Estimating the prevalence of generalized and partial lipodystrophy: findings and challenges

Elaine Chiquette¹
Elif A Oral²
Abhimanyu Garg³
David Araújo-Vilar⁴
Praveen Dhankhar⁵

¹Aegerion Pharmaceuticals, Cambridge, MA, USA; ²Brehm Center for Diabetes Research and Metabolism, Endocrinology and Diabetes Division, Department of Internal Medicine, University of Michigan, Ann Arbor; MI, USA; ³Division of Nutrition and Metabolic Diseases, Department of Internal Medicine, Center for Human Nutrition, University of Texas Southwestern Medical Center, Dallas, TX, USA; ⁴Department of Medicine, UET eM, CIMUS School of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain; ⁵Complete HEOR Solutions (CHEORS), North Wales, PA, USA

Background: Lipodystrophy (LD; non-human immunodeficiency virus [HIV]-associated) syndromes are a rare body of disorders for which true prevalence is unknown. Prevalence estimates of rare diseases are important to increase awareness and financial resources. Current qualitative and quantitative estimates of LD prevalence range from 0.1 to 90 cases/million. We demonstrate an approach to quantitatively estimate LD prevalence (all, generalized, and partial) through a search of 5 electronic medical record (EMR) databases and 4 literature searches.

Methods: EMR and literature searches were conducted from 2012 to 2014. For the EMR database searches (Quintiles, IMS LifeLink, General Electric Healthcare, and Humedica EMR), LD cases were identified by the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code 272.6 (United Kingdom General Practice Research Database used other diagnostic codes to identify LD) plus additional LD-associated clinical characteristics (patients with HIV or documented HIV treatment were excluded). Expert adjudication of cases was used for the Quintiles database only. Literature searches (PubMed and EMBASE) were conducted for each of the 4 major LD subtypes. Prevalence estimates were determined by extrapolating the total number of cases identified for each search to the database population (EMR search) and European population (literature search).

Results: The prevalence range of all LD across all EMR databases was 1.3–4.7 cases/million. For the adjudicated Quintiles search, the estimated prevalence of diagnosed LD was 3.07 cases/million (95% confidence interval [CI], 2.30–4.02), 0.23 cases/million (95% CI, 0.06–0.59) and 2.84 cases/million (95% CI, 2.10–3.75) for generalized lipodystrophy (GL) and partial lipodystrophy (PL), respectively. For all literature searches, the prevalence of all LD in Europe was 2.63 cases/million (0.96 and 1.67 cases/million for GL and PL, respectively).

Conclusion: LD prevalence estimates are at the lower range of previously established numbers, confirming that LD is an ultra-rare disease. The establishment of diagnostic criteria and coding specific to the 4 major LD subtypes and future studies/patient registries are needed to further refine our estimates.

Keywords: adipose tissue, atypical diabetes, dyslipidemia, hypertriglyceridemia, insulin resistance, lipodystrophy, prevalence

Introduction

Physicians in daily practice are tasked with distinguishing an array of rare diseases from more common disorders in their differential diagnosis. Collectively, rare diseases are estimated to impact 6–8% of individuals at some point in their lifetime, greatly reducing patient life expectancy and quality of life.¹ Prevalence estimates of rare diseases are important to increase disease awareness and to marshal sufficient financial resources to improve detection and treatment. The European Union defines rare and ultra-rare
diabetes based on the prevalence rates of <5/10,000 individuals and <1/50,000 individuals, respectively.2,3 The US Food and Drug Administration has set a threshold of <200,000 people in the USA (or <7/10,000 people with a population of ~300 million) for an orphan drug that is intended to treat a rare disease or condition.4

Data for establishing a prevalence estimate may be derived from a variety of disparate information sources that are not standardized or are difficult to combine, including published case reports or systematic reviews, patient registries, expert opinion, and other anecdotal evidence.5 Analyses of claims databases can be very helpful in establishing a prevalence estimate; however, the lack of firmly established and specific diagnostic criteria or coding for a rare disease makes it challenging to know the true number of cases per population sample.1,6,7 Overestimation or underestimation of disease prevalence is likely, as the clinical findings and presentation of a patient with a rare disease may overlap with other common or rare diseases.8,9 Finally, some rare diseases may be more or less prevalent in a given geographic area than others.10,11

