Curr Urol Rep (2017) 18: 68
DOI 10.1007/s11934-017-0714-7

MEN'S HEALTH (A DABAJA, SECTION EDITOR)

Validity of Claims Data for the Identification of Male Infertility

Yash S. Khandwala[1,2] · **Chiyuan A. Zhang**[1] · **Shufeng Li**[3] · **Mark R. Cullen**[4] · **Michael L. Eisenberg**[1]

Published online: 17 July 2017

Abstract
Purpose of Review We sought to determine whether infertile men can accurately be identified within a large insurance claims database to validate its use for reproductive health research.
Recent Findings Prior literature suggests that men coded for infertility are at higher risk for chronic disease though it was previously unclear if these diagnostic codes correlated with true infertility. We found that the specificity of one International Classification of Disease (9th edition) code in predicting abnormal semen parameters was 92.4%, rising to 99.8% if a patient had three different codes for infertility. The positive predictive value was as high as 85%.
Summary The use of claims data for male infertility research has been rapidly progressing due to its high power and feasibility. The high specificity of ICD codes for men with abnormal semen parameters is reassuring and validates prior studies as well as future investigation into men's health.

Keywords Male infertility · Claims data · Epidemiology · Vasectomy · Epidemiology

This article is part of the Topical Collection on *Men's Health*

✉ Michael L. Eisenberg
eisenberg@stanford.edu

[1] Department of Urology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305-5118, USA

[2] University of California San Diego School of Medicine, La Jolla, CA, USA

[3] Departments of Urology and Dermatology, Stanford University School of Medicine, Stanford, CA, USA

[4] Department of Internal Medicine, Stanford University School of Medicine, Stanford, CA, USA

Introduction

Claims data from large healthcare databases increasingly serve as an important source of big data for epidemiological studies. They have served as a cost effective alternative to the administration of patient surveys and arduous chart review [1]. Administrative datasets provide access to the health status of a large, comprehensive group of individuals over an extended period of time that would otherwise be difficult to obtain [2]. Furthermore, the accrual, synthesis, and analysis of this data have the potential to contribute significantly to the contemporary literature on the epidemiology and outcomes of common medical diseases.

The first evidence of claims data utilization for health science research was published in 1979. Over the ensuing two decades, paralleling the rapid advances in computing and data management, insurers such as Medicare and Medicaid began organizing and studying troves of health care data that they had accumulated. Primarily, they were interested in quantifying prevalence of disease, management trends, and overall quality of care with a special focus on cost analyses [3–5]. Recently, administrative data have been used more broadly for disease risk characterization, analysis of morbidity and mortality, hospital length of stay, and complication rates [4, 6, 7].

Claims Data and Chronic Medical Disease

Historically, there has been little confidence by clinicians in the reliability of hospital discharge data for the study of chronic medical diagnoses and procedures. Inconsistencies in coding due to the ambiguity of the International Classification of Disease 9th edition (ICD-9) combined with potential coding errors raised concern over a possible lack of specificity of insurance codes. Each diagnosis or comorbidity may be

ascribed one or more possible codes thus conceivably introducing a type of selection bias. These errors were thought to be systematic and often associated with a secondary goal of maximizing reimbursements [2]. Due to these concerns, researchers attempted to validate the specificity and sensitivity of diagnostic codes in identifying and studying prevalent diseases with varying results.

