

Applicability of existing Gender Scores on German clinical research data: A critical assessment using the Core Dataset of the German Medical Informatics Initiative

Lea Schindler, Hilke Beelich, Elpiniki Katsari, Sylvia Stracke, Dagmar Waltemath

Submitted to: JMIR Medical Informatics
on: March 19, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 31

 Figures 32

 Figure 1..... 33

 Figure 2..... 34

 Multimedia Appendixes 35

 Multimedia Appendix 1..... 36

Applicability of existing Gender Scores on German clinical research data: A critical assessment using the Core Dataset of the German Medical Informatics Initiative

Lea Schindler¹; Hilke Beelich¹; Elpiniki Katsari² Dr med; Sylvia Stracke³ Dr med; Dagmar Waltemath^{1,4} Dr ing

¹ Medical Informatics Laboratory University Medicine Greifswald Greifswald DE

² Heart Surgery University Medicine Greifswald Greifswald DE

³ Internal Medicine A, Nephrology University Medicine Greifswald Greifswald DE

⁴ Core Unit Data Integration Center University Medicine Greifswald Greifswald DE

Corresponding Author:

Dagmar Waltemath Dr ing

Medical Informatics Laboratory
University Medicine Greifswald
Walther-Rathenau-Straße 48
Greifswald
DE

Abstract

Background: Considering sex and gender improves research quality, innovation, and social equity, while ignoring them leads to inaccuracies, and inefficiency in study results. Despite increasing attention on sex- and gender-sensitive medicine, challenges remain in accurately representing gender due to its dynamic and context-specific nature.

Objective: This work aims to contribute to the implementation of a standard for collecting and assessing gender-specific data in German university hospitals and associated research facilities.

Methods: We carried out a review to identify and categorize state-of-the-art gender scores. 22 publications were systematically assessed regarding the applicability and practicability of existing gender scores. Specifically, we evaluated the use of these gender scores on German research data from clinical routine, using the Medical Informatics Initiative core dataset (MII CDS).

Results: Different methods for assessing gender have been proposed, but no standardized and validated gender score is available for health research. Most gender scores target epidemiological or public health research where questions about social aspects and life habits are already part of the questionnaires. However, it is much more challenging to apply concepts for gender scoring on clinical data. The MII CDS, for example, lacks all variables currently being recorded in gender scores. However, some of the required variables are indeed present in clinical routine data, but need to become part of the MII CDS.

Conclusions: To enable gender-specific retrospective analysis of clinical routine data, we recommend to update and expand the MII CDS by including more gender-relevant information. Clinical Trial: Not applicable.

(JMIR Preprints 19/03/2025:74162)

DOI: <https://doi.org/10.2196/preprints.74162>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/preprint/74162>, I will be able to make my manuscript PDF available to anyone. No. Please do not make my accepted manuscript PDF available to anyone.



Original Manuscript

Research Paper

Applicability of existing Gender Scores on German clinical research data: A critical assessment using the Core Data Set of the German Medical Informatics Initiative

Lea Schindler¹, Hilke Beelich¹, Elpiniki Katsari², Sylvia Stracke³, Dagmar Waltemath^{1,4} *,

¹ Medical Informatics Laboratory, University Medicine Greifswald, Greifswald, Germany.

² Heart Surgery, University Medicine Greifswald, Greifswald, Germany.

³ Internal Medicine A, Nephrology, University Medicine Greifswald, Greifswald, Germany.

⁴ Core Unit Data Integration Center, University Medicine Greifswald, Greifswald, Germany.

* corresponding author

E-mail address: lea.schindler@med.uni-greifswald.de

E-mail address: dagmar.waltemath@med.uni-greifswald.de

Phone: +49 3834 86 8370

Full Address

Universitätsmedizin Greifswald

Medical Informatics Laboratory

Walther-Rathenau-Str. 48

17475 Greifswald

Abstract**Background:**

Considering sex and gender improves research quality, innovation, and social equity, while ignoring them leads to inaccuracies, and inefficiency in study results. Despite increasing attention on sex- and gender-sensitive medicine, challenges remain in accurately representing gender due to its dynamic and context-specific nature.

