Data from Routine Primary Healthcare: Opportunities and Threats for Obesity Epidemiological Research

Berta Valente[a, b]   Mónica Rodrigues[a, b]   João Pedro Ramos[a, b, c]
Ana Azevedo[a, b, d]

[a]EPIUnit ITR, Instituto de Saúde Pública da Universidade do Porto, Universidade do Porto, Porto, Portugal; [b]Departamento de Saúde Pública e Ciências Forenses e Educação Médica, Faculdade de Medicina, Universidade do Porto, Porto, Portugal; [c]Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal; [d]Centro de Epidemiologia Hospitalar, Unidade Local de Saúde São João, Porto, Portugal

Introduction

Real-world data are increasingly recognized as important in clinical care and research. Electronic health records (EHRs) serve as a source of real-world data, capturing medical information gathered during healthcare appointments [1]. In primary healthcare settings, where individuals and families are often followed during their life course, EHR data enable the analysis of large samples in a longitudinal perspective. Still, its value depends on the accuracy of data recording to ensure quality and usefulness. High-quality clinical data must meet three key dimensions: completeness, correctness, and currency [2–6]. Completeness requires all expected patient information to be present [3, 4]. Correctness ensures that the data reflect true observations, align with related information, and consistently measure the intended attribute [3, 4]. Currency refers to data being recorded within the relevant time frame [2, 3, 5, 6].

Data from EHR provide a significant opportunity to advance the field of obesity epidemiology, a critical area at the national level. In Portugal, nearly 32% of children aged 6–8 years were living with overweight and 13.5% with obesity in 2022 [7]. In the adult population, the prevalence of overweight and obesity was 38.9% and 28.7%, respectively [8].

Implications of Coding Practices for Overweight and Obesity Research

Recent Portuguese studies on the prevalence of overweight and obesity in both paediatric and adult populations have used data registered in SClinico®, a

Correspondence to:
Berta Valente, berta.valente@ispup.up.pt

national clinical information system managed by central public services. The following studies employed the International Classification of Primary Care (ICPC-2) codes T83 and T82, corresponding to diagnoses of overweight and obesity, respectively.

In adults, a retrospective longitudinal study using data from SClinico® reported a 10.2% prevalence of obesity in the northern region of Portugal, in 2019 [9]. This prevalence was notably lower than the 28.2% and 21.6% reported in population-based studies representative of the Portuguese population [9]. The authors argue that this discrepancy may result from the reliance on EHR data, which may not capture obesity cases not coded with T82 [9].

Since 2018, for adults, SClinico® automatically assigns the T83 and T82 codes during consultations when anthropometric measures are taken or self-reported and the body mass index (BMI) fulfils criteria based on World Health Organization (WHO) cut-offs [9–11]. However, as this study was retrospective, some individuals may not have been captured under this procedure [9]. Additionally, among the 348,536 individuals analysed, 69,098 (19.8%) were excluded due to height (<116.2 cm, >194 cm) and weight (<37.9 kg, >160 kg) values falling outside the minimum and maximum thresholds of plausibility [9].

A study conducted with 5,931 children aged 2–9 years in Portugal's Northern Region reported a prevalence of obesity of 3.4% [10]. The authors encountered some challenges with T82 coding, as 32.4% of children diagnosed with obesity had to be excluded due to implausible weight and height values [10]. Among the remaining 4,008 participants, only 67% met the WHO criteria for obesity, despite all being coded as such [10]. This discrepancy may result from incorrect coding or failure to update EHRs when a child's weight status changes [10]. Additionally, unlike for the adult population, the system does not issue any alerts or automatic coding for BMI values indicating obesity or overweight in children. While the system automatically calculates BMI and assigns a percentile when height and weight data are entered, the interpretation of these data is the sole responsibility of the family doctor, who must individually record the diagnoses in the EHR.

A study examined data from 3,188 children aged 6–8 years and 3,565 adolescents aged 15–17 years in Matosinhos municipality, focusing on obesity prevalence and the accuracy of clinical diagnoses (T82) [11]. Among the 6–8-year-olds, 4.3% were living with overweight and 2.1% with obesity [11]. However, 90.7% of children with overweight and 67% of those with obesity determined by BMI records lacked cor- responding clinical diagnosis codes [11]. Moreover, 26.6% of children with obesity in this age group were incorrectly diagnosed as having overweight (T83) [11]. In the 5–17-year-old group, 7.5% were diagnosed with overweight and 5.6% with obesity [11]. Again, 76.6% of adolescents with overweight and 36.6% with obesity did not have clinical diagnoses (T83 and T82) [11]. These findings reveal a significant gap between BMI-based categorization and clinical diagnoses, emphasizing the need for improved registration, diagnostic accuracy, and awareness among healthcare providers [11].