Utilizing primarily insurance data from Medicare, studies in the 1970s and 1980s concluded that the accuracy of claims data was adequate for only a select number of conditions but the "percentage of agreement" between diagnostic codes and true prevalence of disease was increasing (73.2% in 1977 to 78.2% in 1985) [4, 8]. Thus, over the ensuing decades researchers began validating the use of claims data for each disease individually. In the early 2000s, Lee et al. found that the positive predictive value of insurance codes for heart failure and its comorbidities was above 94% [9]. Steele et al. discovered a similarly effective use of claims data for the measurement of mental health services provided in an ambulatory setting [10]. In 2013, Tessier-Sherman et al. successfully compared the diagnosis of hypertension in occupational medical charts from a US manufacturing company against diagnoses obtained via administrative claims data and found that medical service claims data were highly specific for identifying hypertension [11]. In fact, all 18 diseases represented by the Charlson comorbidity index were found to be significantly specific to insurance codes contained within claims databases [12]. Indeed, there still remains a subset of uninsured patients that is missed when insurance claims are used as the primary source of data thus limiting the sensitivity of this methodology.

Claims Data and Male Infertility

Notably absent among the litany of claims data validation literature is an evaluation of the utility of insurance codes for male factor infertility research. Nearly one in six couples are unable to conceive after at least 1 year of regular sexual intercourse and the male is responsible in at least 30% of the cases [13, 14]. Alarmingly, fecundity may actually be decreasing as men have been producing lower quality semen than their predecessors—a significant public health and anthropologic predicament [14–16]. Furthermore, infertile men have been found to be at increased risk for chronic medical conditions as well as overall mortality [17••, 18, 19]. In 2015, we found that men with an infertility diagnostic code (ICD-9606.x, V26.21) or a procedure code (CPT) suggesting fertility testing were significantly more likely to develop cancer compared to controls [18].

Yet, while claims data has been legitimatized for the study of many chronic and prevalent diseases, its usefulness for infertility remains undetermined. Outside of our work and a paper published in 2007 by Meacham et al. characterizing prevalence and cost of male infertility, men's reproductive health research has relied mostly on small prospective trials or extensive chart review [20]. One major drawback of studies using insurance data has been the exclusion of fertility treatment from insurer reimbursement packages thus disincentivizing providers from accurately coding fertility diagnoses. However, recent legislature such as the Family Building Act of 2005 and 2009 as well as the Medicare Infertility Coverage Act of 2003 and 2005 likely have contributed to improvement in both paternal and maternal reporting [21, 22]. Still, a validation of the reliability of diagnostic codes for predicting true infertility in medical service claims would legitimize prior findings and support future studies on infertile men using claims data.

Specificity of Claims Data for Infertility

Using our institute's internal claims database, STRIDE (Stanford Translational Research Integrated Database Environment), we evaluated whether the diagnosis of male factor infertility accurately identifies men with abnormal semen parameters. We identified men who had a prior semen analysis and collected associated ICD-9 (International Classification of Disease 9th edition) codes—azoospermia (606.0), oligospermia (606.1), infertility due to extratesticular causes (606.8), and unspecified male infertility (606.9). Given that most men (>99%) were evaluated prior to the WHO 5th edition, we used the criteria from the 4th edition manual on semen analysis published in 1999 to define abnormalities of reported semen analyses: volume < 2.0 mL, concentration < 20×10^6/mL, total motility < 50%, and morphology < 14% [23, 24]. Morphology was defined based on the Kruger criteria defined in the laboratory [23]. Semen was considered abnormal if at least one of the four criteria was below the lower reference limit set by the WHO.

A total of 11,068 individuals were included in the analysis. The mean age for the subjects undergoing infertility testing was 37.5 years. Over 90% of the population were between 30 and 49 years of age. 24.8% of the cohort received semen analyses between 1996 and 2000, and 56.6% were diagnosed by reproductive specialists while 11.5% were diagnosed by primary care.

We examined all men with semen data and examined five specific infertility related ICD-9 codes. One thousand two hundred forty (11.2%) patients had at least one infertility related code out of which 1098 had a male infertility diagnosis (i.e., 606.x). Two hundred ten (1.9%) patients were given more than one code but were only included once in the 606.x subgroup. Overall, the mean semen volume for all men with semen data was 3.1 mL, the mean sperm concentration was 63.7 × 10^6/mL, the mean for total motility was 43.6%, and the mean for normal morphology was 10.4%. Patients labeled with a male infertility diagnosis (606.x or V26.21) had a lower mean sperm concentration (49.0 × 10^6/mL, $p < 0.0001$), mean total motility (32.9%, $p < 0.0001$), and

mean morphology (9.77%, $p = 0.014$) than patients without an infertility diagnosis (Table 1).