Objective:

This work aims to contribute to the implementation of a standard for collecting and assessing gender-specific data in German university hospitals and associated research facilities.

Methods:

We carried out a review to identify and categorize state-of-the-art gender scores. 22 publications were systematically assessed regarding the applicability and practicability of existing gender scores. Specifically, we evaluated the use of these gender scores on German research data from clinical routine, using the Medical Informatics Initiative core dataset (MII CDS).

Results:

Different methods for assessing gender have been proposed, but no standardized and validated gender score is available for health research. Most gender scores target epidemiological or public

health research where questions about social aspects and life habits are already part of the questionnaires. However, it is much more challenging to apply concepts for gender scoring on clinical data. The MII CDS, for example, lacks all variables currently being recorded in gender scores. However, some of the required variables are indeed present in clinical routine data, but need to become part of the MII CDS.

Conclusions:

To enable gender-specific retrospective analysis of clinical routine data, we recommend to update and expand the MII CDS by including more gender-relevant information.

Trial Registration: Not applicable.

Keywords: gender score; gender research; review; clinical routine; clinical research data; German Medical Informatics Initiative;

Introduction

Background

Taking sex and gender (see Textbox 1 and 2) into account improves the quality in research and care, and it supports social equity [1]. Considering sex and gender from the outset can lead to new discovery and foster innovation. Ignoring sex and gender, however, leads to inaccuracies, inefficiency and difficulties generalizing results. This is, for example, highlighted by a recent study showing that men with low femininity reported a significant decrease in anxiety during the COVID-19 pandemic, meanwhile women with low femininity reported a significant increase [2].

Textbox 1. Definition biological sex [3, 4].

Biological sex is defined through biological attributes, such as

- o Genetics, i.e. chromosomes, gene expression
- o Hormone levels
- o Anatomy, i.e. internal and external sex organs/reproductive organs

Deviations regarding the hormonal or chromosomal attributes are possible, where the external appearance can be intersex and it is not possible to decide.

Textbox 2. Definition (social) gender [3, 4].

Gender refers to socially constructed roles, behaviors, expectations, and norms associated with being a woman, man, girl or boy. It is formed through biological and psychosexual factors, an individual's social biography, and it is shaped by someone's role in society. Gender influences how people perceive themselves and others, and how they act and interact.

Many healthcare decisions are influenced by the patients' sex and gender, through social norms and experiences. Sex- and gender-sensitive medicine receive increasing attention in international research [5]. However, several publications showed that females are still not considered equally to males and are poorly represented due to biased study designs [6, 7, 8]. It is well known today, that this practice leads to an increase of inequality in medical treatment due to inappropriate therapies.

Moreover, sex- and gender-specific reporting is often deficient, which leads to a lack of reproducibility and reduced effectiveness of research studies [1].

Even though the situation has improved, it can still be challenging to represent psychosocial gender

aspects, as gender is a context-specific construct that is dynamic and multidimensional and can differ across time, geographical regions and societies [9].

Gender scores are developed and used to assess gender. They can be created and applied retrospectively to extract frequently missing gender information from already existing data. For this purpose, gender-sensitive variables in the existing data are determined by expert knowledge or statistical algorithms. Gender scores can then be calculated to predict social gender roles. For example, Lacasse et al. [10] and Nauman et al. [11] each developed a retrospective gender score, applied on specific datasets of population studies. Examples for the corresponding gender-sensitive variables are specific *professions*, like in health care or construction, typical *personality traits*, like agreeableness or neuroticism, but also *loneliness* or *stress* were considered. In general, retrospective scores cannot capture participants' current gender identity (see Textbox 4).

Textbox 3. Definition birth-assigned sex (BAS) [3, 4].

Birth-assigned sex (BAS) refers to the information, that can be found on official documents, like a birth certificate and that was assigned based on the child's external anatomy/ sex organs.

BAS is often measured by a binary choice of response, i.e. male/female.

Biological sex is not a dichotomy, nor is BAS, where diversity is given.

Textbox 4. Definition gender identity [3, 4].

Gender identity is determined through the individual self-perception and the sexual identity of a person, which may differ from someone's physical appearance or biological sex.