Lipodystrophy (LD) syndromes are rarely inherited or acquired disorders characterized by near-total or partial loss of body fat.12,13 Excluding LD in human immunodeficiency virus (HIV)-infected patients, LD syndromes have been classified into 4 major subtypes: acquired generalized lipodystrophy (AGL) and congenital generalized lipodystrophy (CGL), referred to here as generalized lipodystrophy (GL), and acquired partial lipodystrophy (APL) and familial partial lipodystrophy (FPL), referred to as partial lipodystrophy (PL). Patients with GL syndromes have markedly low serum leptin levels, which reflects the widespread loss of adipose tissue and contributes to metabolic abnormalities (severe insulin resistance, lipoatrophic diabetes, hypertriglyceridemia, hepatic steatosis, and others) that increase patient morbidity and mortality. Although patients with PL may have variable loss of adipose tissue with corresponding variable levels of serum leptin, they also suffer from metabolic complications.13,14

Awareness of GL and PL is low due to its rarity, making an accurate estimate of disease prevalence very difficult. Patients with GL often have diabetes that is difficult to manage, and they may bear some dysmorphic features and coarse facies (the latter also a characteristic of acromegaly and gigantism).13,15,16 The overlap of clinical presentation with more common diseases may complicate the diagnosis. Similarly, patients with PL may also have diabetes that is difficult to manage and a physical appearance that may resemble an overweight or obese patient with metabolic syndrome or type 2 diabetes or a patient with Cushing’s syndrome.17,18 In patients with FPL type 1, a form of severe insulin resistance characterized by a lack of fat in the extremities, truncal obesity, and having components of the metabolic syndrome, there may be a polygenic contribution of up to 53 different loci that are also known to be associated with more common insulin-resistant states and obesity-related conditions.19 Including patients with FPL type 1 in a prevalence estimate, along with other well-defined monogenic forms of FPL that are more rare, could mistakenly inflate a prevalence estimate. Similarly, the lack of precise diagnostic criteria makes it hard to firmly establish the diagnosis of PL; patients may get missed, and this also complicates prevalence estimates. Finally, acquired forms of GL and PL have also been associated with autoimmune diseases, such as dermatomyositis, systemic lupus erythematosus, autoimmune hepatitis, and others.14,16,20

Not surprisingly, currently available worldwide and European prevalence estimates for the 4 major subtypes of LD combined are wide ranging and varied (~0.1–90 cases/million) depending on the information source (previously published reviews, expert opinion, or figures cited by the European Medicines Agency or Orphanet) and the methodology used for computation of the estimate (qualitative or quantitative).15,21-29 For a rare disease such as LD, documentation of disease prevalence is an important requirement mandated by many regulatory agencies for obtaining orphan drug designation.5,30 We aimed to demonstrate an approach to quantitatively estimate the prevalence of all LD, GL, and PL through a search of 5 electronic medical record (EMR) databases and 4 literature searches.

Methods

EMR database searches
A total of 5 EMR databases were individually searched from a period of 2012–2014—4 maintained in the USA (Quintiles, IMS LifeLink Health Claims, General Electric [GE] Healthcare, and Humedica EMR) and one maintained in the UK (United Kingdom General Practice Research Database [UK GPRD]) (Table 1). Collectively, the 5 EMR databases represent a total population of ~139 million deidentified patient records derived from a large number of medical institutions and providers. The level of patient detail was variable across each EMR database. At the time of the study, all databases captured basic demographic information and allowed for the search of International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnostic codes (the UK GPRD used a system of searchable diagnostic coding that was unrelated to ICD-9-CM), but not all EMRs captured or linked patient records to laboratory...
data or physician notes that would lend additional support to review or confirm a patient’s diagnosis.