The WHO criteria for analyzing semen quality were used to determine percentage of patients with abnormal semen parameters. Overall, 2228 men (20.1%) had semen volume less than 2.0 mL, 1855 men (16.8%) had sperm concentration less than 20×10^6/mL, 5722 men (51.7%) had total motility less than 50%, and 5474 men (49.5%) had less than 14% normal sperm morphology (Table 1).

Next, diagnostic accuracy of individual diagnosis codes for identifying abnormal semen analysis was calculated. Table 2 shows that the specificity for the group ranged from 92% to greater than 99%. When three or more 606.x codes were documented, the specificity further increased to 99.8%. The specificity was extremely high for all diagnostic codes and even higher when multiple diagnoses were coded.

Positive predictive values for all codes had a large range from 65 to 85%. The diagnostic code 606.1 was the best predictor of an abnormal semen analysis, and V26.21 (fertility testing) was the worst predictor. More stringent algorithms looking at only patients with more than one, two, or three medical claims do not appreciably change either the positive or negative predictive values.

The data was further stratified to determine associations between claims data and specific abnormal semen parameters. Both abnormal concentration and motility were evaluated due to their relevance to the diagnosis of azoospermia and oligospermia. Table 3 displays the results for concentration and motility, respectively. Similar to the previous analysis, individual semen parameters display a high specificity. The diagnostic codes 606.0 and 606.1 have positive predictive values for abnormal motility of greater than 90%.

Charts of patients with the diagnosis of azoospermia or oligospermia without the corresponding sperm concentrations that mirror those findings (e.g., concentration > 0 for azoospermia and concentration > 20 million/mL for oligospermia) were individually reviewed to determine the reason for possible miscoding. In patients coded with 606.0 (azoospermia) but with sperm concentration greater than 0, 58% (42 out of 73 men) had undergone either a vasectomy or a vasectomy reversal explaining the discrepancy. Thirty-two percent (23 men) had been given an incorrect diagnostic code and were not truly azoospermic. For patients coded with 606.1 (oligospermia) and a sperm concentration greater than 20 million per mL, 47% (14 out of 30 men) had multiple semen analyses with at least one resulting in low sperm concentration. Forty-three percent (13 men) were either misdiagnosed or miscoded.

Discussion

The ability of claims data to correctly identify cases of infertility has not previously been evaluated. In the current report, we compared semen quality with diagnostic codes to determine if an association exists. The specificity of ICD-9 coding for azoospermia (606.0), oligospermia (606.1), infertility due to extra-testicular causes (606.8), and unspecified male infertility were all greater than 90%. These results suggest that the claims data successfully identified infertile men defined by an abnormal semen analyses. Patients with three or more ICD-9 codes indicating infertility had a 99.8% chance of having an abnormal semen analysis—which is functionally confirmatory. This trend continued when looking exclusively at abnormalities in concentration and motility. The sensitivity of claims data in diagnosing infertility was not included in this study as not all patients in the cohort received semen analyses; it would be inaccurate to assume that the untested were all fertile. Nevertheless, among those tested, it seems that each ICD-9 diagnosis code is reasonably sensitive in identifying infertility as each code captures about two thirds of men with abnormal tests. Future studies including only men who have undergone testing are required to accurately quantify the sensitivity of diagnostic codes.