Prospective assessments can enhance the collection of gender-specific data from the outset of a study by incorporating further social data or information regarding the identity. Fraser [12], and Lagos and Compton[13], for example, evaluated prospective Two-Step gender identity measures on huge studies. The Two-Step differentiates between the current gender identity and birth-assigned sex (BAS) (see Textbox 3 and 4). The given examples show, that gender scores are mostly used in surveys and studies and hardly used on clinical routine data.

All 38 university hospitals in Germany are part of the Medical Informatics Initiative [14] (MII), which serves to close the gap between research and routine care. A core dataset (CDS) is being developed by the MII consortia and the participating university hospitals to foster interoperability and allow a shared use of clinical routine data across Germany. Data Integration Centers collect and process clinical routine data and make it accessible to researchers for the secondary use.

Despite the above mentioned existing approaches for gender assessment, a gold standard is still missing for how sex and gender should be considered in research with clinical research data. A practicable score should be balanced, comprehensive and easily applicable. Interestingly, no gender score has so far been established as an international standard. In practice, outdated scores are often used without considering proper validation and potential biases [15]. We therefore conducted a systematic evaluation of gender scores regarding their applicability in health research. In this article, the term gender score refers to the assessment of gender in general (social gender and gender identity), and the term gender identity measure refers exclusively to gender identity.

Goal of our Work

In the scope of this work, we:

- (1) identify and categorize state-of-the-art gender scores;
- (2) systematically assess their applicability and practicability; and
- (3) evaluate their applicability on German research data from clinical routine (see Section MII CDS).

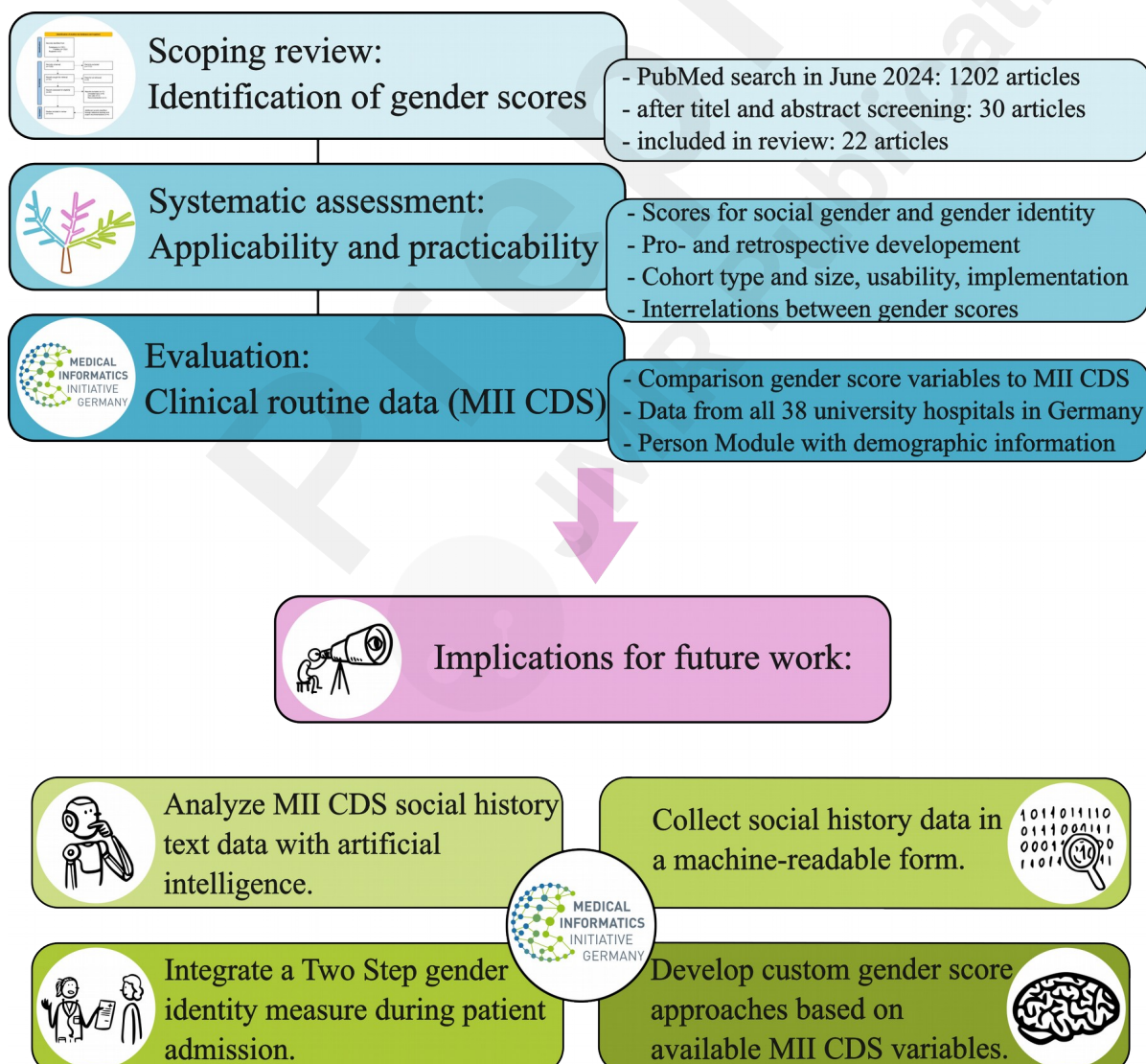
The process of our work is illustrated in Figure 1.