Because of variability in the data collected and disease terminology used across databases, search criteria were developed that were specific to each database; however, the overall search strategy used for each database was mostly similar. First, patient criteria that characterized LD were defined before the search of each database (Table 1). Second, the database was queried in the following sequence to generate a total number of patients with LD: 1) search for ICD-9-CM code 272.6 for LD (other diagnostic codes were used for UK GPRD); 2) removal of patients with documented HIV or HIV treatment; and 3) removal of patients who did not meet the mentioned above clinical characteristics or diagnostic criteria often associated with LD. Clinician experts reviewed physician notes for the phrases “lipodystrophy,” “lipoatrophy,” or “lipoatrophic.”

**Abbreviations:** BMI, body mass index; EMR, electronic medical record; HbA1c, glycated hemoglobin; HIV, human immunodeficiency virus; ICD-9-CM, Information Classification of Diseases, Ninth Revision, Clinical Modification; LD, lipodystrophy; NASH, nonalcoholic steatohepatitis; T2DM, type 2 diabetes mellitus; TGs, triglycerides; UK GPRD, United Kingdom General Practice Research Database.

**Notes:** aClinical characteristics or diagnostic criteria often associated with LD. EMR databases were queried in a 3-step process: 1) search for ICD-9-CM code 272.6 for LD (other diagnostic codes were used for UK GPRD); 2) removal of patients with documented HIV or HIV treatment; and 3) removal of patients who did not meet the mentioned above clinical characteristics or diagnostic criteria often associated with LD. bClinician experts reviewed physician notes for the phrases “lipodystrophy,” “lipoatrophy,” or “lipoatrophic.”

<table>
<thead>
<tr>
<th>EMR database (location)</th>
<th>Size (million)</th>
<th>Characteristics of EMR database patient population</th>
<th>Database-specific inclusion criteria in addition to ICD-9-CM code 272.6*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quintiles (USA)</td>
<td>~17</td>
<td>Database of patient-level data from large, centralized national network of outpatient offices involving &gt;725 member institutions and &gt;39,000 providers</td>
<td>Expert adjudication*</td>
</tr>
<tr>
<td>IMS LifeLink Health Claims (USA)</td>
<td>~70</td>
<td>Database composed of commercial health plan information from enrollees in managed care plans involving ~80–90% of hospitals, doctors, and Fortune 100 companies throughout the USA</td>
<td>&lt;65 years with T2DM or any lipid disorder, diagnosed by an endocrinologist or pediatrician</td>
</tr>
<tr>
<td>GE Healthcare (USA)</td>
<td>~30</td>
<td>Database composed of independent physician practices, academic medical centers, hospitals, and large integrated delivery networks</td>
<td>Presence of 1 or 2 additional diagnostic markers: T2DM, High TGs, NASH</td>
</tr>
<tr>
<td>Humedica EMR (USA)</td>
<td>~12</td>
<td>A patient-based health care system EMR database of inpatient, outpatient, and pharmacy claims</td>
<td>Non-obese (BMI &lt;30 kg/m²) and any of the following: Uncontrolled diabetes (T2DM diagnosis, age ≤60–65 years, and HbA₁c ≥6.5%), High TGs (no T2DM, age ≤60 years, TGs &gt;350–400 mg/dL), or NASH (no T2DM, age &lt;40 years, and TGs &lt;400 mg/dL)</td>
</tr>
<tr>
<td>UK GPRD (UK)</td>
<td>~10</td>
<td>Contains longitudinal data from the EMRs of patients from a large sample of general practices within the UK</td>
<td>≥2 diagnosis claims for T2DM or High TGs or chronic nonalcoholic liver conditions</td>
</tr>
</tbody>
</table>

**Table 1** EMR database characteristics and search details
not enough information was available to confirm or refute the diagnosis, were excluded from the count. Given the rigor of the review, the cases counted as LD were not ambiguous cases, and there was usually enough detail in the medical records to indicate that they were seen at reference centers such as the National Institutes of Health, University of Texas Southwestern, or University of Michigan for diagnosis, or that a consultation was arranged for verification. Although quite rigorous, this approach may have resulted in an underestimate of the prevalence.