Despite our findings of high ICD-9 code specificity, we identified misdiagnoses in a fraction of patients attributed to, among other things, variation in semen parameters between

Table 1 Semen analysis results per diagnostic code

	Number	Mean ± standard deviation				% abnormal			
		Volume	Concentration	Motility	Morphology	Volume	Concentration	Motility	Morphology
None	9828	3.1 ± 1.9	65.5 ± 52.3	44.9 ± 23.1	10.4 ± 5.8	19.6%	15.3%	50.8%	49.6%
V26.21	142	3.0 ± 1.4	65.8 ± 50.9	42.7 ± 24.8	11.1 ± 7.0	19.7%	14.1%	56.3%	51.4%
606.0	222	3.3 ± 5.8	12.5 ± 30.5	8.0 ± 16.8	6.2 ± 4.5	31.1%	42.3%	49.6%	28.4%
606.1	165	3.4 ± 1.9	16.2 ± 19.7	13.7 ± 16.1	5.6 ± 4.3	21.8%	53.3%	77.0%	58.2%
606.8	40	3.1 ± 2.0	31.4 ± 42.7	20.8 ± 24.4	9.1 ± 5.7	30.0%	30.0%	50.0%	45.0%
606.9	903	2.9 ± 1.6	57.0 ± 52.7	38.3 ± 25.0	10.2 ± 6.8	23.6%	24.1%	59.1%	51.6%
606.x	1098	3.0 ± 3.0	46.8 ± 51.9	31.6 ± 26.1	9.6 ± 6.7	24.8%	29.9%	59.0%	48.0%
All	11,068	3.1 ± 2.0	63.7 ± 52.5	43.6 ± 23.8	10.4 ± 5.9	20.1%	16.8%	51.7%	49.5%

Table 2 Specificity and predictive value of claims data

	Specificity (%)	PPV (%)	NPV (%)
V26.21	98.8 [98.5,99.1]	65.5 [57.1,73.3]	38.1 [37.2,39.0]
606.x	92.3 [91.5,93.1]	79.8 [68.1,73.5]	39.1 [38.1,40.1]
606.0	98.4 [98.0,98.8]	69.8 [63.3,75.8]	38.2 [37.3,39.1]
606.1	99.4 [99.1,99.6]	84.9 [78.5,90.0]	38.4 [37.5,39.3]
606.8	99.7 [99.5,99.9]	70 [53.5,83.4]	38.1 [37.2,39.0]
606.9	93.5 [92.7,94.2]	69.6 [66.4,72.5]	38.7 [37.8,39.7]
One 606.x code	92.4 [91.6,93.2]	69.4 [66.5,72.2]	39.1 [38.1,40.1]
Two 606.x codes	98.8 [98.5,99.2]	75.3 [68.4,81.4]	39.1 [38.1,40.1]
Three or more 606.x codes	99.8 [99.6,99.9]	67.9 [47.7,84.1]	39.1 [38.1,40.1]

tests [25, 26]. However, miscoding also contributes to misclassification. One explanation for miscoding is confusion regarding the definition of the ICD-9 code 606—it may not be clear that 606 implies 606.0 (azoospermia) rather than "unspecified male infertility." The rare incongruity between diagnosis and semen analysis can satisfactorily be explained in most cases and the overall specificity of the data is not significantly affected.

While not all men receive a diagnosis of male factor infertility, there was an association between infertility diagnoses and semen quality. Not only did men with male infertility diagnoses have worse semen quality than those without, but semen quality also varied significantly based on diagnosis. Men with an infertility diagnosis of 606.x or V26.21 had lower sperm concentration, motility, and morphology than fertile men.

The specificity of diagnostic codes as shown in our work as well as Tessier-Sherman et al.'s corroborate prior literature using the assumption that claims data can accurately identify men with certain chronic medical diseases. The high specificity of ICD-9 codes with abnormal semen analyses allows researchers to use claims data to confidently identify and study a large, population of infertile men. Furthermore, data from the National Survey of Family Growth (NSFG) suggests that 7.5% of all sexually experienced men have at some point presented for infertility care with a significant proportion being insured [27••]. The number of men captured within the insurance datasets is sizeable and likely to grow [22].