This work aims to contribute to the necessary implementation of a standard for collecting and assessing gender-specific data in German university hospitals and the respecting research facilities, therefore we formulated four action steps for an extension in health research to enable gender-specific analyses (see Figure 1).

Figure 1. Overview of the implemented methods. (1) Identification of recently used gender scores through a scoping review (2) Systematic assessment of the retrieved gender scores with regard to various aspects of their applicability and practicability, and (3) Evaluation of the applicability of gender scores variables on clinical routine data, specifically the MII CDS.

As implications for future work, we propose four action steps to update the MII CDS and enable gender-specific analyses. The goal is to improve gender-specific data collection in or extraction from clinical information systems, and are further described in the Discussion.

MII CDS: Medical Informatics Initiative core dataset.



Related Work

The focus of our scoping review protocol published in JMIR Review Protocols [16] was the applicability and practicability in health research. Related to the work described in this paper, Horstmann et al. [17] published a review on recently used instruments for the of sex and gender in health research. The work of Miani et al. [18] summarizes epidemiological aspects in gender scores. The reviews included publications until 2020 and 2021 respectively. For this work, we identified the latest scores and therefore deliberately limited our search to publications between 2019 and 2024. 13 of the 18 articles from our primary result set were published after 2020.

In parallel to our work on the systematic review of recent gender scores, another systematic review about the operationalization of gender via composite gender scores in epidemiological studies by Ballering et al. [19] was published in 2024. The authors searched 3 databases (PubMed, Web of Science, and CINAHL) and identified 24 articles with a total of 26 gender scores developed in Europe or North America. They aggregated information on gender scores regarding author, year of publication, methodology, cohort information, and the included variables of the respective gender score, and found that many variables overlapped across multiple studies.

Ballering et al. [19] criticize theory-driven approaches — developed solely based on expert knowledge — as experts were not free from bias and this could potentially enhance sexism or other biases. The authors further claim that data-driven approaches — using statistical models to develop gender scores — only predicted female or male sex through identified psychosocial variables, and did not consider gender itself. Psychosocial variables, that do not associate with sex are excluded and the quality of the gender scores is strongly dependent on the quality of the dataset. Data-driven approaches might be good to personalize healthcare, as gendered variables point out differences between women and men that could lead to improved treatment decisions and patient care.

The work of Ballering et al. [19] is a great foundation for our work, and we were able to contribute more insights on the applicability on clinical routine data. We had an overlap of 13 articles on both sides, with 13 out of 22 of our review and 13 out of 24 articles from the review of Ballering et al. [19], i.e. there is an overlap of over 50% (59% ours, and 54% theirs).

We found further articles regarding gender identity, that can be applied from the outset of a study or data collection, which could be of great use to implement gender information in clinical routine.