To arrive at a prevalence estimate of all LD for each EMR database, the total number of identified cases was divided by the total population of the respective EMR database (Table 1). For the Quintiles search with expert adjudication, it was possible to estimate the prevalence for diagnosed cases of GL (AGL and CGL) and PL (APL and FPL) in addition to all LD.

The use of the data from EMR databases did not require approval from an institutional review board. The EMR databases consisted of de-identified patient data.

**Literature searches**

PubMed and EMBASE literature searches were conducted through May 2012. Search terms and time frames for the literature searches are provided in Table 2. References identified by the searches were evaluated, and review articles or clinical case reports that summarized cases of LD in the European Union community population were retained. References that summarized cases of LD from countries outside of the European Union, repeated or updated presentations of the same patient, or general reviews of related syndromes or preclinical reviews were excluded.

A prevalence estimate was determined through a 3-step process. First, the total number of European cases identified for all LD (AGL + CGL + APL + FPL), GL (AGL + CGL), and PL (APL + FPL) was determined. Second, because it has been estimated that only 25% of all LD cases have been reported in the literature, the total number of cases identified was multiplied by 4. Third, the total number of cases for all LD, GL, and PL (after adjustment for underreporting) was divided by the total European Union community population as of 2012 figure (507,751,512 people).

**Statistics**

Prevalence estimates for each EMR database and literature search were summarized descriptively and presented as the number of cases per million. For the Quintiles EMR adjudicated search, 95% confidence intervals (CIs) were constructed using the Clopper–Pearson method based on the exact binomial distribution and then scaled to 1,000,000 patients. CIs were obtained using readily available software. A summary prevalence estimate combining the data from all 5 EMR databases was not possible because of differences in the search criteria used (each database used its own medical terminology), the lack of resources to perform an adjudicated search of each EMR, and the inability to minimize the double counting of patient cases across each database.

**Results**

The range of worldwide prevalence of all LD as determined by the simple EMR database searches was 1.3–4.7 cases/million (Figure 1). For the adjudicated Quintiles database search, the estimated prevalence of all diagnosed LD cases was 3.07 cases/million (95% CI, 2.30–4.02). When separated into diagnosed GL and PL, the estimated prevalence was 2.6 cases/million (95% CI, 1.9–3.4). The estimated prevalence for all LD cases was 4.7 cases/million (95% CI, 3.9–5.5).
0.23 (95% CI, 0.06–0.59) and 2.84 (95% CI, 2.10–3.75) cases/million, respectively.

Literature searches for AGL, CGL, APL, and FPL identified 7, 31, 11, and 40 publications, respectively, that met the search criteria (Figure 2). The total number of all LD, GL, and PL cases was 334, 122, and 212, respectively. When adjusted for underreporting and extrapolated to the European Union community population (507,751,512 people), the estimated prevalence of all LD was 2.63 cases/million (Figure 3). The estimated prevalence of GL and PL was 0.96 and 1.67 cases/million, respectively.

**Discussion**

By evaluating the prevalence of LD from the 5 EMR database searches and 4 literature searches, we have demonstrated that the range of prevalence estimates for all LD (1.3–4.7 cases/million), GL (0.2–1.0 cases/million), and PL (1.7–2.8 cases/million) is smaller and narrower than the range of previously established estimates from multiple and disparate information resources (~0.1–90 cases/million). Our findings confirm that LD is an ultra-rare disease, as it is well below the thresholds established for an ultra-rare disease set forth by the European Union (<1/50,000 individuals). A recent report on the natural history of 33 patients with CGL from Turkey estimated a local prevalence of 0.5 cases/million for CGL. This estimate of CGL compares well with our estimate of 0.2–1.0 cases/million for GL, which includes both AGL and CGL.