These studies mark a significant advance in access to data for national-level research on the prevalence of overweight and obesity in primary care. However, at this stage, and with the currently available data, relying solely on T83 or T82 codes may not provide accurate estimates of overweight and obesity rates in the population using primary healthcare.

EHR in Obesity Research: Advantages and Risks

The use of EHR holds great potential, particularly due to automated data collection, which can reduce healthcare professionals' workload while providing more comprehensive information. However, data errors may introduce bias in the studies results [12]. Common errors in anthropometric measures recorded in EHR include unit misclassification, adding or omitting digits, and swapping height and weight values [13].

As secondary data, it is crucial to guarantee the identification and correction of potential errors to produce robust, unbiased findings [14]. Automated data cleaning presents significant potential for research involving large EHR databases. For example, using *growthcleanr*, an easy-to-use tool for cleaning large-scale paediatric and adult height and weight data, offers several advantages, these include algorithms tailored for both paediatric and adult populations, efficiency in handling big data, and the ability to process extraneous and carried-forward measurements [14]. Using an exponentially weighted moving average, this tool identifies outliers and detects implausible values or errors through longitudinal data analysis. Being open source and specifically designed for EHR data, *growthcleanr* can process large datasets efficiently through parallel computing. However, users must be familiar with R for smooth execution, and the exponentially weighted moving average method requires at least two measurements to ensure stability.

Another challenge in current practice is the fragmentation of EHR systems, with a lack of integration between the software accessible to different categories of healthcare professionals. This fragmentation can affect the accuracy of weight and height, which are essential for calculating BMI and assessing overweight and obesity rates. Additionally, periodic assessments of anthropometric measures would help ensure the information remains up to date.

The layout and functionality of the EHR user interface, and the interactive computer screens used by clinicians, can significantly influence the quality of data collected. A well-designed interface enhances data quality by providing content, features, and functionalities that support accurate documentation [2, 15, 16]. The effectiveness of SClinico® could be greatly improved if all members of the healthcare team had access to the same interface, with information entered in specific, structured fields. This approach would reduce redundancy and enhance the efficiency of consultations by minimizing repetitive tasks, such as taking measurements.

Better integration of EHR systems could also ensure that anthropometric data are more consistently recorded, reducing reliance on self-reported data, which can lead to BMI miscalculations and inaccuracies in obesity prevalence estimates. Direct links between software and clinical scales for data import could further improve data accuracy and continuity. Additionally, incorporating clear prompts and visual distinctions between fields for measured and self-reported data could help prevent mixing of these sources in the system. For the paediatric population, a notification asking whether to add a diagnosis after entering anthropometric data could also prove beneficial.

Opportunities and Strategies

Addressing these issues may significantly improve the reliability and validity of epidemiological studies on obesity. Misclassification and underdiagnosis not only hide the true extent of the problem but also impede the implementation of effective interventions and policies [17]. There is a critical need for better integration of standardized software protocols for obesity diagnosis, along with rigorous data collection and documentation practices. Only through such concerted efforts can we ensure that real-world data accurately reflect the epidemiological landscape and inform evidence-based strategies to curb this public health emergency.

Indeed, Dispatch No. 12634/2023, issued on December 11, 2023, in Diário da República, mandates an Integrated Care Model for Obesity Prevention and Treatment, which consists of a tailored care approach targeting individuals with obesity [18]. This may constitute a great opportunity to address the arguments raised as it proposes establishing a "Specialized and Multidisciplinary Obesity Consultation" within primary healthcare settings, aimed at individuals with obesity with or without comorbidities, overseen by a multidisciplinary healthcare team [18]. Implementing an automatic referral system for individuals coded with obesity could further streamline access to this program.

Conclusion

Adaptations to information systems are necessary for accurate categorization, and ongoing training for professionals is crucial. Therefore, we advocate for the involvement of healthcare professionals and academics in the development, implementation, and monitoring of the software used in clinical practice as their expertise and first-hand insights can ensure that these tools are both practical and effective. Furthermore, Dispatch No. 12634/2023 offers a valuable opportunity to train the healthcare workforce in anthropometric assessment, to raise awareness about the systematic and accurate coding of T83 and T82, as well as to assess and refine the structured pathway through which people with obesity receive care in primary healthcare settings.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Author Contributions

Berta Valente, Mónica Rodrigues, and João Pedro Ramos were responsible for design, writing, content review, and approval of the manuscript. Ana Azevedo was responsible for the critical review with important intellectual contribution.