Methods

Scoping Review

We conducted a scoping review to identify articles that report on the development and implementation of gender scores in health-related studies (see Figure 1). It resulted in 22 articles, which were identified through a title and abstract screening followed by a full text evaluation. Moreover, we developed a data-charting form to extract information on article type and applicability of the gender score in health research. The review protocol was published in JMIR Research Protocols [16].

Assessment of Applicability and Practicability

To assess the applicability and practicability of gender scores we followed a systematic approach (see Figure 1). In the scope of this work we extended the data-charting process to fully assess the applicability of gender scores on clinical routine data.

Social Gender and Gender Identity

First, we distinguished between scores for gender identity (see Table 1) versus social gender (see Table 2 and 3). Gender identity (see Textbox 4) is usually measured in a prospective way, because it is a purely subjective perception and it is not possible to extract this information, when it was not collected. However, social gender includes norms that are expected in society, stereotypes and certain behaviors (see Textbox 2) and it can be measured pro- and retrospectively. Typical behavior can be analyzed and interpreted after data collection (e.g. specific hobbies, occupational segregation, doing housework or odd jobs). A typical female or male lifestyle or behavior will lead to different exposure to risk factors for accidents or disease. Social expectations will also influence the process of recovery.

Prospective and Retrospective Scores

Second, we distinguished between prospective (see Table 1+2) and retrospective (see Table 3) scores. It is thoughtful to include gender variables from the outset in a study, which results in a prospective gender score and is easy to apply to the study dataset. But the huge amount of existing data drives the development of retrospective gender scores, where gender-specific information can be abstracted in hindsight.

Cohort type and size

Third, it needs to be considered, whether gender scores themselves are representative. In practice, some gender scores are developed on restricted datasets, for instance, for specific patient groups or other samples that are not representative (e.g. older age group, workers or a specific diagnosis). This is underlined by the finding of Wandschneider et al. [9] who showed that gendered practices vary between eastern and western Germany, reflecting a different social development due to historical political context. And as gender norms, stereotypes or behaviors vary between societies, the relevant gender variables can differ across social groups, countries, regions, or even between different societies within a country.

Furthermore, since retrospective scores depend on statistical modeling we also consider the sample size the models were based on. Another analysis we performed is to further classify the background under which the score was developed. We took into account whether the research group used patient or population data to categorize which gender score might be more appropriate for the MII CDS. We also considered the continent, where the data collection took place, which is important for consistency due to varying gender-sensitive variables depending on the society or geographical location.

Usability

Forth, we examined the validity, usability, and practicability of the gender scores. Validity is the main issue because it is of high importance that the score is representative regarding gender. Usability refers to how well the score is established, because if a new standard should be introduced it is of high importance to use an accepted model in order to make the scores as usable, comparable and possible to apply in large scale studies and assessments. Practicability refers to how difficult they would be to integrate in health research.

Implementation

The fifth point we included in our assessment, is the level of implementation of scores. To apply

scores systematically on existing data a theoretical model is not sufficient. Instead, it is necessary to have the executable code of a model to integrate it directly. Therefore, we further investigated the applicability of gender scores from the technical viewpoint to evaluate whether models can, realistically, be implemented in a clinical setting. We use the following categories to describe this aspect: Code published — refers to a gender score for which an executable model was published; Data published — refers to a gender score for which the underlying data is available and a score could be reproduced; Statistical parameters published — refers to a models for which all statistical parameters were published and the model could be constructed from them; Not published — a model for which not all relevant parameters are available. We consider “Code published” the gold standard regarding applicability. “Data published” and “Parameters published” on the other hand, might contain sufficient information to construct an executable model. In the scope of this study, we consider code or data available “on request” as unavailable. However, in practice there might still be limitations if the description of the article is not concise enough to reproduce all steps.

Interrelations

Based on the extracted data we performed further analysis to better understand the currently available gender scores. We noticed early on that direct dependencies exist between gender scores. Therefore, we extracted and modeled different types of dependencies between gender scores (see Figure 2). This analysis allows us to investigate how scores were developed and whether improvements were made to existing gender scores.