Determining a prevalence estimate for LD was not an easy undertaking. One major limitation related to the EMR database searches was the use of the nonspecific LD ICD-9-CM diagnostic code (Table 3). Codes specific for the 4 major LD subtypes have not been established because the diagnostic criteria have not been finalized. Through the years, several expert opinions and reviews have been published suggesting diagnostic criteria for the 4 major subtypes of LD, but a unified consensus statement has been elusive to achieve. In the adjudicated Quintiles database search, it was observed that the ICD-9-CM code 272.6 was overused by dermatology and plastic surgery in the USA, causing many false-positive hits and illustrating the need for more accurate forms of verification of true cases. Localized LD has the same ICD-9-CM code and can confound prevalence estimates, especially in patients with diabetes mellitus receiving insulin therapy. It could also be argued that pairing the ICD-9-CM code 272.6 with a metabolic disorder or a diagnostic marker, as was done in the EMR database searches, may have excluded patients with LD who did not have active metabolic abnormalities or abnormalities that were presently controlled.

A second major limitation in establishing a prevalence estimate for the EMR database searches was the variability in information sources. Humedica allowed for ranges of biochemical laboratory values to be searched in addition to the ICD-9-CM code for LD; thus, a more thorough search for LD cases was possible. Other EMR databases were more limited in their search capabilities due to the level of data captured within each database. The Quintiles database, for example, allowed individual patient records to be searched; however, not all patient records contained detailed physician notes, thus

---

**Figure 2** Flow of literature search results.

*Note:* Garg has estimated that only 25% of all LD cases are reported.

**Abbreviations:** AGL, acquired generalized lipodystrophy; APL, acquired partial lipodystrophy; CGL, congenital generalized lipodystrophy; EU, European Union; FPL, familial partial lipodystrophy; LD, lipodystrophy.
making expert adjudication more difficult. It is possible that our estimate from Quintiles underestimated the true prevalence due to only including patients with complete records for adjudication. A final point to variability was the lack of standardization of medical terminology between individual databases. The development of different search criteria for each database in our study could have introduced bias to the prevalence estimates. A third limitation to the EMR database searches was that databases were held and maintained in the USA and the UK. Samples from databases in other regions of the world where LD cases are known to occur would help to add to the range of databases held and maintained in the USA and the UK.

A second limitation is the use of a wide range of dates in the literature search (>50 years). It is possible that some patients who were included in the estimate may have been deceased. A third limitation was the decision to limit the literature search strategy to only the 4 major subtypes of LD. The exclusion of novel forms of LD that have not been fully characterized or that fit into the classification scheme of 4 major subtypes may have resulted in an underestimate of all LD prevalence.

One limitation concerning the literature searches is that despite efforts to avoid counting the same patient more than once, there is always the possibility that this occurred due to the lack of sufficient demographic details across publications. The number of cases derived from the literature search was based on the assumption that only 25% of cases are actually reported in the literature; however, this commonly used assumption has never been fully confirmed by a separate study because of the rarity of the disease. A second limitation is the use of a wide range of dates in the literature search (>50 years). It is possible that some patients who were included in the estimate may have been deceased. A third limitation was the decision to limit the literature search strategy to only the 4 major subtypes of LD. The exclusion of novel forms of LD that have not been fully characterized or that fit into the classification scheme of 4 major subtypes may have resulted in an underestimate of all LD prevalence.

It should also be stressed that the purpose of this paper was not to refine the prevalence of the genotypes for the genetic forms of LD, as these data are not readily available in the EMRs. Therefore, we did not specify the prevalence of the reported genotype from the specific papers in the medical literature when this information was sometimes available (but not uniformly). There are distinctive genetic causes for the different forms of LD,13 and more specific attention should be given to estimate the true prevalence of the different genotypes, which exceeds the scope of this paper. Finally, a general limitation in both the EMR database searches and the literature searches relates to the heterogeneity of LD syndromes (especially PL syndromes where patients with FPL type 1 may have been included) and the lack of specific diagnostic criteria to establish a firm diagnosis and, thus, a reliable prevalence estimate.