These studies support and substantiate the movement towards using large claims datasets in medical research studying male infertility. With the increasing use of EMR systems, claims data is becoming more comprehensive and datasets more prevalent. Patient care is more centralized than in the past, and procurement of large databases through insurance providers is becoming easier. Financial incentives for physicians to include all relevant diagnostic codes also allow for greater accuracy compared with surveys or interviews, which may be susceptible to recall bias [1, 2]. Thus, administrative datasets are vital sources of information and have the potential to play a substantial role in clinical research. These datasets have already been used to uncover associations between diseases such as testicular cancer and infertile men, and will undoubtedly continue to foster similar research in the near future [28].

However, the data in this report is from a single institution and thus describes an association that may only be generalized to academic centers similar in population and setting. For example, at our medical center, urologists are one of many trained specialists in diagnosing and treating male infertility. Nonetheless, the distribution of providers caring for infertile

Table 3 Association between claims data and semen parameters

Diagnostic code	Abnormal sperm concentration			Abnormal sperm motility		
	Specificity (%)	PPV (%)	NPV (%)	Specificity (%)	PPV (%)	NPV (%)
V26.21	98.7	14.3	82.7	98.8	57.6	46.6
606.x	92.6	33.5	84.4	93.6	67.3	48.0
606	99.6	70.2	83.4	99.8	91.7	47.0
606.1	99.5	66.7	83.4	99.9	95.5	47.1
606.8	99.8	40.0	82.8	99.8	71.4	46.6
606.9	92.8	25.3	83.4	93.5	62.1	47.3
One 606.x code	92.3	28.8	84.4	93.1	63.5	48.0
Two 606.x codes	99.0	44.4	84.4	99.3	79.2	48.0
Three or more 606.x codes	99.9	41.2	84.4	99.9	82.4	48.0

men is unlikely disparate from any other large academic medical center in the USA. In addition, diagnostic coding is a relatively subjective manner of documentation that has the potential to be influenced by financial incentives and physician error [11]. This has mostly been accounted for through the inclusion and sub-stratification of all relevant codes associated with the diagnosis of infertility. Also, the use of WHO criteria to evaluate semen quality was used to define infertility despite questions regarding its validity. Less stringent criteria might alter the specificity of the data.

Conclusion

The use of insurance claims for unlocking valuable retrospective health data has become commonplace in medical research. The current review contributes infertility to the growing list of medical diseases for which claims data has been validated as an effective means for conducting clinical research. The high specificity in identifying infertile men supports the findings of prior studies by Meacham et al. that assessed prevalence and financial impact of male infertility as well as our group's previous findings that infertility may be associated with an increased risk of chronic medical disease and cancer [17••, 20, 28]. Future research will hopefully strengthen the association between claims data and infertility and further elaborate on the implications for men's health.

Compliance with Ethical Standards

Conflict of Interest Yash S. Khandwala, Chiyuan A. Zhang, Shufeng Li, Mark R. Cullen, and Michael L. Eisenberg each declare no potential conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
•• Of major importance