One relation between gender scores, captures cases, in which an existing gender score is used directly or adapted in another score (see Figure 2, blue branch (2)). For instance, the gender score of Pelletier et al. [20] includes the Bem Sex Role Inventory (BSRI) [21]. De Breij et al. [22] adapted the score of Smith and Koehoorn [23], but changed “caring for children” to “informal caregiving” as the data was available in their dataset (see Figure 2, blue branch (2)).

Moreover, there are development relations (see Figure 2, pink branch (3)), where one gender score was developed based on the approach, for instance the statistical model, of another research group. For example, Gisinger et al. [24] with their statistical model of principal component analysis and logistic regression were inspired by the approach of Pelletier et al. [20].

Applicability on Clinical Routine Data

To assess whether gender scores are applicable on existing data and whether they can be integrated in clinical routine data, we selected a representative sample to evaluate the gender scores against (see Figure 1). The MII CDS [14] (see Background) is being collected across all 38 university hospitals in Germany. It contains interoperable, consolidated data items with a high degree of standardization in line with international standards to make collaborative data analysis practicable and efficient.

The CDS [25] consists of basic and extension modules that are subject to iterative refinement. The basic modules are generic and the extension modules are relevant for specific medical fields and disciplines such as oncology or intensive care. We used the *Person Module*, one of the early basic modules, which contains demographic information such as name, date of birth, address and administrative gender as stated in official documents. The CDS has a comprehensive description of how the data is structured and how it can be accessed, which allows us to directly compare it against gender scores. The information is available and can be compared based on a sample dataset [26]. Overall, this makes the CDS a suited choice for studying how gender scores can be applied in realistic clinical data collection.

Retrospective scores required that we compared the variables contained in the gender scores against

the variables available in clinical routine data. To perform this comparison we systematically analyzed all variables contained in the sample dataset of the MII CDS against the variables included in the statistical models of gender scores.

Results

A table with more detailed information on our results is provided in the supplementary material.

Scoping Review

Our review results are categorized into: prospective development (Table 1-2) and retrospective development (Table 3), and between gender scores (Table 2-3) and gender identity measures (Table 1).

Articles that overlap with the already published review of Ballering et al. [2024] are marked with an asterisk (*) in front of the authors' names.

Table 1. Prospective gender identity measures [12, 13, 27, 28].

gender identity measures	
Author (year), description	Questions and answer options
Lagos and Compton (2021): Two-Step gender identity measure [13]	1. What sex were you assigned at birth? – male, female, intersex 2. What is your current gender? - woman, man, transgender, a gender not listed here (space for answer)
Fraser et al. (2020): Open-ended gender measure [12]	1. What is your gender? (open-ended item)
McGuire et al. (2019): Genderqueer Identity Scale [27]	1. I am non binary, genderqueer, or an identity other than male or female. 2. I don't want to be seen in the gender binary (as either male or female) 3. I try to deliberately confuse people about whether I am male or female. 4. I try to do things that are masculine and feminine at the same time. 5. I enjoy it when people are not sure if I am male or female 6. The way I think about my gender has always been the same. 7. My gender comes naturally from within me. 8. My gender is something I have spent a lot of time figuring out. 9. The way I show my gender changes depending on who I am with. 10. The way I think about my gender has been influenced by experiences in my life. 11. The way I think about my gender will probably continue to change further as I age

	12. I have done research about gender theory and gender roles. 13. I try to convince others that society should not insist on a gender binary. 14. I try to convince others that society expects people to be too gender conforming. 15. Around me, I make sure people are free to express whatever gender roles they want. 16. The way I show my gender is important because I push society to question traditional gender roles. 17. I encourage others to be more open minded about gender and gender roles. 18. In the future, my gender expression will be traditional. 19. In the future, it will upset me if people misgender me. 20. The way I show my gender will probably be mostly the same from day to day. 21. In the future, I expect that people will rarely question my gender. 22. In the future, I think my gender will be fluid or change over time. 23. I will have a non-traditional gender role (be gender non-conforming).
Lombardi and Banik (2016): Two-Step gender identity measure [28]	1. What is your sex or gender? (Check ALL that apply) - male, female, other: please specify (space for answer) 2. What sex were you assigned at birth? (Check one) - male, female, unknown or question not asked, decline to state

Table 2. Prospective gender scores regarding social gender [20, 21, 29, 30, 31, 32, 33].
BSRI: Bem Sex Role Inventory; PAQ: Personal Attributes Questionnaire.

