Establishment of a large-scale oral disease registry (NDCS-ODR) in a national specialty center

John Rong Hao Tay[1,2,3†*], Gustavo G Nascimento[3,4,5†], Jasmine Sio Hong Ho[3‡], Narayanan Ragavendran[1,6], Bryan Wei Ren Yeo[1,6], Sean Shao Wei Lam[1,6,7], Hanimi Reddy Kallam[1,6], Marco A Peres[1,3,4*]

[1]*Health Services Research and Population Health Programme, Duke-NUS Medical School, Singapore, Singapore.*

[2]*Department of Restorative Dentistry, National Dental Centre Singapore, Singapore.*

[3]*National Dental Research Institute Singapore, National Dental Centre Singapore, Singapore, Singapore*

[4]*Oral Health Academic Clinical Programme, Duke-NUS Medical School, Singapore, Singapore, Singapore*

[5]*School of Dentistry, University of Utah, Salt Lake City, Utah, USA.*

[6]*Health Services Research Centre, SingHealth, Singapore, Singapore.*

[7]*School of Computing and Information Systems, Singapore Management University, Singapore, Singapore*

*Correspondence:

Marco A Peres

Health Services Research and Population Health Programme, Duke-NUS Medical School, 8 College Road, Singapore 169857, Singapore. Email: marco.peres@duke-nus.edu.sg

John Rong Hao Tay

Health Services Research and Population Health Programme, Duke-NUS Medical School, 8 College Road, Singapore 169857, Singapore. Email: john.tay.r.h@singhealth.com.sg

[†]These authors contributed equally to this work and share first authorship.

[‡]Data manager at the National Dental Centre Singapore at the time of the study.

Abstract

This study describes the establishment of the National Dental Center Singapore Oral Disease Registry (NDCS-ODR), a large-scale, electronic health records registry designed to capture real-world data on oral diseases. The NDCS-ODR was developed to standardize and integrate oral health data within Singapore Health Services, the country's largest healthcare cluster. Its development, governance, and data architecture are described, with an overview of individuals with oral diseases recorded in the registry. Data collection from 2013 to June 2025 has been completed. As of June 2025, the NDCS-ODR comprises 229,249 unique patients, with a mean (SD) age of 49.1 (19.5) years and an approximately equal sex distribution. Most were of Chinese ethnicity (77.6%), and Singapore citizens (92.5%). Clinical variables indicated substantial disease and treatment burden, with a mean of 7.9 (7.7) missing teeth, 4.8 (6.2) restored surfaces, and 2.8 (3.4) restored teeth per patient. Among 108,517 recorded periodontal diagnoses, Stage III periodontitis (2018 EFP/AAP Classification) and severe chronic periodontitis (1999 Classification of Periodontal Diseases and Conditions) were most common. The NDCS-ODR represents Singapore's first large-scale, real-world oral disease registry embedded within a national specialty center, demonstrating the feasibility of leveraging electronic health record data for research and service evaluation.

Keywords: electronic data processing, epidemiology, big data, data science, dentistry, oral health

Introduction

Oral diseases are among the most prevalent non-communicable diseases worldwide [1]. The World Health Organization estimates that nearly 3.7 billion people, half of the world population, are affected by untreated caries, periodontitis, edentulism, and other oral disorders. Furthermore, caries and periodontitis are the two most common oral diseases and major causes of tooth loss, and can compromise nutrition and reduce quality of life [2, 3]. Their consequences extend beyond physical health, influencing psychosocial well-being and increasing mortality [4-6]. By 2050, the global burden is projected to rise substantially, with severe periodontitis affecting more than 1.5 billion people and

edentulism exceeding 660 million cases [7]. Similarly, the prevalence of permanent caries is projected to increase by more than 20% [8]. Poor oral health also exerts a considerable economic burden worldwide; in 2019, the combined global costs of oral diseases were estimated at approximately US$710 billion, reflecting substantial healthcare spending on treatment as well as productivity losses [9].

In Singapore, oral health challenges remain significant. Although complete edentulism affects about 13% of adults aged 60 years and above, a recent national survey reported that dental caries affects 85% of the population, with 35% having at least one untreated carious lesion [10, 11]. Periodontitis is highly prevalent, affecting 78% of the population, with 16% presenting with its severe form [10]. As one of the fastest-ageing societies globally [12], preventing tooth loss in its population is a key factor for preserving nutrition, well-being, and healthy longevity.

High-quality real-world data are essential for understanding the determinants and outcomes of oral disease, evaluating the quality of care, informing healthcare policy, and planning the future workforce. In oral health, routinely collected clinical data from electronic health records provide data that enable benchmarking of outcomes, detection of unwarranted variation, and the targeting of service improvements toward populations and practices with the greatest need [13, 14]. Singapore Health Services (SingHealth) functions as Singapore's largest healthcare cluster and operates as an Academic Medical Centre. Within this network, initiatives such as the Perioperative and Anesthesia Subject Area Registry (PASAR) [15], the SingHealth Diabetes Registry (SDR) [16], and the SingHealth COPD and Asthma Data Mart (SCDM) [17] illustrate how electronic health record–integrated registries can advance observational research, quality improvement, and data-driven innovation. These systems rely on data extraction from electronic health records, harmonized data architectures, rigorous quality assurance, and privacy-preserving de-identification processes. Establishing a comparable infrastructure for oral health would allow systematic study of disease patterns, trajectories, and clinical outcomes in routine care settings.

SingHealth's enterprise information technology ecosystem already supports a wide spectrum of administrative, clinical, and operational needs through its data warehouse, known as the Electronic

Health Intelligence System (eHints), which consolidates data across SingHealth institutions [18]. Leveraging this framework would facilitate the linkage and integration of oral health data with other health records, enabling the investigation of oral and broader health outcomes.

The National Dental Centre Singapore (NDCS) is the country's flagship specialist center for oral healthcare. It is the largest oral healthcare center in Singapore and one of five national specialty centers within the SingHealth cluster. By utilizing NDCS's electronic health records, the NDCS Oral Disease Registry (NDCS-ODR; hereafter referred to as ODR) catalogs diagnoses, procedures, and outcomes for individuals receiving care at the center. This paper introduces the methodology for establishing the ODR. The objectives are: (1) to provide an overview of individuals with oral diseases captured in the ODR, (2) to describe the technical and governance framework used to create the real-world data repository, and (3) to illustrate its application through a focused evaluation of data accuracy and completeness for periodontal records between 2013 and 2025. Periodontal disease was selected as the initial focus because it combines structured data with considerable analytical complexity. Accurate periodontal surveillance relies on tooth-level measurements recorded at multiple sites per tooth, which are often incomplete, require examiner calibration to minimize measurement variability, and must account for a major shift in diagnostic frameworks over the observation period. Its strong relevance to tooth survival and systemic health outcomes further makes it a clinically valuable first test of the ODR's capacity to support epidemiology, quality evaluation, and cross-registry linkage.