References

1 Pavão J, Bastardo R, Pereira LT, Oliveira P, Costa V, Martins AI, et al. Considerations on the usability of SClínico. In: Cliquet A Jr, Wiebe S, Anderson P, Saggio G, Zwiggelaar R, Gamboa H, et al, editors. Proceedings of the 11th international joint conference on biomedical engineering systems and technologies (BIOSTEC 2018), 2018 Jan 19-21, Madeira, Portugal. Berlin: Springer; 2019.

2 Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013; 20(1):144–51. https://doi.org/10.1136/amiajnl-2011-000681

3 Bian J, Lyu T, Loiacono A, Viramontes TM, Lipori G, Guo Y, et al. Assessing the practice of data quality evaluation in a national clinical data research network through a systematic scoping review in the era of real-world data. J Am Med Inform Assoc. 2020; 27(12):1999–2010. https://doi.org/10.1093/jamia/ocaa245

4 Weiskopf NG, Bakken S, Hripcsak G, Weng C. A data quality assessment guideline for electronic health record data reuse. EGEMS. 2017;5(1):14. https://doi.org/10.5334/egems.218

5 Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. EGEMS. 2016;4(1):1244. https://doi.org/10.13063/2327-9214.1244

6 Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997;4(5):342–55. https://doi.org/10.1136/jamia.1997.0040342

7 Rito A, Mendes S, Figueira I, Faria MC, Carvalho R, Santos T, et al. Childhood obesity surveillance initiative: COSI Portugal 2022. Lisboa: Instituto Nacional de Saúde Doutor Ricardo Jorge; 2023.

8 Barreto M, Gaio V, Kislaya I, Antunes L, Rodrigues AP, Silva AC, et al. [1st national health examination survey (INSEF 2015): health status]. Lisboa: Instituto Nacional de Saúde Doutor Ricardo Jorge; 2016. Portuguese.

9 Páscoa R, Teixeira A, Henriques TS, Monteiro H, Monteiro R, Martins C. Characterization of an obese population: a retrospective longitudinal study from real-world data in northern Portugal. BMC Prim Care. 2023; 24(1):99. https://doi.org/10.1186/s12875-023-02023-7

10 Amorim AFC. [Childhood obesity in the northern regional health administration (Portugal): analysis of real-world data] master's thesis. Porto: Faculty of Medicine of the University of Porto; 2023.

11 Pereira CM, Silva CH, Oliveira RM, Príncipe RM, Valpaços EJ. [Prevalence of overweight children in primary health care users in Matosinhos: a comparison of Body Mass Index with classification in clinical records]. Rev Port Endocrinol Diabetes Metab. 2020;15(1–2).

12 Smith N, Coleman KJ, Lawrence JM, Quinn VP, Getahun D, Reynolds K, et al. Body

weight and height data in electronic medical records of children. Int J Pediatr Obes. 2010; 5(3):237–42. https://doi.org/10.3109/17477160903268308

13 Roche A. Growth, maturation, and body composition: the fels longitudinal study 1929–1991. New York: Cambridge University Press; 1992.

14 Lin PID, Rifas-Shiman SL, Aris IM, Daley MF, Janicke DM, Heerman WJ, et al. Cleaning of anthropometric data from PCORnet electronic health records using automated algorithms. JAMIA Open. 2022; 5(4):ooac089. https://doi.org/10.1093/jamiaopen/ooac089

15 Johnson SB, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, et al. An electronic health record based on structured narrative. J Am Med Inform Assoc. 2008;15(1):54–64. https://doi.org/10.1197/jamia.M2131

16 Hasan S, Padman R. Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. AMIA Annu Symp Proc. 2006;2006:324–8.

17 Ramos E, Lopes C, Oliveira A, Barros H. Unawareness of weight and height: the effect on self-reported prevalence of overweight in a population-based study. J Nutr Health Aging. 2009;13(4):310–4. https://doi.org/10.1007/s12603-009-0028-7

18 Despacho n.o 12634/2023, de 11 de dezembro. Diário da República Série II. n.o; 2023-12-11. Vol. 237; p. 99–101.