Prospective gender scores of social gender	
Author (year), description	Gender score variables
* Pohrt et al. (2022) [29]	1. Level of education 2. Civil status 3. Personal net income 4. Households primary earner status 5. Responsibility for caring for other people 6. Responsibility educating and upbringing people 7. Responsibility for picking up people when they become ill 8. Hours per week spent on housework 9. Being mainly responsible for housework, 10. Social support 11. Personal Attributes Questionnaire 12. Bem Sex Role Inventory: masculinity score 13. Bem Sex Role Inventory: femininity score 14. Perceived social standing
Tibubos et al.	1. Not at all emotional - very emotional

(2022): Personal Attributes Questionnaire-8 [30]	2. Not at all aware of feelings of others -very aware of feelings of others 3. Never gives up easily - gives up very easily 4. Not at all self-confident - very self-confident 5. Feels superior - feels very inferior 6. Not at all understanding of others - very understanding of others 7. Very warm in relations with others - very cold in relations with others 8. Goes to pieces under pressure - stands up well under pressure
Demuth et al. (2021) [31]	The questionnaire is not available.
Nielsen et al. (2021) [32]	1. Caregiver strain 2. Work strain 3. Independence 4. Risk-taking 5. Emotional intelligence 6. Social support 7. Discrimination
* Pelletier et al. (2015) [20]	1. Primary earner status 2. Personal income 3. Number of hours per week doing housework 4. Primary responsibility for doing housework 5. Level of stress at home 6. Bem Sex Role Inventory: masculinity score 7. Bem Sex Role Inventory: femininity score
Spence et al. (1975): Personal Attributes Questionnaire (PAQ) [33]	Male score: 1. Independent 2. Not easily influenced 3. Good at sports 4. Not excitable, minor crisis 5. Active 6. Competitive 7. Skilled in business 8. Knows ways of world 9. Adventurous 10. Outspoken 11. Interested in sex 12. Makes decisions easily 13. Not give up easily 14. Outgoing 15. Acts as leader 16. Intellectual 17. Self confident 18. Feels superior 19. Takes a stand 20. Ambitious 21. Stands up under pressure

	22. Forward 23. Not timid Female score: 1. Emotional 2. Not hide emotions 3. Considerate 4. Grateful 5. Devotes self to others 6. Tactful 7. Strong conscience 8. Gentle 9. Helpful to others 10. Kind 11. Aware, other feelings 12. Neat 13. Creative 14. Understanding 15. Warm to others 16. Likes children 17. Enjoys art and music 18. Expresses tender feelings Sex specific score: 1. Aggressive (M) 2. Dominant (M) 3. Likes math and science (M) 4. Excitable, major crisis (F) 5. Home-oriented (F) 6. Mechanical aptitude (M) 7. Needs approval (F) 8. Feelings hurt (F) 9. Cries easily (F) 10. Loud (M) 11. Religious (F) 12. Sees self running show (M) 13. Needs for security (F)
Bem (1974): Bem Sex Role Inventory (BSRI) [21]	Masculinity score 1. Acts as a leader 2. Aggressive 3. Ambitious 4. Analytical 5. Assertive 6. Athletic 7. Competitive 8. Defends own beliefs 9. Dominant 10. Forceful 11. Has leadership abilities 12. Independent

13. Individualistic
14. Makes decisions easily
15. Masculine
16. Self-reliant
17. Self-sufficient
18. Strong personality
19. Willing to take a stand
20. Willing to take risks

Femininity score:

1. Affectionate
2. Cheerful
3. Childlike
4. Compassionate
5. Does not use harsh language
6. Eager to soothe hurt feelings
7. Feminine
8. Flatterable
9. Gentle
10. Gullible
11. Loves children
12. Loyal
13. Sensitive to the needs of others
14. Shy
15. Soft spoken
16. Sympathetic
17. Tender
18. Understanding
19. Warm Yielding

Social desirability score (neutral):

1. Adaptable
2. Conceited
3. Conscientious
4. Conventional
5. Friendly
6. Happy
7. Helpful
8. Inefficient
9. Jealous
10. Likable
11. Moody
12. Reliable
13. Secretive
14. Sincere
15. Solemn
16. Tactful
17. Theatrical
18. Truthful
19. Unpredictable

20. Unsystematic

Table 3. Retrospective gender scores regarding social gender [9, 10, 11, 22, 23, 24, 34, 35, 36, 37, 38].