Materials and Methods

Source system

The registry's data source is the proprietary Electronic Dental Record (EDR) used at NDCS—developed by Integrated Health Information Systems (IHiS, now known as Synapxe). The extracted registry's domain-specific EDR data captures structured clinical information, including demographics, diagnoses, procedures, tooth condition, and periodontal charting (probing pocket depth [PPD], recession [REC] and bleeding on probing [BOP]). It does not include radiographic, administrative or

pharmacy data, and free text notes are not extracted for the registry. Figure 1 illustrates the data architecture and workflow from EDR to ODR.

Fig 1. Data architecture.
Abbreviations: DB, database; EDR, Electronic Dental Record; ODR, Oral Disease Registry.

Administration and governance

The ODR currently includes all patients who had a dental visit at NDCS between January 2013 and June 2025. The registry was established under CIRB exemption 348/2021 with a waiver of informed consent because only routinely collected, anonymized clinical data were used. Governance follows Singapore's Personal Data Protection Act, the Human Biomedical Research Act (HBRA), and the Declaration of Helsinki. Each patient is assigned a unique identifier (NRIC for citizens and permanent residents), which serves as a linkage key across visits and enables integration with other healthcare databases within SingHealth. NRICs were encrypted before storage; a linkage file permitting re-identification requires approval from the SingHealth Office of Deputy Group Chief Medical Informatics Officer (Research); and all data are stored in secure, access-controlled environments. Analysts access de-identified data through virtual desktops within the SingHealth network using corporate credentials and devices, ensuring that no data are stored or processed outside the secure infrastructure.

Researchers seeking to use ODR data follow a structured and transparent process. They begin by submitting a proposal outlining their intended analysis, which is shared with ODR investigators for feedback and potential collaboration. Once endorsed, the researcher submits a formal data request to the ODR Executive Committee for approval. External collaborators may also participate by contacting the ODR Principal Investigator, who reviews and circulates the proposal among ODR investigators for consideration.

System development

Initial discussions were held with the Principal Investigator, data manager, NDCS Chief Medical Informatics Officer, and Synapxe engineers to define the project scope and data requirements. EDR Subject area variables were identified and mapped to EDR's metadata repository or data dictionary to understand the relationships between its elements to other data, usage and formats. Screens from the NDCS EDR were systematically reviewed to identify and confirm the exact front-end fields required for extraction at record-level from a larger dataset. These were subsequently traced to their corresponding back-end tables and compiled into a User Requirements Document outlining clinical variables and inclusion rules, and a Functional and Technical Specification detailing data elements, linkage keys (tooth, site, and patient level), and data flow architecture.

The team designed the registry schema and implemented a two-step extract–transform–load process:

(i) Cohort identification, which defined eligibility criteria (i.e., patient with an initial assessment visit note), and

(ii) Data import and mapping, which integrated the relevant clinical elements.

A data harmonization step reconciled discrepancies between front-end variable names and back-end storage formats. ETL scripts were implemented and validated through internal integration testing to ensure data lineage and field-level accuracy. User acceptance testing involved three independent testers—two clinicians and a data manager—who evaluated both functionality and data fidelity using a synthetic test cohort of 45 mock patients, each with 10 scheduled visits. The UAT primarily focuses on validating data constraints and rules established across the EDR datasets, checking on the consistency and completeness of data to ensure data integrity is maintained. Test visits were distributed across fixed weekdays to emulate real operational patterns such as clinic scheduling and end-of-day data loads. Testing comprised three key domains: a) functional verification: ensuring screens and queries retrieved the expected fields, and that diagnostic and odontogram logic operated as designed (the odontogram refers to the graphical charting interface in the EDR that records each tooth's surface condition and periodontal charting; b) data verification: confirming that record counts and patient-timepoints in the ODR matched EDR extracts under identical inclusion criteria, with field-level checks on presence,

surface condition, PPD, REC, BOP, and c) soak testing: performed 10 weeks after UAT initiation to confirm that newly generated visits on the scheduled days were correctly captured by the ETL runs, simulating ongoing production operations.

Data collected

For each eligible visit, the ODR captured:

- Demographics: age, date of birth, sex, ethnicity, residency status, Singapore Housing Index, a proxy measure of socioeconomic position based on the type of public housing (e.g., 1-2 room, 3 room, 4-5 room, or private housing) [19].
- Smoking history
- Diagnosis codes
- Examination chart variables: complaints, dental history, extra-oral and intra-oral findings
- Odontogram data: presence or absence of each tooth, tooth and surface condition (including caries, restorations, crowns, root-filled teeth, and prosthetic status), and periodontal charting (PD, REC and BOP for six sites per tooth)

Data quality assessment

Inclusion criteria and extraction rules for periodontal cases

Records entered the registry for periodontal cases through the ETL process if they had: (1) an initial assessment visit or periodontics treatment note; (2) a periodontal diagnosis recorded; (3) non-empty odontogram data; and (4) odontogram data that match the visit date within ±1 day. Variables extracted included tooth presence/absence, surface conditions, PPD, REC, BOP, treatment codes, demographics, and smoking status.

Accuracy of periodontitis classification

To assess diagnostic accuracy of periodontitis cases, the registry randomly sampled records meeting two criteria: (1) complete periodontal charting (PD, REC and BOP) and (2) a recorded periodontal diagnosis from patients with a documented periodontal treatment note, indicating clinical management

185 within the periodontal unit at NDCS. Each selected case was independently classified by a calibrated
186 examiner using only the periodontal charting data, blinded to the original recorded diagnosis.
187 Classification criteria were clarified in advance through discussions with a senior epidemiologist to
188 ensure consistency. Agreement between the examiner's classification and the registry diagnosis was
189 then assessed.

