by fused lab
ia. Cystoscopy appeared normal. She apparently did not have any clitoromegaly wit
h the clitoris measuring about 1 centimeter in length. At that time, the decision
 was made to wait with the surgical repair of her probable urogenital sinus and a
s mentioned, she presents today with her parents to discuss when this surgery cou
ld be performed. She currently has very occ  urinary symptoms. About once a month
, she feels an urge to go to the bathroom, but then "nothing comes out." This is
usually when she is rushing to go to the bathroom. These complaints began about s
ix months ago. She has very rare incontinence, less than once a year, usually occ
urring at night or when she is very tired. She goes to the bathroom about six to
seven times a day. She denies any dribbling or burning with urination. She has on
e to two bowel movements a day without any complaints. She denies any history of
UTI.         is currently premenarchal. She does have some breast development, but
 is not wearing a bra yet.

Her past medical history is significant for her history of CAH.

Birth History: She was born preterm at    weeks of gestation and was hospitalized
 in the NICU after birth. She did have a normal female karyotype of 46 XX as well
 as a pelvic ultrasound showing reportedly a normal female anatomy. She has been
followed for her CAH most recently by Pediatric Endocrinology at the
           and it appears that her CAH is well controlled at this time.

Current medications are Cortef 10 milligrams in the morning, 2.5 milligrams in th
```

Figure 1. Search results for the DSD terms using EMERSE at Hospital A. All identifiers have been redacted. Search results are presented in three views for assisting with rapid review. (A) The highest level displays a 'heat map' of results, with rows representing different patients and columns representing different components of the medical record. Sections with "hits" are highlighted; darker colors represent more documents with a hit. (B) Clicking on a cell in the heat map displays a summary of the documents for a specific patient, with summaries around the hits shown. Here alternating shaded rows represents a single document for a patient. (C) Clicking on a specific document summary brings up the full document with all search terms still highlighted. Note that the "xy" highlighted in 'hydroxylase' is actually a false positive result, as it was intended to identify karyotypes. Options were available in EMERSE to reduce this type of false hit, but they were not used in this study.

doi:10.1371/journal.pone.0108702.g001

documents, made chart review highly efficient. The search terms themselves were not meant to find mention of specific diagnoses (e.g., Swyer syndrome, complete androgen insensitivity syndrome, etc) but rather high-level physical and genetic testing characteristics that can help identify features consistent with a DSD. This is because many times the actual diagnosis may not be clearly described in the clinical notes but the phenotypic characteristics will provide supporting evidence for such a diagnosis. The diagnoses falling under the umbrella of DSD are complex [45], so using a more general approach without excessively specific search terms was determined by the DSD team to be of greatest value.

Both hospitals (B and C) obtained assistance from their local medical information technology (IT) teams to run the database queries and obtain the necessary data for chart review. They first used the set of ICD-9 codes, date ranges, and dates of birth to search their administrative databases to identify a large, initial cohort of potentially eligible patients. The IT teams were then instructed to further identify potential cases from this larger cohort who had at least one of the keyword terms in their notes (Table 3). They were instructed to search in a manner that matched any subset of text. Thus, a search for "karyotyp" would identify words such as "karyotype," "karyotyped," "karyotypes," and "karyotyping." Sophisticated search engines can sometimes handle these variations in word endings (referred to as stemming), but traditional database searches do not. Such database searches typically use structured query language (SQL) queries using the LIKE command for string matching (e.g., SELECT patient WHERE text LIKE '%karyotyp%'). Regular expressions, a powerful approach for complex pattern matching in text, can be used for more advanced searching in databases, but requires an additional level of expertise. To ease the review of cases, we asked that the keywords be highlighted in the documents so that the chart review could be more efficient, as was the case for EMERSE at Hospital A; this required additional computer code to be written. These keywords were displayed as snippets, or excerpts, with approximately 3 to 5 additional words on either side to provide context yet maintain brevity for rapid reviewing. An example of how these snippets from Hospital A were displayed in EMERSE can be seen in Figure 1B. Similarly, text snippets that were reviewed from data obtained from Hospital B are shown in Figure 2 for comparison. While not completely identical, the SQL used in hospitals B and C was similar to the SQL that was being used at hospital A by the EMERSE system in production use at that time. It is worth pointing out that while the ease of use of EMERSE would help users save time, our goal in this study was not to compare time efficiency but rather to compare the completeness of patient case identification using the standard approaches versus the more advanced approach that leveraged SQL.

Manual Chart Review

All cases identified via the traditional and informatics approaches were manually reviewed by trained research assistants, MG, and DES for additional inclusion and exclusion criteria. Because of the larger number of cases initially identified using the informatics approach, the chart review was performed in two stages. In the initial chart review, ineligible cases (based on requirements of the parent study) were removed such as those with developmental delay, isolated distal hypospadias, cloacal exstrophy, Turner syndrome, and Klinefelter syndrome. The final chart review required a more detailed inspection of cases including a review of a description of the genitals, laboratory values, and other diagnoses.

Table 2. ICD-9 codes used to identify an initial cohort of potentially eligible patients.