1. Adamson DM, Chang S, Hansen LG. Health research data for the real world: the MarketScan databases. New York Thompson … 2006;24.
2. Steinberg EP, Whittle J, Anderson GF. Impact of claims data research on clinical practice. Int J Technol Assess Health Care. 1990;6:282–7.
3. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease: comparing administrative data and self-reports. Med Care. 1997;35.
4. Bjarndóttir M V, Czerwinski D. Applications of insurance claims data in healthcare. Healthc. Anal. From Data to Knowl. to Healthc.Improv. 2016.
5. Ferver K, Burton B, Jesilow P. The use of claims data in healthcare research. Open Public Health J. 2009;2:11–24.
6. Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data: an overview. Med Care. 2002;40.
7. Klabunde CN, Harlan LC, Warren JL. Data sources for measuring comorbidity: a comparison of hospital records and medicare claims for cancer patients. Med Care. 2006;44:921–8.
8. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 1992;82:243–8.
9. Lee DS, Donovan L, Austin PC, et al. Data for use in outcomes research of Co in administrative and clinical data for use in outcomes research. 2002;43:182–188.
10. Steele LS, Glazier RH, Lin E, Evans M. Using administrative data to measure ambulatory mental health service provision in primary care. Med Care. 2004;42:6.
11. Tessier-Sherman B, Galusha D, Taiwo OA, Cantley L, Slade MD, Kirsche SR, et al. Further validation that claims data are a useful tool for epidemiologic research on hypertension. BMC Public Health. 2013;13:51.
12. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. J Clin Epidemiol. 2004;57:131–41.
13. Nallella KP, Sharma RK, Aziz N, Agarwal A. Significance of sperm characteristics in the evaluation of male infertility. Fertil Steril. 2006;85:629–34.
14. Taylor A. ABC of subfertility: extent of the problem. BMJ. 2003;327:434–6.
15. Povey AC, Clyma JA, McNamee R, Moore HD, Baillie H, Pacey AA, et al. Modifiable and non-modifiable risk factors for poor semen quality: a case-referent study. Hum Reprod. 2012;27:2799–806.
16. Redmon JB, Thomas W, Ma W, et al. Semen parameters in fertile US men: the study for future families. Andrology. 2013;1:806–14.
17.•• Eisenberg ML, Li S, Cullen MR, Baker LC. Increased risk of incident chronic medical conditions in infertile men: analysis of U.S. claims data. Fertil Steril. 2016;105:0015–282. **This study utilized a national insurance database to evaluate men's health after receiving fertility evaluation and found that these individuals have a higher risk of developing diabetes, ischemic heart disease, alcohol abuse, and drug abuse. More importantly, this study represents one of the growing number of investigations of infertile men utilizing claims data to identify subjects—a previously unvalidated methodology**
18. Eisenberg ML, Li S, Brooks JD, Cullen MR, Baker LC. Increased risk of cancer in infertile men: analysis of U.S. claims data. J Urol. 2015;193:1596–601.
19. Jensen TK, Jacobsen R, Christensen K, Nielsen NC, Bostofte E. Good semen quality and life expectancy: a cohort study of 43,277 men. Am J Epidemiol. 2009;170:559–65.
20. Meacham RB, Joyce GF, Wise M, Kparker A, Niederberger C. Male infertility. J Urol. 2007;177:2058–66.
21. Rajeshuni N. Infertility: a plague gone unnoticed. Stanford J. Public Heal. 2013.
22. Schmidt L. Effects of infertility insurance mandates on fertility. J Health Econ. 2007;26:431–46.
23. WHO. WHO Laboratory Manual for the examination and processing of human semen. World Health. 2010;5:286.
24. Organization WH. WHO Laboratory Manual for the examination of human semen and sperm-cervical mucus interaction. 1999.

25. Oshio S, Ashizawa Y, Yotsukura M, et al. Individual variation in semen parameters of healthy young volunteers. Arch Androl. 2004;50:417–25.
26. Mallidis C, Howard EJ, Baker HW. Variation of semen quality in normal men. Int J Androl. 1991;14:99–107.
27.•• Anderson J, Farr S, Jamieson D, Warner L, Macaluso M. Infertility services reported by men in the United States: national survey data. Fertil Steril. 2009;91:2466–70. **This important study emphasized the substantial prevalence of men seeking care for infertility (7.7%). Administrative databases have the potential to capture a large percentage of these patients with high specificity and thus may be used as a powerful investigational tool in the future**
28. Eisenberg ML, Li S, Brooks JD, Cullen MR, Baker LC. Increased risk of cancer in infertile men: analysis of U.S. claims data. J Urol. 2015;193:1596–601.