190 Two diagnostic systems were used over the observation period: the 1999 Classification of Periodontal
191 Diseases and Conditions (CPDC) classification (used up to April 2019), and the 2018 European
192 Federation of Periodontology/American Academy of Periodontology (EFP/AAP) classification (used
193 from May 2019). For the latter, only staging was assessed due to a lack of radiographs. Classifications
194 for 1999 CPDC follow Armitage [20], while those for the 2018 EFP/AAP classification follow Tonetti
195 et al [21]. Power calculations indicated that 35 cases per period were sufficient to estimate a 90%
196 agreement rate with 95% confidence. Consequently, 35 cases were randomly sampled within each
197 period, resulting in a total of 140 cases stratified approximately equally across four periods to account
198 for clinician turnover and potential variations in diagnostic practices: January 2013-February 2016 (38
199 months), March 2016-April 2019 (38 months), May 2019-June 2022 (38 months), and July 2022-June
200 2025 (36 months).

201 *Completeness of periodontal charting*

202 Completeness was assessed at both the site- and patient-level. The analysis began with quantifying the
203 total number of missing sites across the entire dataset to provide an overall measure of data quality. The
204 proportion of records with 100% charting completeness was then determined, where completeness was
205 defined as non-missing values for probing depth (PD), gingival recession (REC), and bleeding on
206 probing (BOP) at all six sites per present tooth. To examine changes in charting quality over the course
207 of treatment per patient, the percentage of patients with fully complete charts at their first two visits was
208 evaluated. Analyses for the current phase are limited to the first two visits per patient. Future updates
209 will incorporate additional visits once further data extraction is completed. In parallel, a multiple-
210 imputation approach was developed and validated to address incomplete periodontal charting in this

registry, enabling bias-reduced estimation of disease prevalence and longitudinal patterns from routinely collected data [22].

Results

Characteristics of the ODR cohort

From 2013 to mid-2025, the ODR captured 229,249 unique patients (Table 1). The sex distribution was roughly balanced throughout, and the cohort's composition was predominantly Chinese (~78%) and Singapore citizens (~93%). The cohort aged over time: mean age rose steadily from the late-40s in the to the mid-50s by 2023–2025. Smoking status was dominated by never-smokers across all years, with a modest decline in current smoking over time. Most clinical variables showed right-skewed distributions: median values were zero and means were low, indicating that while most patients had no recorded events for these measures, a smaller proportion exhibited higher counts (Table 2).

Table 1. Characteristics of patients in the NDCS Oral Disease Registry, 2013–2025.

Characteristic	**Overall (N= 229,249)**	**2013 (N= 5,317)**	**2014 (N= 17,237)**	**2015 (N= 17,414)**	**2016 (N= 21,522)**	**2017 (N= 20,840)**	**2018 (N= 20,784)**	**2019 (N= 20,333)**	**2020 (N= 16,859)**	**2021 (N= 21,069)**	**2022 (N= 20,060)**	**2023 (N= 18,369)**	**2024 (N= 19,325)**	**2025* (N= 10,120)**
Age, mean (SD)	49.1 (19.5)	46.3 (18.8)	46.5 (19.2)	47.9 (18.9)	46.8 (19.1)	47.0 (19.4)	47.6 (19.6)	49.1 (19.5)	49.9 (19.0)	48.9 (19.0)	50.6 (19.4)	52.6 (19.8)	52.6 (20.0)	53.0 (20.2)
Gender														
Female	114,306 (49.9%)	2,749 (51.7%)	8,726 (50.6%)	8,740 (50.2%)	10,469 (48.6%)	10,122 (48.6%)	10,117 (48.7%)	9,921 (48.8%)	8,175 (48.5%)	10,579 (50.2%)	10,178 (50.7%)	9,485 (51.6%)	9,862 (51.0%)	5,183 (51.2%)
Male	114,943 (50.1%)	2,568 (48.3%)	8,511 (49.4%)	8,674 (49.8%)	11,053 (51.4%)	10,718 (51.4%)	10,667 (51.3%)	10,412 (51.2%)	8,684 (51.5%)	10,490 (49.8%)	9,882 (49.3%)	8,884 (48.4%)	9,463 (49.0%)	4,937 (48.8%)
Ethnicity														
Chinese	177,827 (77.6%)	4,040 (76.0%)	13,119 (76.1%)	13,399 (76.9%)	16,537 (76.8%)	16,199 (77.7%)	16,215 (78.0%)	15,847 (77.9%)	12,770 (75.7%)	16,075 (76.3%)	15,842 (79.0%)	14,620 (79.6%)	15,258 (79.0%)	7,906 (78.1%)
Malay	19,842 (8.7%)	486 (9.1%)	1,560 (9.1%)	1,541 (8.8%)	1,930 (9.0%)	1,789 (8.6%)	1,924 (9.3%)	1,817 (8.9%)	1,543 (9.2%)	1,768 (8.4%)	1,616 (8.1%)	1,382 (7.5%)	1,615 (8.4%)	871 (8.6%)
Indian	19,362 (8.4%)	464 (8.7%)	1,504 (8.7%)	1,493 (8.6%)	1,876 (8.7%)	1,817 (8.7%)	1,611 (7.8%)	1,666 (8.2%)	1,570 (9.3%)	2,019 (9.6%)	1,594 (7.9%)	1,444 (7.9%)	1,481 (7.7%)	823 (8.1%)
Others	12,218 (5.3%)	327 (6.2%)	1,054 (6.1%)	981 (5.6%)	1,179 (5.5%)	1,035 (5.0%)	1,034 (5.0%)	1,003 (4.9%)	976 (5.8%)	1,207 (5.7%)	1,008 (5.0%)	923 (5.0%)	971 (5.0%)	520 (5.1%)
Residency Status														
Citizen	212,000 (92.5%)	4,811 (90.5%)	15,744 (91.3%)	15,990 (91.8%)	19,903 (92.5%)	19,383 (93.0%)	19,366 (93.2%)	19,006 (93.5%)	15,403 (91.4%)	19,100 (90.7%)	18,517 (92.3%)	17,206 (93.7%)	18,102 (93.7%)	9,469 (93.6%)
PR	11,253 (4.9%)	239 (4.5%)	787 (4.6%)	795 (4.6%)	969 (4.5%)	877 (4.2%)	933 (4.5%)	885 (4.4%)	927 (5.5%)	1,427 (6.8%)	1,149 (5.7%)	864 (4.7%)	903 (4.7%)	498 (4.9%)
Others	5,991 (2.6%)	267 (5.0%)	705 (4.1%)	629 (3.6%)	649 (3.0%)	579 (2.8%)	484 (2.3%)	441 (2.2%)	529 (3.1%)	542 (2.6%)	394 (2.0%)	299 (1.6%)	320 (1.7%)	153 (1.5%)
Missing	5 (0.0%)	0 (0.0%)	1 (0.0%)	0 (0.0%)	1 (0.0%)	1 (0.0%)	1 (0.0%)	1 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Smoker														
Current	16,697 (7.3%)	397 (7.5%)	1,399 (8.1%)	1,508 (8.7%)	1,906 (8.9%)	1,605 (7.7%)	1,512 (7.3%)	1,362 (6.7%)	1,205 (7.1%)	1,522 (7.2%)	1,358 (6.8%)	1,029 (5.6%)	1,255 (6.5%)	639 (6.3%)
Never	187,780 (81.9%)	4,201 (79.0%)	13,649 (79.2%)	14,047 (80.7%)	17,411 (80.9%)	16,864 (80.9%)	16,944 (81.5%)	16,578 (81.5%)	13,810 (81.9%)	17,471 (82.9%)	16,768 (83.6%)	15,377 (83.7%)	16,205 (83.9%)	8,455 (83.5%)
Former	5,780 (2.5%)	159 (3.0%)	596 (3.5%)	556 (3.2%)	638 (3.0%)	536 (2.6%)	514 (2.5%)	487 (2.4%)	414 (2.5%)	505 (2.4%)	403 (2.0%)	265 (1.4%)	439 (2.3%)	268 (2.6%)
Missing	18,992 (8.3%)	560 (10.5%)	1,593 (9.2%)	1,303 (7.5%)	1,567 (7.3%)	1,835 (8.8%)	1,814 (8.7%)	1,906 (9.4%)	1,430 (8.5%)	1,571 (7.5%)	1,531 (7.6%)	1,698 (9.2%)	1,426 (7.4%)	758 (7.5%)