As investigations into rare diseases are inherently challenged because of a lack of patient cases to draw from, we consider the large number of information sources used in this study to be a strength. By casting a wide net with a search of 5 large claims databases held in the USA and the UK and 4 literature searches using PubMed and EMBASE, we have demonstrated that the range of possible prevalence was narrower and at the lower end of previously available estimates (~0.1–90 cases/million). A second strength of the study was the use of expert adjudication of LD cases in the Quintiles database search. The prevalence of all diagnosed cases of LD identified by Quintiles was 3.07 cases/million, which was approximately in the middle of estimates derived from the other 4 EMR databases using the simple search. To
obtain a more precise prevalence estimate, it would have been ideal to have applied this additional layer of adjudication to the search of all 5 EMR databases and to include a search of undiagnosed cases of LD, but this was not possible due to the enormity of such an effort.

The finalization of diagnostic criteria and the establishment of diagnostic coding specific to the 4 major subtypes of LD will greatly help to refine future prevalence estimates derived from the database searches. Furthermore, there is a need to better differentiate FPL type 1 from the rest of FPL cases, as it appears to be distinctly different in etiology and creates confusion on the prevalence of the condition in general. The hope for more precise diagnostic codes assumes that endocrinologists and nonspecialist clinicians will be sufficiently informed and aware of LD to use the code and correctly distinguish from other conditions that may bear a similar clinical presentation to the LD syndromes or be able to differentiate between various LD subtypes. The establishment of registries is one recommended part of regional and national plans to increase the awareness of rare diseases and to provide the natural history of the disease.\(^\text{1}\) Data collected from a multicenter global registry initiative organized by major academic centers studying these conditions and the upcoming Metreleptin Effectiveness and Safety Registry (MEASuRE; NCT02325674) will help to improve the identification of new patients with LD and increase overall disease awareness. Prospective studies are also needed to validate algorithms independent of diagnostic codes to further define the prevalence of LD.

**Conclusion**

LD syndromes are a rare body of diseases for which prevalence is currently estimated to be 1.3–4.7 cases/million. By evaluating multiple information sources, we have demonstrated that the range of possible prevalence is smaller and narrower than the range reported in previous estimates (0.1–90 cases per million). Our methodology describes one approach to estimate the prevalence of a rare disease, which may be of use to other research groups investigating the prevalence of rare diseases. The true prevalence of LD is unknown, but future research will help to further refine and inform the estimates.

**Acknowledgments**

Robert Schupp, PharmD, CMPP, of inScience Communications, Springer Healthcare (Philadelphia, PA, USA), provided medical writing support funded by Aegerion Pharmaceuticals. Some of the data from this paper were presented in an abstract/poster form at the 2016 European Conference on Rare Diseases & Orphan Products in Edinburgh, Scotland; the 2015 American Diabetes Association Annual Meeting in Boston, MA; and the 2015 International Society for Pharmacoeconomics and Outcomes Research Annual Meeting in Philadelphia, PA. Funding of the study and development of the paper were supported by Aegerion Pharmaceuticals.

**Author contributions**

All authors contributed toward data analysis, drafting and critically revising the paper and agree to be accountable for all aspects of the work.

**Disclosure**

EC is an employee and stockholder of Aegerion Pharmaceuticals. EAO has worked as a consultant/advisor to and has received grant/drug support from Amylin, Bristol-Myers Squibb, AstraZeneca, and Aegerion Pharmaceuticals (manufacturers of metreleptin through the years); Akcea Therapeutics; and Ionis Pharmaceuticals. Unrelated to the field of LD, EAO has received nonmaterial support from Boehringer Ingelheim and grant support from GI Dynamics. EAO is partially supported by R01 DK088114. AG has received research support from Aegerion Pharmaceuticals and Ionis Pharmaceuticals and has worked as a consultant for Aegerion Pharmaceuticals, Biomedical Insights, ClearView Healthcare Partners, Gerson Lehrman Group, and Smith-Solve. DA-V has nothing to disclose. PD is an employee of Complete HEOR Solutions, which has done contract work for Aegerion Pharmaceuticals. The authors report no other conflicts of interest in this work.

**References**