ICD-9 code	Description
752	Congenital anomalies of genital organs
752.4	Abnormalities of cervix, vagina, and external female genitalia
752.40	Unspecified anomaly of cervix, vagina, and external female genitalia
752.49	Other anomalies of cervix, vagina, and external female genitalia
752.61	Hypospadias
752.64	Micropenis
752.69	Other penile anomalies
752.7	Indeterminate sex and pseudohermaphroditism
255.2*	Adrenogenital disorders (eg, congenital adrenal hyperplasia)
259.5	Androgen Insensitivity Syndrome (AIS) (partial and complete)

*This specific code was restricted to female patients only.
doi:10.1371/journal.pone.0108702.t002

Results

The numbers of cases identified through both the traditional and informatics approaches are listed in Table 1. While the traditional approach was not used at Hospital A, this approach yielded only a relatively small number of cases at the other two hospitals, with 28 and 14 cases identified at Hospital B and C, respectively. By contrast, the number of cases ascertained using the informatics approach was substantially higher: 77 cases were identified at Hospital B using the informatics approach, representing a nearly three-fold increase in study patients. For Hospital C the informatics approach identified 226 patients eligible for the study, a 16-fold increase in study patients.

Using ICD-9 codes as an initial screen to identify patients yielded many more potential cases that had to be ruled out. At Hospital A, 8.0% of the ICD-9 cases were eventually found to be truly eligible, compared to 2.7% for Hospital B and 6.4% for Hospital C.

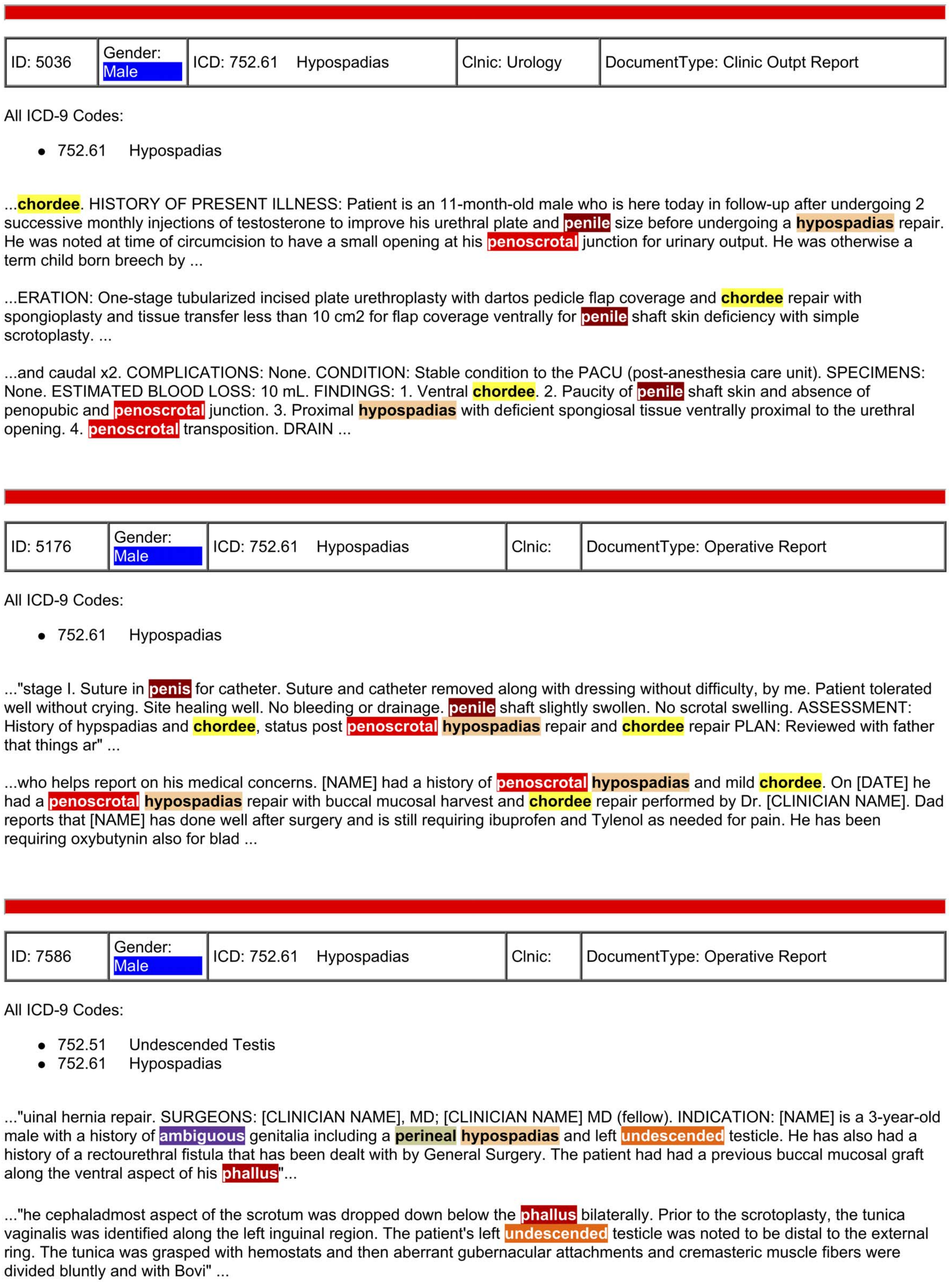
Venn Diagrams displaying the number of cases found jointly by the traditional and informatics approaches, and those cases only ascertained by one of the two approaches are shown in Figure 3. The informatics approach identified many more cases than did the traditional approach, and many of those patients would not have been identified using only the traditional approach. Conversely, the traditional approach identified a small number of cases that the informatics approach missed. At Hospital B, nine cases were identified only with the traditional approach whereas only one such case was identified with the traditional approach at Hospital C.