SHI (Socioeconomic indicator)														
<3.0	14,785 (6.4%)	327 (6.2%)	1,089 (6.3%)	1,178 (6.8%)	1,371 (6.4%)	1,285 (6.2%)	1,348 (6.5%)	1,409 (6.9%)	1,085 (6.4%)	1,199 (5.7%)	1,261 (6.3%)	1,204 (6.6%)	1,320 (6.8%)	709 (7.0%)
3.0 to <4.0	38,372 (16.7%)	900 (16.9%)	2,908 (16.9%)	3,113 (17.9%)	3,546 (16.5%)	3,441 (16.5%)	3,465 (16.7%)	3,403 (16.7%)	2,825 (16.8%)	3,455 (16.4%)	3,376 (16.8%)	3,081 (16.8%)	3,215 (16.6%)	1,644 (16.2%)
4.0 to <6.0	91,347 (39.8%)	2,019 (38.0%)	6,609 (38.3%)	6,821 (39.2%)	8,525 (39.6%)	8,312 (39.9%)	8,247 (39.7%)	8,335 (41.0%)	6,744 (40.0%)	8,512 (40.4%)	7,970 (39.7%)	7,298 (39.7%)	7,805 (40.4%)	4,150 (41.0%)
6.0	82,900 (36.2%)	2,010 (37.8%)	6,426 (37.3%)	6,101 (35.0%)	7,868 (36.6%)	7,587 (36.4%)	7,555 (36.4%)	7,008 (34.5%)	6,098 (36.2%)	7,802 (37.0%)	7,344 (36.6%)	6,673 (36.3%)	6,862 (35.5%)	3,566 (35.2%)
Missing	1,845 (0.8%)	61 (1.1%)	205 (1.2%)	201 (1.2%)	212 (1.0%)	215 (1.0%)	169 (0.8%)	178 (0.9%)	107 (0.6%)	101 (0.5%)	109 (0.5%)	113 (0.6%)	123 (0.6%)	51 (0.5%)

Footnotes:

(1) Values represent unique patients, counted once across the entire study period even if they were seen in multiple years.

(2) Singapore Housing Index (SHI) is a building-level socioeconomic indicator based on the weighted average number of rooms per dwelling within a residential building (range 1-6), with higher values indicating higher socioeconomic position.

*Data for 2025 are up to June.

Abbreviations: N, number; PR, permanent resident, SD, standard deviation, SHI, Singapore Housing Index.

Table 2. Summary statistics of clinical variables recorded.

Variable	Minimum	Mean	Median (P50)	P25	P75	P90	P95	Maximum	SD
Missing teeth	0	7.87	5	3	10	20	26	32	7.71
Crowns	0	0.53	0	0	0	2	3	32	1.49
Implants	0	0.08	0	0	0	0	0	30	0.62
Root-filled teeth	0	0.15	0	0	0	0	1	25	0.66
Prosthetic units	0	0.93	0	0	0	3	6	32	2.55
Carious surfaces	0	1.21	0	0	1	4	6	91	2.96
Carious teeth	0	0.74	0	0	1	2	4	27	1.60
Restored surfaces	0	4.77	3	0	7	13	17	105	6.22
Restored teeth	0	2.81	2	0	4	7	10	29	3.36

Footnote: All values represent counts per patient at baseline visit.
Abbreviations: SD, standard deviation; P25/P50/P75, 25th/50th/75th percentile.

Periodontal diagnoses and service activity

The overall distribution of periodontal diagnoses is presented in Table 3, with the full list in S1 Table. Among cases classified with the 2018 EFP/AAP system, Stage III disease predominated, with both localized and generalized presentations represented across grades. For diagnoses recorded under the 1999 CPDC Classification, localized and generalized severe chronic periodontitis were the most common categories. Periodontal service activity by year and visit type is summarized in S2 Table. Visit counts increased over time, with relatively lower counts observed in the earlier years of the registry period, followed by a marked increase from 2019 onwards. A transient drop in visits was observed in 2020, consistent with reduced clinical activity during the COVID-19 pandemic, followed by recovery in subsequent years.

Table 3. Distribution of periodontal diagnoses in the ODR.

Diagnosis category	N	%
EFP/AAP Stage I–II periodontitis	2,814	2.59
EFP/AAP Stage III periodontitis	27,051	24.93
EFP/AAP Stage IV periodontitis	9,188	8.47
CPDC severe chronic periodontitis	13,131	12.10
CPDC mild/moderate chronic periodontitis	8,542	7.87
CPDC aggressive periodontitis	460	0.42
Gingivitis (localized/generalized)	9,995	9.21
Peri-implantitis/Peri-implant mucositis	2970	2.74
Acute periodontal lesions (abscess / endo-periodontal)	5,896	5.43
Other periodontal conditions*	5,603	5.16
Unspecified periodontal conditions	22867	21.07
Overall	108,517	100.00

Footnote: All values represent counts per patient at baseline visit. Other periodontal conditions include drug-induced gingival overgrowth; occlusal trauma; gingival recession; periodontitis modified by systemic disease. See S1 Table for a full listing and counts.
Abbreviations: CPDC, 1999 Classification of Periodontal Diseases and Conditions classification; EFP/AAP, 2018 European Federation of Periodontology/American Academy of Periodontology classification.