Additional details about the DSD cases that were verified as having a true DSD at both Hospitals B and C are provided as Supporting Information tables. Tables S1, S2, S3, S4, S5, and S6 report the combined counts for both Hospitals B and C. Tables S7, S8, S9, S10, S11 and S12 report the counts for Hospital B, and Tables S13, S14, S15, S16, S17 and S18 reports the counts for

Table 3. Keywords used to highlight relevant information in the clinical documents identified from the ICD-9 search.

DSD Keywords	
46 XX	CAH
46 XY	congenital adrenal hyperplasia
46-XX	gonad
46-XY	hypospad
XO	labia
XX	mosaic
XY	penile
ambig	penis
chordee	phall
karyotyp	prader
penoscrotal	urogenital sinus
perineal	viriliz
severe hypospad	
undescended	

Keywords and phrases in the left column were used in conjunction with the ICD-9 code (752.61) for hypospadias; whereas those in both columns were used in conjunction with the balance (ie, non-hypospadias) of ICD-9 codes listed in Table 2. Note that because of the searching process, a keyword search for a term such as ''gonad'' would also identify terms such as ''gonads'', ''gonadic'', and ''gonadal''.
doi:10.1371/journal.pone.0108702.t003



ID: 5036	Gender: Male	ICD: 752.61 Hypospadias	Clnic: Urology	DocumentType: Clinic Outpt Report

All ICD-9 Codes:

- 752.61 Hypospadias

...chordee. HISTORY OF PRESENT ILLNESS: Patient is an 11-month-old male who is here today in follow-up after undergoing 2 successive monthly injections of testosterone to improve his urethral plate and penile size before undergoing a hypospadias repair. He was noted at time of circumcision to have a small opening at his penoscrotal junction for urinary output. He was otherwise a term child born breech by ...

...ERATION: One-stage tubularized incised plate urethroplasty with dartos pedicle flap coverage and chordee repair with spongioplasty and tissue transfer less than 10 cm2 for flap coverage ventrally for penile shaft skin deficiency with simple scrotoplasty. ...

...and caudal x2. COMPLICATIONS: None. CONDITION: Stable condition to the PACU (post-anesthesia care unit). SPECIMENS: None. ESTIMATED BLOOD LOSS: 10 mL. FINDINGS: 1. Ventral chordee. 2. Paucity of penile shaft skin and absence of penopubic and penoscrotal junction. 3. Proximal hypospadias with deficient spongiosal tissue ventrally proximal to the urethral opening. 4. penoscrotal transposition. DRAIN ...

ID: 5176	Gender: Male	ICD: 752.61 Hypospadias	Clnic:	DocumentType: Operative Report

All ICD-9 Codes:

- 752.61 Hypospadias

..."stage I. Suture in penis for catheter. Suture and catheter removed along with dressing without difficulty, by me. Patient tolerated well without crying. Site healing well. No bleeding or drainage. penile shaft slightly swollen. No scrotal swelling. ASSESSMENT: History of hypspadias and chordee, status post penoscrotal hypospadias repair and chordee repair PLAN: Reviewed with father that things ar" ...

...who helps report on his medical concerns. [NAME] had a history of penoscrotal hypospadias and mild chordee. On [DATE] he had a penoscrotal hypospadias repair with buccal mucosal harvest and chordee repair performed by Dr. [CLINICIAN NAME]. Dad reports that [NAME] has done well after surgery and is still requiring ibuprofen and Tylenol as needed for pain. He has been requiring oxybutynin also for blad ...

ID: 7586	Gender: Male	ICD: 752.61 Hypospadias	Clnic:	DocumentType: Operative Report

All ICD-9 Codes:

- 752.51 Undescended Testis
- 752.61 Hypospadias

..."uinal hernia repair. SURGEONS: [CLINICIAN NAME], MD; [CLINICIAN NAME] MD (fellow). INDICATION: [NAME] is a 3-year-old male with a history of ambiguous genitalia including a perineal hypospadias and left undescended testicle. He has also had a history of a rectourethral fistula that has been dealt with by General Surgery. The patient had had a previous buccal mucosal graft along the ventral aspect of his phallus"...

..."he cephaladmost aspect of the scrotum was dropped down below the phallus bilaterally. Prior to the scrotoplasty, the tunica vaginalis was identified along the left inguinal region. The patient's left undescended testicle was noted to be distal to the external ring. The tunica was grasped with hemostats and then aberrant gubernacular attachments and cremasteric muscle fibers were divided bluntly and with Bovi" ...

Figure 2. Search results for the DSD terms at Hospital B. These snippets of text with DSD keywords highlighted are from three patients that were identified using the generic ICD-9 code 752.61 (hypospadias). By reviewing the highlighted text, the study team was able to determine that these patients were likely to exhibit a DSD rather than isolated hypospadias.

doi:10.1371/journal.pone.0108702.g002

Hospital C. For each group of tables, the counts are broken down by various regions of the Venn diagrams as described in the manuscript.

One case of complete androgen insensitivity syndrome (CAIS) was identified using the traditional approach without the use of an ICD-9 code, so this was not detected using the informatics

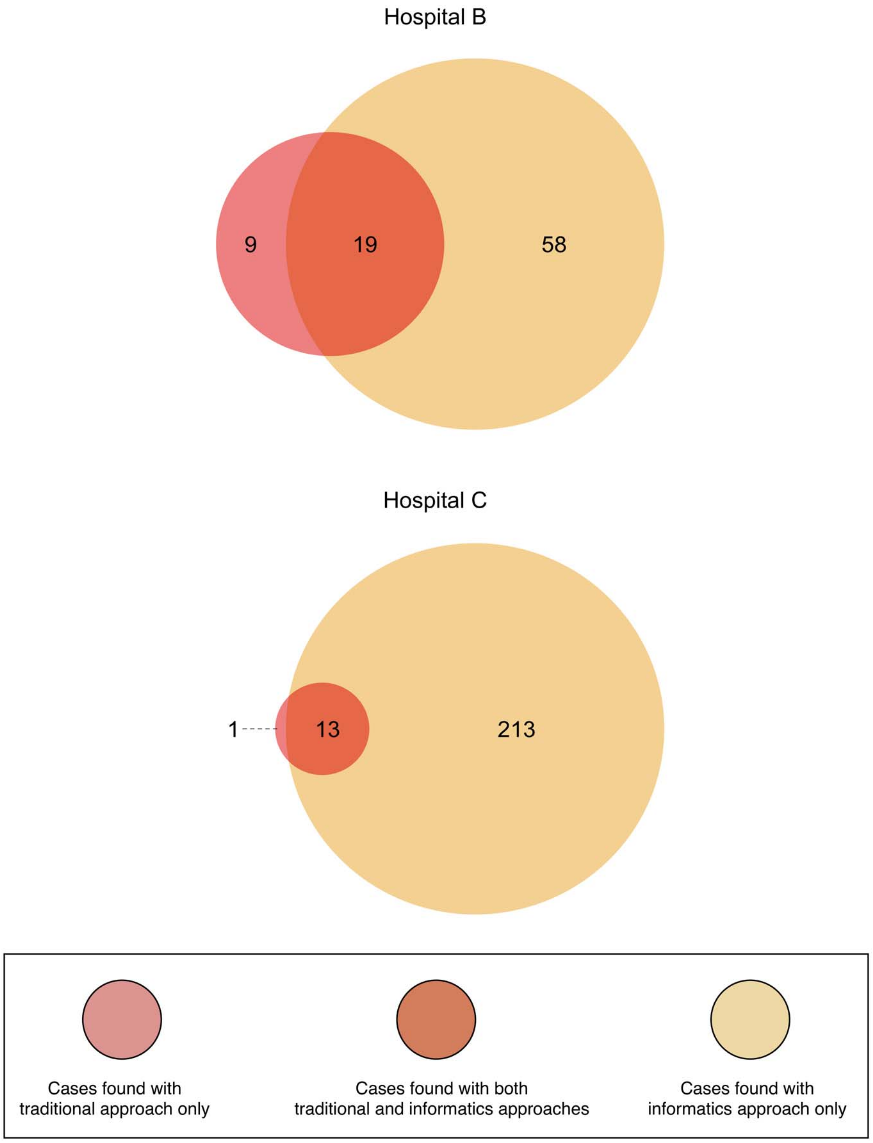


Figure 3. Cases of DSD found using the traditional and informatics approaches. Cases of DSD found using the traditional and informatics approaches, showing the overlap of cases found jointly by both approaches and those found distinctly with either approach. Additional details about the cases can be found in the Supporting Information tables.
doi:10.1371/journal.pone.0108702.g003

approach. For all of the other codes, the informatics approach found the same number or more cases (Table S19), ranging from 1 case each of ICD-9 code 752.4 (Unspecified Congenital Anomaly of Cervix, Vagina, and External Female Genitalia) via both the traditional and informatics approaches to a 25-fold increase in the number of cases of the non-specific ICD-9 code 752.69 (Other Penile Anomalies) found with the informatics approach.

Discussion

Rare diseases can often be heterogeneous in their presentation [46–48], and it can be challenging to identify affected persons in a comprehensive and systematic manner [49]. Thus, the difficulty in obtaining adequate patient cohorts may lead to less rigorous recruitment practices which, itself, can introduce bias [50]. Case ascertainment in DSD has historically been challenging due to a lack of standardized definitions of what actually constitutes a DSD. Disagreements about definitions persist [51,52], suggesting that additional focus should be placed on lessening the impact of other shortcomings in the cohort identification process that could lead to non-representative samples. As registries are being developed to comprehensively capture cases of rare diseases, including DSD [53], it will become increasingly important to ensure that the patients are included using unbiased strategies [54].