Diagnosis agreement

Agreement between examiner-derived and recorded periodontal diagnoses varied by period (Table 4). The highest concordance was observed during the final years when the 1999 CPDC classification was still in use and in the most recent period, with lower agreement noted in the earliest ODR years and during the initial implementation of the 2018 EFP/AAP classification. After combining all diagnoses, unweighted Cohen's kappa indicated substantial agreement: $\kappa = 0.84$ (95% CI 0.78–0.90).

Table 4. Agreement between examiner-derived and recorded periodontal diagnoses across three time periods.

Time Period	N	Agreement (%)	95% CI
January 2013-February 2016	35	80.0	66.7-93.3
March 2016-April 2019	35	91.4	82.2-100.0
May 2019-June 2022	35	82.9	70.4-95.3
July 2022-June 2025	35	91.4	82.2-100.0

Abbreviations: N, number; CI, confidence interval.

Data completeness

Data completeness was low, with REC showing the highest proportion of missing data (Table 5). Completeness improved at follow-up compared with initial visits, likely because second visits typically corresponded to periodontal treatment appointments, who recorded more comprehensive charting. Nevertheless, PPD and BOP were only slightly more than half complete. At the site level, REC similarly exhibited the greatest degree of incompleteness across parameters (S3 Table).

Table 5. Completeness of periodontal charting.

	BOP	**PPD**	**REC**
All odontograms: 100% site-level completeness	55,648 / 276,907 (20.1%)	69,303 / 276,907 (25.0%)	35,563 / 276,907 (12.8%)
First visit only	29,547 / 229,249 (12.9%)	41,928 / 229,249 (18.3%)	17,794 / 229,249 (7.8%)
Second visit only	16,082 / 35,042 (45.9%)	17,460 / 35,042 (49.8%)	10,331 / 35,042 (29.5%)

Footnote: Completeness is defined as non-missing values for all six sites per tooth within each odontogram.
Abbreviations: BOP, bleeding on probing; PPD, probing pocket depth; REC, gingival recession.

Discussion

This study describes the establishment of the ODR, including its governance, data architecture, and the processes used to extract and harmonize structured electronic dental record data between 2013 and 2025. In the focused periodontal evaluation, examiner-derived and recorded diagnoses showed substantial agreement, supporting the use of registry-coded periodontitis diagnoses for epidemiologic analyses. However, periodontal charting completeness—particularly for gingival recession—was limited, indicating that analyses relying on full clinical charting may be affected by under-recording and should incorporate strategies to address missing data and documentation variability.

Routine electronic health records accumulate large volumes of patient data and can be utilized to generate real-world evidence about oral diseases. There is increasing recognition that secondary use of electronic health record data can improve the knowledge base in dentistry, enabling questions to be addressed regarding treatment effectiveness and the links between oral and overall health [23]. Clinical

trials, while rigorous, are costly and often involve highly selective populations that may not reflect real-world treatment effects [24]. Administrative or claims databases, on the other hand, capture all patient interactions with the healthcare system but lack important clinical details such as examination findings, diagnostic test results, and prescribed medications [25]. Administrative databases are also limited in their ability to explore clinical and contextual factors underlying oral health disparities, such as differences in disease burden and access to care across population groups [26]. Although data leakage or misclassification can occur, electronic health records occupy a middle ground by combining the clinical richness of chart data with the large scale that claims data provide.

Although large-scale electronic dental record repositories are still relatively uncommon globally, several initiatives show that routinely collected dental data can be curated for epidemiological and outcomes research. In the United States, the BigMouth Dental Data Repository was established in 2012 using electronic data from four academic dental schools; it initially aggregated 1.1 million patient records[23], and has since expanded to 18 schools with over 6 million records [27]. Practice-based research networks have also demonstrated the feasibility of using standardized electronic dental record data across distributed clinics for multi-site observational studies [28]. In Nordic countries, national dental registries—such as Denmark's National Child Odontology Registry [29] and Sweden's Dental Health Register [30]—further illustrate how structured dental data can support population surveillance research.

In contrast, the ODR is a single-center registry based on structured data. It leverages SingHealth's enterprise data infrastructure for warehousing and governance, but currently encompasses only one of Singapore's two national specialty centers. Despite its narrower scope, the ODR includes more than 229,000 unique patients over 12 years and provides detailed, tooth- and site-level odontogram data on caries and periodontal status. As the ODR is embedded within an integrated health system, it enables linkage of oral health data with medical databases, capabilities not available in existing dental repositories such as BigMouth. Therefore, the ODR provides unique opportunities to study oral and systemic relationships.

The ODR has approximately equal numbers of male and female patients and an ethnic distribution comprising predominantly Chinese individuals. The mean age increased over the study period, suggesting that older adults are increasingly seeking specialist oral care. These demographic trends are broadly consistent with findings from national-representative oral health surveys [10], although the ODR included fewer Malay and permanent-resident patients, and proportionally more citizens. The differences in residency status compared to Singapore proportion are likely attributable to referral pathways and subsidy structures within the public healthcare system. Clinical variables were markedly skewed, with median counts for many parameters at zero and means close to zero. This likely reflects a mix of patients with low prior treatment burden and those attending for initial assessment or treatment planning, where treatment needs had not yet been recorded.

Only about one-quarter of odontograms had complete charting for PPD and one-fifth for bleeding on probing (BOP), and just 12.8% were complete for REC. These findings are broadly consistent with evaluations from other dental repositories such as BigMouth, which similarly reported high completeness for demographic variables but variable completeness for clinical measures, often influenced by provider-level differences. Incomplete periodontal charting is associated with systematic underestimation of disease burden [31]. However, analyses using ODR periodontal records showed that multiple imputation can substantially reduce this bias, particularly for mild-to-moderate disease categories, while preserving overall diagnostic patterns observed in the registry [22]. Agreement between examiner-derived and recorded diagnoses varied across time periods, with higher concordance observed in 2016–2019 and the most recent period. This likely reflected clinicians' familiarity with the 1999 CPDC classification and ongoing calibration exercises linked to concurrent clinical studies. The subsequent dip in agreement may be explained by the transition to the newer 2018 EFP/AAP diagnostic framework and the associated learning curve during its early implementation. However, the overall level of agreement was substantial, demonstrating that periodontal diagnoses within the registry were largely concordant.