The traditional approaches for case identification in the multi-institutional collaboration described in the current study were varied, but seemed to mimic traditional manual searches through paper charts, which are neither efficient nor comprehensive. But when we standardized the approaches by using techniques that are enabled through the use of EHRs, the number of patients identified was substantially higher. This suggests that while many diseases are rare, they might not always be as rare as it may seem when a more comprehensive approach is used to identify patients. Further, this more inclusive approach is likely to increase statistical power from larger study samples and reduce selection bias that may occur when single clinics or clinicians "hand-pick" study participants. A strength of our study was that we demonstrated the effectiveness of the informatics approach using clinical notes generated through multiple EHRs. Thus, even though there may be variations in hospitals, or even among clinical groups with respect to documentation practices and conventions, the informatics approach was still able to identify more cases than the traditional approach.

ICD-9 codes are often used, but insufficient, for identifying rare diseases [19], in large part because there is not a specific code for each disease or phenotype. This was also evident in our study where we initially used ICD-9 codes but then reduced the many candidate cases through additional keyword searches. The transition in the U.S. from ICD-9 to ICD-10 codes has the potential to improve the way diagnoses are coded, but it remains to be seen how precise clinicians will be when using the codes. For example, the ICD-10 code Q54.2 ("hypospadias, penoscrotal") would potentially be indicative of a patient with a DSD, but if clinicians choose to use the more generic and much less specific code Q54.9 ("hypospadias, unspecified"), then further review of the records will still be required to confirm that the case meets definitional requirements of the category DSD. In our current study, we found patients with hypospadias to be very challenging in terms of identifying DSD cases since it has such a broad phenotype. A recent study conducted in Norway (where ICD-10 has been in use since 1997) [55] used a combination of ICD codes to identify patients but still required a manual chart review to verify and systematically identify all patients with congenital adrenal hyperplasia [56]. This should not be surprising given the known challenges of accurately assigning codes for both the ICD-9 [16–18] and ICD-10 systems [57–63]. One study specifically acknowledged that "the implementation of ICD-10 coding has not significantly improved the quality of administrative data relative to ICD-9-CM." [64].

The keywords used in our study did not by themselves narrow down the cases to a significant extent, but by highlighting them in the text it allowed our reviewers to focus on those specific sections of text and allowed them to rapidly identify the key concepts to help make the eligibility determination. That is why the number of cases identified with ICD-9 alone did not differ much from those identified with ICD-9 plus the text matches. However, the review of the cases that occurred because of the text matches allowed for the cases to be effectively narrowed to only appropriate ones. Additionally, because we were able to identify and highlight keywords across all documents for each patient, we were able to ensure a more comprehensive approach to chart review while maintaining efficiency. Future work should explore the use of negation or exclusions (e.g., not highlighting 'labia' in the context of 'labia were normal in appearance') as a way to potentially reduce the number of false positive terms highlighted. EMERSE already supports a simple version of exclusions, but it was not used in this study. Additionally, comparison of this informatics-based, but relatively simple, approach to more advanced computational natural language processing techniques would be worthwhile.

While our informatics approach increased the number of cases identified, there were a small number of cases that were missed and were only identified using the traditional approaches. One patient, for example, was seen only at an affiliated satellite clinic that had its own EHR, and had never been seen at the main hospital from which the study recruitment was being conducted. The clinical DSD team was nonetheless aware of the patient and identified them for inclusion. This raises an issue that even when using comprehensive approaches for identifying patients at large health centers, additional efforts may be required to comprehensively identify the overall population of patients with rare diseases in a given region.

While the three study settings were not precisely comparable in size and volume of procedures, all were full tertiary care children's hospitals with comprehensive specialty services including both medical and surgical services. All three have specialty services that diagnose and treat patients with DSD, and conduct research on DSD. It should be noted that at Hospital A, we used a well-established tool, EMERSE, that had functionality we attempted to replicate at the other two institutions, and with modest effort we were able to implement some of the core features of EMERSE at both institutions, yielding a much larger number of cases identified. EMERSE provides a simple user interface that makes running searches quick and efficient, but our goal in this paper was to focus on the portability and efficacy of this approach at other institutions rather than the usability or speed of EMERSE itself. The method proved to be easily reproducible at both other institutions with help from the IT teams. Further, if EMERSE itself had been installed at the two other institutions, the analyses could have been carried out without any help from the IT groups once the software installation was complete.

The version of EMERSE used at Hospital A for the study described here has been replaced by a newer version that integrates data from an older, locally developed EHR system as well as clinical documents from the new vendor system (Epic). EMERSE is available at no cost for academic use at other institutions, and interested investigators can contact the authors for additional information about the system. A demonstration version of the system is available at http://project-emerse.org.

Our approach for case ascertainment should be applicable to other rare disease research, and may help standardize the way in which cases are identified across multiple institutions. To achieve this goal, we suggest that future studies provide additional details describing their case ascertainment approaches to reduce the likelihood that selection biases are responsible for variability in results across studied populations. Detailed reporting of methodology is already required for many clinical trials using frameworks such as the Consolidated Standards of Reporting Trials (CONSORT) [65]. One way this type of reporting could be improved would be for investigators to provide the list of keywords or search terms used (Table 3). Indeed, other studies that have used similar approaches have listed the search terms used [31,33,38,40,41] and some authors have noted the benefits of using systems such as EMERSE to ensure standardized and reproducible results by searching in a systematic and unbiased manner [66,67].

Conclusion

To the best of our knowledge, this is the first time that an informatics-based approach has been used for the systematic identification of cases of DSD in an electronic health record. The results of our study demonstrate that traditional approaches for case ascertainment in DSD are limited and may introduce bias. The use of EHRs alone may not lead to a reduction in this bias unless additional strategies are utilized with these systems to more comprehensively identify patients for study inclusion. Applying more rigorous and reproducible informatics-based approaches for case ascertainment is now feasible and should be considered as research teams recruit patients for studies of rare diseases.

Supporting Information

Table S1 Hospital B + Hospital C: Patients identified by all methods.
(PDF)

Table S2 Hospital B + Hospital C: Patients identified by Standard Method.
(PDF)

Table S3 Hospital B + Hospital C: Patients identified by Informatics.
(PDF)

Table S4 Hospital B + Hospital C: Patients identified only by Standard Method.
(PDF)

Table S5 Hospital B + Hospital C: Patients identified only by Informatics.
(PDF)

Table S6 Hospital B + Hospital C: Patients identified by Informatics and Standard Method.
(PDF)

Table S7 Hospital B: Patients identified by all methods.
(PDF)

Table S8 Hospital B: Patients identified by Standard Method.
(PDF)

Table S9 Hospital B: Patients identified by Informatics.
(PDF)

Table S10 Hospital B: Patients identified only by Standard Method.
(PDF)

Table S11 Hospital B: Patients identified only by Informatics.
(PDF)

Table S12 Hospital B: Patients identified by Informatics and Standard Method.
(PDF)

Table S13 Hospital C: Patients identified by all methods.
(PDF)

Table S14 Hospital C: Patients identified by Standard Method.
(PDF)

Table S15 Hospital C: Patients identified by Informatics.
(PDF)

Table S16 Hospital C: Patients identified only by Standard Method.
(PDF)

Table S17 Hospital C: Patients identified only by Informatics.
(PDF)

Table S18 Hospital C: Patients identified by Informatics and Standard Method.
(PDF)

Table S19 Collapsed/merged ICD-9 codes for Hospitals B and C combined.
(PDF)

Acknowledgments

The authors wish to thank: Barry A. Kogan, Alexandra L. Quittner, Kalina Benedict, Mirranda Boshart, Karl Bosse, Antonio Chaviano, Lynn Holevinski, Bruce Jackson, Patricia Logan, Elizabeth McCauley, Elana Mendelowitz, Theresa Meyer, Michael Miller, Linda Ramsdell, Jacqueline Rau, George Sales, and Claire-Marie Varga.

Author Contributions

Conceived and designed the experiments: DAH MG DES. Performed the experiments: DAH MG DES. Analyzed the data: DAH MG DES. Contributed reagents/materials/analysis tools: DAH MG DES. Contributed to the writing of the manuscript: DAH MG DES.

References

1. Lee PA, Houk CP, Ahmed SF, Hughes IA (2006) Consensus statement on management of intersex disorders. International Consensus Conference on Intersex. Pediatrics 118: e488–500.
2. Arboleda VA, Lee H, Sanchez FJ, Delot EC, Sandberg DE, et al. (2013) Targeted massively parallel sequencing provides comprehensive genetic diagnosis for patients with disorders of sex development. Clin Genet 83: 35–43.
3. Magritte E (2012) Working together in placing the long term interests of the child at the heart of the DSD evaluation. J Pediatr Urol 8: 571–575.
4. Sandberg DE, Gardner M, Cohen-Kettenis PT (2012) Psychological aspects of the treatment of patients with disorders of sex development. Semin Reprod Med 30: 443–452.
5. Tamar-Mattis A, Baratz A, Baratz Dalke K, Karkazis K (2013) Emotionally and cognitively informed consent for clinical care for differences of sex development. Psychology and Sexuality: 1–12.
6. Lux A, Kropf S, Kleinemeier E, Jurgensen M, Thyen U (2009) Clinical evaluation study of the German network of disorders of sex development (DSD)/intersexuality: study design, description of the study population, and data quality. BMC Public Health 9: 110.
7. Jurgensen M, Hiort O, Holterhus PM, Thyen U (2007) Gender role behavior in children with XY karyotype and disorders of sex development. Horm Behav 51: 443–453.
8. Michala L, Goswami D, Creighton SM, Conway GS (2008) Swyer syndrome: presentation and outcomes. BJOG 115: 737–741.
9. Wolfe-Christensen C, Fedele DA, Kirk K, Mullins LL, Lakshmanan Y, et al. (2013) Caregivers of children with a disorder of sex development: associations between parenting capacities and psychological distress. Journal of Pediatric Urology.
10. Brinkmann L, Schuetzmann K, Richter-Appelt H (2007) Gender assignment and medical history of individuals with different forms of intersexuality: evaluation of medical records and the patients' perspective. J Sex Med 4: 964–980.
11. Parisi MA, Ramsdell LA, Burns MW, Carr MC, Grady RE, et al. (2007) A Gender Assessment Team: experience with 250 patients over a period of 25 years. Genet Med 9: 348–357.
12. Hullmann SE, Fedele DA, Wolfe-Christensen C, Mullins LL, Wisniewski AB (2011) Differences in adjustment by child developmental stage among caregivers of children with disorders of sex development. Int J Pediatr Endocrinol 2011: 16.