Lessons learnt

Electronic health records often operate in fragmented and inefficient systems, with limited interoperability and substantial data processing required to extract usable information. The lack of consistent structure and standardization further compounds these inefficiencies that hinder integration and secondary analysis [32]. Our experience revealed several operational challenges. First, initial assessment visit and treatment notes must be standardized and completed consistently. The relatively low number of recorded periodontal visits prior to 2019 likely reflects inconsistent use of structured periodontal note templates rather than true service volume. Second, data extraction depended on structured fields; free-text notes and radiographs were excluded, limiting capture of clinical data. Radiographs and prescriptions are stored in separate systems, further complicating data linkage.

Third, electronic health records are primarily designed for patient care rather than research, and substantial effort was required to harmonize variable names, map relationships from tooth to site to patient, and resolve inconsistencies. Finally, registry development requires close involvement of diverse stakeholders. This includes researchers, data managers, engineers, and clinicians familiar with the electronic health records system. Strong administrative leadership is also critical to navigate governance processes and ensure timely approvals. Clear governance structures and sufficient resource support are also necessary to maintain the registry's long-term viability. The results from the registry can also be utilized for clinical audits, for example, by identifying gaps in documentation, and standardizing charting practices.

Limitations and Strengths

The ODR is not without limitations. The ODR represents a single national specialty center and may not represent the broader community disease burden; the cohort is skewed toward citizens, and individuals seeking specialist care. Data completeness is limited. A particular challenge in dentistry is the inherently hierarchical structure of the oral cavity, where each patient has multiple teeth, each with several charted sites. This making comprehensive charting time-consuming in routine workflows. Moreover, the 1999 CPDC and 2018 EFP/AAP classifications require complete site-level data, a standard difficult to achieve in everyday clinical practice. Standardizing charting procedures and clinician training at NDCS will therefore be essential to improve data completeness and diagnostic reliability. A substantial

proportion of visits carried unspecified periodontal codes, which is expected as many of these encounters were likely managed by clinicians from other specialties as part of routine triage, rather than by periodontists assigning definitive diagnoses. Smoking history was self-reported and may also have been under-recorded.

Nevertheless, this study has strengths. It covers more than a decade of care for a large number of unique patients. The use of a unique national identifier for each citizen or permanent resident (NRIC) enables linkage to other SingHealth and national registries, providing opportunities to explore oral–systemic relationships. Standardization of data elements within the ODR enhances data quality and facilitates interoperability across healthcare domains. The calibrated diagnosis demonstrated substantial agreement, suggesting that clinical charting can be used reliably for epidemiological research.

Another strength of this study is the use of the SHI as a proxy for socioeconomic position [19]. In Singapore, home ownership and housing type serve as well-established socioeconomic indicators [33]. Individual-level data such as income or education are not routinely captured in electronic health records and may be subject to reporting and social desirability biases when self-reported. Area-level indices, such as neighborhood-based socioeconomic position measures, have been developed locally, but remain prone to ecological fallacy [34]. The SHI overcomes these limitations by providing building-level granularity, which retains privacy while approximating individual socio-economic status [35]. It is derived from open data Application Programming Interfaces (APIs) and represents a weighted average number of rooms per residential unit, scaled from 1 to 6. While it does not distinguish between private housing types (e.g., condominiums or landed properties), it preserves the ordinal housing-value gradient across the population (S4 Table). The SHI has been validated and shown to correlate with population-level indicators such as bystander cardiopulmonary resuscitation rates [36] and vaccination uptake [37].

Future directions

To our knowledge, the ODR is the first large-scale oral disease registry in Asia established within a national specialty center. Through this, we hope to contribute to the growing literature on real-world oral health research and encourage collaborations that span registries and health systems. Ongoing work

will extend the ODR to include data from other dental specialties (e.g., pediatric dentistry). Incorporation of clinical images and prescription data, as well as natural language processing of free-text notes will enhance data richness and research utility. We hope for the registry to be refreshed at six-month intervals, subject to operational feasibility and resource availability.

The utilization of the ODR represents a large-scale effort to integrate electronic oral health records within a major academic medical center in Singapore. This registry provides a valuable platform for research and healthcare service evaluation. Future expansion of its data scope and linkages will further enhance its utility, supporting quality assessment, care improvement, and targeted interventions aimed at reducing the burden of oral disease nationally.

Acknowledgements

JRHT received PhD funding from the Singapore National Medical Research Council (NMRC) Research Training Fellowship (RTF24jan-0007). We also thank Bee Tin Goh, Chu Guan Koh, Li Qian, Dian Yi Chow, and Huihua Li for their valuable contributions and support in the development and implementation of the Oral Disease Registry.

Availability of data

The data that support the findings of this study are available from National Dental Centre Singapore. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the authors upon reasonable request, with the permission of National Dental Centre Singapore.

References

1. Bernabe E, Marcenes W, Abdulkader RS, Abreu LG, Afzal S, Alhalaiqa FN, et al. Trends in the global, regional, and national burden of oral conditions from 1990 to 2021: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet. 2025;405(10482):897-910. doi: 10.1016/S0140-6736(24)02811-3.

2. Lexomboon D, Kumar A, Freyland S, Xu W, Sandborgh-Englund G. Is poor chewing ability a risk factor for malnutrition? A six-year longitudinal study of older adults in Sweden. The Journal of nutrition, health and aging. 2025;29(6):100554. doi: https://doi.org/10.1016/j.jnha.2025.100554.

3. McGrogan C, Matcham F, Dawes H, Ang CS, Cartner H, Branley-Bell D. Exploring relationships between chewing difficulty, eating experience, pain, and well-being: A cross-sectional self-report study. Clinical Nutrition Open Science. 2024;57:177-202. doi: https://doi.org/10.1016/j.nutos.2024.07.013.

4. Tay JRH, Cooray U, Chan A, Tonetti MS, Nascimento GG, Peres MA. Perceived Masticatory Function and Mortality: A Causal Study. Journal of Dental Research. 0(0):00220345251363851. doi: 10.1177/00220345251363851. PubMed PMID: 40947960.