13. Kohler B, Kleinemeier E, Lux A, Hiort O, Gruters A, et al. (2012) Satisfaction with genital surgery and sexual life of adults with XY disorders of sex development: results from the German clinical evaluation study. J Clin Endocrinol Metab 97: 577–588.
14. Rosenthal R, Rosnow RL (2009) Artifacts in behavioral research: Robert Rosenthal and Ralph L. Rosnow's classic books: a re-issue of Artifact in behavioral research, Experimenter effects in behavioral research and The volunteer subject. New York: Oxford University Press. xv, 886 p.
15. Cooke CR, Joo MJ, Anderson SM, Lee TA, Udris EM, et al. (2011) The validity of using ICD-9 codes and pharmacy records to identify patients with chronic obstructive pulmonary disease. BMC Health Serv Res 11: 37.
16. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, et al. (2005) Measuring diagnoses: ICD code accuracy. Health Serv Res 40: 1620–1639.
17. Reker DM, Rosen AK, Hoenig H, Berlowitz DR, Laughlin J, et al. (2002) The hazards of stroke case selection using administrative data. Med Care 40: 96–104.
18. Rhodes ET, Laffel LM, Gonzalez TV, Ludwig DS (2007) Accuracy of administrative coding for type 2 diabetes in children, adolescents, and young adults. Diabetes Care 30: 141–143.
19. Sickbert-Bennett EE, Weber DJ, Poole C, MacDonald PD, Maillard JM (2010) Utility of International Classification of Diseases, Ninth Revision, Clinical Modification codes for communicable disease surveillance. Am J Epidemiol 172: 1299–1305.
20. Morrison Z, Fernando B, Kalra D, Cresswell K, Sheikh A (2013) National evaluation of the benefits and risks of greater structuring and coding of the

electronic health record: exploratory qualitative investigation. J Am Med Inform Assoc.

21. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, et al. (2011) Data from clinical notes: a perspective on the tension between structure and flexible documentation. J Am Med Inform Assoc 18: 181–186.
22. Zheng K, Mei Q, Yang L, Manion FJ, Balis UJ, et al. (2011) Voice-dictated versus typed-in clinician notes: linguistic properties and the potential implications on natural language processing. AMIA Annu Symp Proc 2011: 1630–1638.
23. Embi PJ, Weir C, Efthimiadis EN, Thielke SM, Hedeen AN, et al. (2013) Computerized provider documentation: findings and implications of a multisite study of clinicians and administrators. J Am Med Inform Assoc 20: 718–726.
24. Christensen T, Grimsmo A (2008) Instant availability of patient records, but diminished availability of patient information: a multi-method study of GP's use of electronic patient records. BMC Med Inform Decis Mak 8: 12.
25. Jha AK (2011) The promise of electronic records: around the corner or down the road? JAMA 306: 880–881.
26. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF (2008) Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform: 128–144.
27. Stanfill MH, Williams M, Fenton SH, Jenders RA, Hersh WR (2010) A systematic literature review of automated clinical coding and classification systems. J Am Med Inform Assoc 17: 646–651.
28. Turchin A, Kolatkar NS, Grant RW, Makhni EC, Pendergrass ML, et al. (2006) Using regular expressions to abstract blood pressure and treatment intensification information from the text of physician notes. J Am Med Inform Assoc 13: 691–695.
29. Abhyankar S, Demner-Fushman D, Callaghan FM, McDonald CJ (2014) Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis. J Am Med Inform Assoc.
30. Ford E, Nicholson A, Koeling R, Tate AR, Carroll J, et al. (2013) Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text? BMC Med Res Methodol 13: 105.
31. Singh B, Singh A, Ahmed A, Wilson GA, Pickering BW, et al. (2012) Derivation and validation of automated electronic search strategies to extract Charlson comorbidities from electronic medical records. Mayo Clin Proc 87: 817–824.
32. Hanauer DA (2006) EMERSE: The Electronic Medical Record Search Engine. AMIA Annu Symp Proc: 941.
33. Seyfried L, Hanauer DA, Nease D, Albeiruti R, Kavanagh J, et al. (2009) Enhanced identification of eligibility for depression research using an electronic medical record search engine. Int J Med Inform 78: e13–18.
34. Yang L, Mei Q, Zheng K, Hanauer DA (2011) Query log analysis of an electronic health record search engine. AMIA Annu Symp Proc 2011: 915–924.
35. Zheng K, Mei Q, Hanauer DA (2011) Collaborative search in electronic health records. J Am Med Inform Assoc 18: 282–291.
36. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, et al. (2010) Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 17: 124–130.
37. Al-Holou WN, Terman SW, Kilburg C, Garton HJ, Muraszko KM, et al. (2011) Prevalence and natural history of pineal cysts in adults. J Neurosurg 115: 1106–1114.
38. Asmar R, Beebe-Dimmer JL, Korgavkar K, Keele GR, Cooney KA (2013) Hypertension, obesity and prostate cancer biochemical recurrence after radical prostatectomy. Prostate Cancer Prostatic Dis 16: 62–66.
39. Hanauer DA, Ramakrishnan N, Seyfried LS (2013) Describing the relationship between cat bites and human depression using data from an electronic health record. PLoS One 8: e70585.
40. Jensen KM, Davis MM (2013) Health care in adults with Down syndrome: a longitudinal cohort study. J Intellect Disabil Res 57: 947–958.
41. Patrick SW, Davis MM, Sedman AB, Meddings JA, Hieber S, et al. (2013) Accuracy of hospital administrative data in reporting central line-associated bloodstream infections in newborns. Pediatrics 131 Suppl 1: S75–80.
42. Choi SW, Stiff P, Cooke K, Ferrara JL, Braun T, et al. (2012) TNF-inhibition with etanercept for graft-versus-host disease prevention in high-risk HCT: lower TNFR1 levels correlate with better outcomes. Biol Blood Marrow Transplant 18: 1525–1532.
43. Paczesny S, Braun TM, Levine JE, Hogan J, Crawford J, et al. (2010) Elafin is a biomarker of graft-versus-host disease of the skin. Sci Transl Med 2: 13ra12.
44. Hanauer DA, Englesbe MJ, Cowan JA, Jr., Campbell DA (2009) Informatics and the American College of Surgeons National Surgical Quality Improvement Program: automated processes could replace manual record review. J Am Coll Surg 208: 37–41.
45. Houk CP, Hughes IA, Ahmed SF, Lee PA (2006) Summary of consensus statement on intersex disorders and their management. International Intersex Consensus Conference. Pediatrics 118: 753–757.
46. Aaronson IA (2011) Terminology for disorders of sex development: clarity or confusion? J Urol 185: 388–389.
47. Groman JD, Meyer ME, Wilmott RW, Zeitlin PL, Cutting GR (2002) Variant cystic fibrosis phenotypes in the absence of CFTR mutations. N Engl J Med 347: 401–407.
48. Knowles MR, Durie PR (2002) What is cystic fibrosis? N Engl J Med 347: 439–442.
49. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, et al. (2009) Clinical research for rare disease: opportunities, challenges, and solutions. Mol Genet Metab 96: 20–26.
50. Lilford RJ, Thornton JG, Braunholtz D (1995) Clinical trials and rare diseases: a way out of a conundrum. BMJ 311: 1621–1625.
51. Wit JM, Ranke MB, Kelnar CJH (2007) Disorders of Sex Development (Dsd). Hormone Research 68 (Supplement 2): 21–26.
52. Arboleda VA, Sandberg DE, Vilain E (2014) DSDs: genetics, underlying pathologies and psychosexual differentiation. Nat Rev Endocrinol In Press.
53. Hiort O, Wunsch L, Cools M, Looijenga L, Cuckow P (2012) Requirements for a multicentric multidisciplinary registry on patients with disorders of sex development. J Pediatr Urol 8: 624–628.
54. Cox K, Bryce J, Jiang J, Rodie M, Sinnott R, et al. (2013) Novel associations in disorders of sex development: findings from the I-DSD Registry. J Clin Endocrinol Metab.
55. Povl Munk-Jørgensen AB (1999) Implementation of ICD-10 in the Nordic countries. Nordic Journal of Psychiatry 53: 5–9.
56. Nermoen I, Husebye ES, Svartberg J, Lovas K (2010) Subjective health status in men and women with congenital adrenal hyperplasia: a population-based survey in Norway. Eur J Endocrinol 163: 453–459.
57. Casez P, Labarere J, Sevestre MA, Haddouche M, Courtois X, et al. (2010) ICD-10 hospital discharge diagnosis codes were sensitive for identifying pulmonary embolism but not deep vein thrombosis. J Clin Epidemiol 63: 790–797.
58. Farzandipour M, Sheikhtaheri A, Sadoughi F (2010) Effective factors on accuracy of principal diagnosis coding based on International Classification of Diseases, the 10th revision (ICD-10). International Journal of Information Management 30: 78–84.
59. Langner I, Mikolajczyk R, Garbe E (2011) Regional and temporal variations in coding of hospital diagnoses referring to upper gastrointestinal and oesophageal bleeding in Germany. BMC Health Services Research 11: 193.
60. Nilsson G, Petersson H, Ahlfeldt H, Strender LE (2000) Evaluation of three Swedish ICD-10 primary care versions: reliability and ease of use in diagnostic coding. Methods Inf Med 39: 325–331.
61. Quach S, Blais C, Quan H (2010) Administrative data have high variation in validity for recording heart failure. Can J Cardiol 26: 306–312.
62. Stausberg J, Lehmann N, Kaczmarek D, Stein M (2008) Reliability of diagnoses coding with ICD-10. International Journal of Medical Informatics 77: 50–57.
63. Wockenfuss R, Frese T, Herrmann K, Claussnitzer M, Sandholzer H (2009) Three- and four-digit ICD-10 is not a reliable classification system in primary care. Scand J Prim Health Care 27: 131–136.
64. Quan H, Li B, Saunders LD, Parsons GA, Nilsson CI, et al. (2008) Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. Health Serv Res 43: 1424–1441.
65. Schulz KF, Altman DG, Moher D (2010) CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 8: 18.
66. DeBenedet AT, Saini SD, Takami M, Fisher LR (2011) Do clinical characteristics predict the presence of small bowel angioectasias on capsule endoscopy? Dig Dis Sci 56: 1776–1781.
67. DiMagno MJ, Spaete JP, Ballard DD, Wamsteker EJ, Saini SD (2013) Risk models for post-endoscopic retrograde cholangiopancreatography pancreatitis (PEP): smoking and chronic liver disease are predictors of protection against PEP. Pancreas 42: 996–1003.