5. Tay JRH, Nascimento GG, Chan A, Malhotra R, Tonetti MS, Peres MA. Estimating the causal impact of chewing disability on depressive symptoms mediated by loneliness: A longitudinal marginal structural model study of older adults in Singapore. Innovation in Aging. 2025:igaf100. doi: 10.1093/geroni/igaf100.

6. Wei X, Zhuang L, Zhang X, Deng K, Chen R, Liu B, et al. Association of Masticatory Function With Mortality in Older Adults. Int Dent J. 2025;75(6):103901. Epub 20250915. doi: 10.1016/j.identj.2025.103901. PubMed PMID: 40957156; PubMed Central PMCID: PMCPMC12464697.

7. Nascimento GG, Alves-Costa S, Romandini M. Burden of severe periodontitis and edentulism in 2021, with projections up to 2050: The Global Burden of Disease 2021 study. J Periodontal Res. 2024;59(5):823-67. Epub 20240827. doi: 10.1111/jre.13337. PubMed PMID: 39192495.

8. Alves-Costa S, Romandini M, Nascimento GG. Lip and Oral Cancer, Caries and Other Oral Conditions: Estimates From the 2021 Global Burden of Disease Study and Projections up to 2050. J Periodontal Res. 2025;60(6):544-58. Epub 20250618. doi: 10.1111/jre.13421. PubMed PMID: 40530949; PubMed Central PMCID: PMCPMC12312818.

9. Jevdjevic M, Listl S. Global, Regional, and Country-Level Economic Impacts of Oral Conditions in 2019. J Dent Res. 2025;104(1):17-21. Epub 20241113. doi:

10.1177/00220345241281698. PubMed PMID: 39535193; PubMed Central PMCID: PMCPMC11662506.

10. Sim CPC, Lee YH, Sim YF, Wong J, Soh SH, Goh BT, et al. Findings from the 2019 nationally representative oral health survey for adults in Singapore. Community Dent Oral Epidemiol. 2024;52(3):281-91. doi: 10.1111/cdoe.12961. PubMed PMID: 38747365.

11. Ministry of Health. Edentulism 2024. Available from: https://www.moh.gov.sg/newsroom/edentulism.

12. Malhotra R, Bautista MAC, Müller AM, Aw S, Koh GCH, Theng YL, et al. The Aging of a Young Nation: Population Aging in Singapore. Gerontologist. 2019;59(3):401-10. doi: 10.1093/geront/gny160. PubMed PMID: 30517628.

13. Hummel R, Bruers J, van der Galiën O, van der Sanden W, van der Heijden G. Outcome measures for oral health based on clinical assessments and claims data: feasibility evaluation in practice. BMC Oral Health. 2017;17(1):125. doi: 10.1186/s12903-017-0410-5.

14. Choi SE, Pandya A, White J, Mertz E, Normand SL. Quality Measure Adherence and Oral Health Outcomes in Children. JAMA Netw Open. 2024;7(1):e2353861. Epub 20240102. doi: 10.1001/jamanetworkopen.2023.53861. PubMed PMID: 38289601; PubMed Central PMCID: PMCPMC10828912.

15. Abdullah HR, Lim DYZ, Ke Y, Salim NNM, Lan X, Dong Y, et al. The SingHealth Perioperative and Anesthesia Subject Area Registry (PASAR), a large-scale perioperative data mart and registry. Korean J Anesthesiol. 2024;77(1):58-65. Epub 20231108. doi: 10.4097/kja.23580. PubMed PMID: 37935575; PubMed Central PMCID: PMCPMC10834714.

16. Lim DYZ, Chia SY, Abdul Kadir H, Mohamed Salim NN, Bee YM. Establishment of the SingHealth Diabetes Registry. Clin Epidemiol. 2021;13:215-23. Epub 20210316. doi: 10.2147/clep.S300663. PubMed PMID: 33762850; PubMed Central PMCID: PMCPMC7982443.

17. Lam SSW, Fang AHS, Koh MS, Shantakumar S, Yeo SH, Matchar DB, et al. Development of a real-world database for asthma and COPD: The SingHealth-Duke-NUS-GSK COPD and Asthma Real-World Evidence (SDG-CARE) collaboration. BMC Med Inform Decis Mak. 2023;23(1):4. Epub

20230109. doi: 10.1186/s12911-022-02071-6. PubMed PMID: 36624490; PubMed Central PMCID: PMCPMC9830781.

18. Synapxe. SingHealth-Synapxe Electronic Health Intelligence System 2025. Available from: https://www.synapxe.sg/healthtech/data-and-analytics/electronic-health-intelligence-system.

19. Lim DYZ, Wong TH, Feng M, Ong MEH, Ho AFW. Leveraging open data to reconstruct the Singapore Housing Index and other building-level markers of socioeconomic status for health services research. Int J Equity Health. 2021;20(1):218. Epub 20211003. doi: 10.1186/s12939-021-01554-8. PubMed PMID: 34602083; PubMed Central PMCID: PMCPMC8489093.

20. Armitage GC. Development of a classification system for periodontal diseases and conditions. Ann Periodontol. 1999;4(1):1-6. doi: 10.1902/annals.1999.4.1.1. PubMed PMID: 10863370.

21. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. J Periodontol. 2018;89 Suppl 1:S159-s72. doi: 10.1002/jper.18-0006. PubMed PMID: 29926952.

22. Liu P, Tay JRH, Peres MA. Recovering incomplete clustered binary data in longitudinal electronic health records using multiple imputation

23. Walji MF, Kalenderian E, Stark PC, White JM, Kookal KK, Phan D, et al. BigMouth: a multi-institutional dental data repository. J Am Med Inform Assoc. 2014;21(6):1136-40. Epub 20140703. doi: 10.1136/amiajnl-2013-002230. PubMed PMID: 24993547; PubMed Central PMCID: PMCPMC4215035.

24. Raittio E, Baelum V. Assessing the Impact of Periodontal Therapy on Tooth Loss: A Register-Based Longitudinal Study in Denmark. Community Dent Oral Epidemiol. 2025. Epub 20250926. doi: 10.1111/cdoe.70026. PubMed PMID: 41013953.

25. Schneeweiss S, Desai RJ, Ball R. A future of data-rich pharmacoepidemiology studies: transitioning to large-scale linked electronic health record + claims data. Am J Epidemiol. 2025;194(2):315-21. doi: 10.1093/aje/kwae226. PubMed PMID: 39013780; PubMed Central PMCID: PMCPMC11815500.

26. Tiwari T, Patel JS, Nascimento GG. Big Data and Oral Health Disparities: A Critical Appraisal. J Dent Res. 2025;104(2):119-30. Epub 20241204. doi: 10.1177/00220345241285847. PubMed PMID: 39629938.

27. UTHealth Houston. BigMouth Dental Data Repository 2025. Available from: https://www.uth.edu/bigmouth/#:~:text=About%20BigMouth.

28. Thyvalikakath TP, Duncan WD, Siddiqui Z, LaPradd M, Eckert G, Schleyer T, et al. Leveraging Electronic Dental Record Data for Clinical Research in the National Dental PBRN Practices. Appl Clin Inform. 2020;11(2):305-14. Epub 20200429. doi: 10.1055/s-0040-1709506. PubMed PMID: 32349142; PubMed Central PMCID: PMCPMC7190390.

29. Nygaard N, Ängquist L, Belstrøm D, Stankevic E, Hansen T, Olsen A, et al. The national child odontology registry (SCOR): a valuable resource for odontological and public health research. BMC Oral Health. 2023;23(1):608. doi: 10.1186/s12903-023-03199-1.

30. Ljung R, Lundgren F, Appelquist M, Cederlund A. The Swedish dental health register - validation study of remaining and intact teeth. BMC Oral Health. 2019;19(1):116. Epub 20190617. doi: 10.1186/s12903-019-0804-7. PubMed PMID: 31208416; PubMed Central PMCID: PMCPMC6580593.

31. Susin C, Kingman A, Albandar JM. Effect of partial recording protocols on estimates of prevalence of periodontal disease. J Periodontol. 2005;76(2):262-7. doi: 10.1902/jop.2005.76.2.262. PubMed PMID: 15974851.

32. Blumenthal S. Improving Interoperability between Registries and EHRs. AMIA Jt Summits Transl Sci Proc. 2018;2017:20-5. Epub 20180518. PubMed PMID: 29888033; PubMed Central PMCID: PMCPMC5961768.

33. Low LL, Wah W, Ng MJ, Tan SY, Liu N, Lee KH. Housing as a Social Determinant of Health in Singapore and Its Association with Readmission Risk and Increased Utilization of Hospital Services. Front Public Health. 2016;4:109. Epub 20160530. doi: 10.3389/fpubh.2016.00109. PubMed PMID: 27303662; PubMed Central PMCID: PMCPMC4884736.

34. Earnest A, Ong MEH, Shahidah N, Chan A, Wah W, Thumboo J. Derivation of indices of socioeconomic status for health services research in Asia. Preventive Medicine Reports. 2015;2:326-32. doi: https://doi.org/10.1016/j.pmedr.2015.04.018.

35. Danesh J, Gault S, Semmence J, Appleby P, Peto R. Postcodes as useful markers of social class: population based study in 26 000 British households. Bmj. 1999;318(7187):843-4. doi: 10.1136/bmj.318.7187.843. PubMed PMID: 10092262; PubMed Central PMCID: PMCPMC27800.

36. Ho AFW, Ting PZY, Ho JSY, Fook-Chong S, Shahidah N, Pek PP, et al. The Effect of Building-Level Socioeconomic Status on Bystander Cardiopulmonary Resuscitation: A Retrospective Cohort Study. Prehosp Emerg Care. 2023;27(2):205-12. Epub 20220428. doi: 10.1080/10903127.2022.2061094. PubMed PMID: 35363103.

37. Seah SJ, Lim GH, Yoon S, Bee YM. Sociodemographic, health and behavioral factors associated with influenza vaccination uptake among patients with diabetes in Singapore: A cohort registry-based study. Prev Med Rep. 2025;56:103136. Epub 20250612. doi: 10.1016/j.pmedr.2025.103136. PubMed PMID: 40599239; PubMed Central PMCID: PMCPMC12208936.

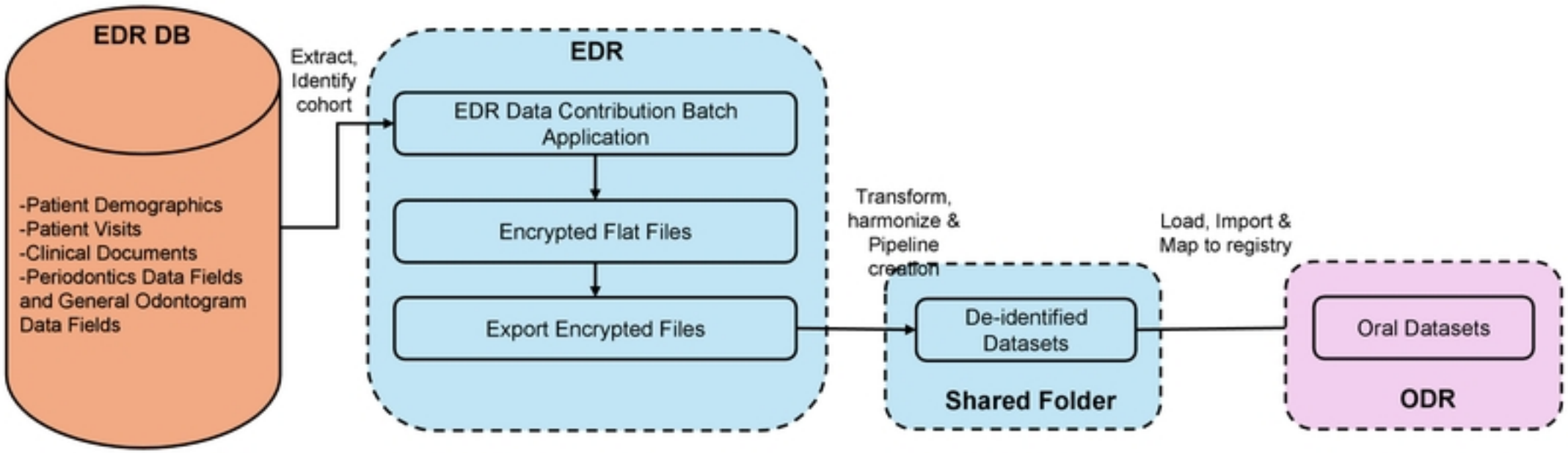
EDR DB
-Patient Demographics
-Patient Visits
-Clinical Documents
-Periodontics Data Fields and General Odontogram Data Fields
Extract, Identify cohort
EDR
EDR Data Contribution Batch Application
Encrypted Flat Files
Export Encrypted Files
Transform, harmonize & Pipeline creation
De-identified Datasets
Shared Folder
Load, Import & Map to registry
Oral Datasets
ODR

Fig 1