ICD-10-CA: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Canada

Retrospective gender scores of social gender	
Author (year)	Gender score variables
* Cipriani et al. (2024) [34]	1. Hostile behavior during childhood 2. Hostile behavior during adulthood 3. Having at least a secondary school diploma 4. Having at least a bachelor's degree 5. Sleep satisfaction 6. Sleep efficacy 7. Having private housing 8. Sexual violence during childhood
* Gisinger et al. (2023) [24]	1. Level of education 2. Household income 3. Household size 4. Civil status 5. Perceived life stress 6. Sense of belonging to community
* Teterina et al. (2023): ICD-10-CA diagnostic codes with highest effects for predicting male/female gender [35]	Male score: 1. Fall on and from scaffolding 2. Occupant of heavy transport vehicle injured in noncollision transport accident, driver, traffic accident 3. Driver of other all-terrain or other off road motor vehicle injured in traffic accident 4. Motorcycle rider injured in collision with fixed or stationary object, driver, traffic accident 5. Asymptomatic human immunodeficiency virus (HIV) infection status 6. Motorcycle rider injured in collision with car, pick-up truck or van, driver, traffic accident 7. Motorcycle rider injured in noncollision transport accident, driver, nontraffic accident 8. Motorcycle rider injured in unspecified nontraffic accident 9. Fracture of malar and maxillary bones, LeFort 2, closed 10. Contact with nonpowered hand tool Female score: 1. Assault by spouse or partner 2. Aesthetic sports 3. Physical abuse 4. Horse riding sports 5. Animal-rider or occupant of animal-drawn vehicle injured by fall from or being thrown from animal or animal-drawn

	vehicle in noncollision accident 6. Fall while being carried or supported by other persons 7. Other specified gymnastic and aesthetic sports and recreational activity 8. Bitten or struck by dog 9. Animal-rider or occupant of animal-drawn vehicle injured in other and unspecified transport accidents 10. Problems in relationship with spouse or partner
* Vader et al. (2023): masculine gender score [36]	1. Work and education (compared with partner) 2. Informal care: household chores, odd jobs, taking care of sick people 3. Lifestyle: physical intensity, type of sport, smoking, alcohol 4. Emotions: emotional problems, nervousness, energetic and vibrant, exhausted and tired
* De Breij et al. (2022) [22]	1. Working hours 2. Income 3. Occupation segregation 4. Education 5. Informal caregiving 6. Time spent on household chores
* Wandschneider et al. (2022) [9]	1. A person who is living with their partner for the long term should get married. 2. Children below the age of 6 suffer if their mother works. 3. A same-sex couple can raise a child just as well as a man and woman. 4. It would be good for society if transgender people were recognized as normal. 5. Working experience part-time employment 6. Hours/weekday housework 7. Hours/weekday repairs 8. Hours/weekday leisure, hobbies 9. Worried about global terrorism 10. Worried about crime in Germany 11. Satisfaction with housework 12. Willingness to take risks 13. Worried about own retirement pension
* Nauman et al. (2021) [11]	1. Chronic stress 2. Civil status 3. Risk-taking behavior 4. Agreeableness 5. Neuroticism 6. Extraversion 7. Conscientiousness 8. Loneliness 9. Level of education
* Yuan et al. (2021) [37]	1. Risk willingness 2. Loneliness 3. Less participation in household tasks

	4. Regular drinking 5. Depression
* Ballering et al. (2020) [38]	1. Leisure activities (e.g. type of sport) 2. Occupation-related components (e.g. profession, housewife/-husband) 3. Time spend on household tasks 4. Time spend on odd jobs 5. Lifestyle (e.g. dieting, preparing dinner, alcohol, smoking) 6. Experiencing long-term difficulties or negative life events 7. Personality traits and emotions (e.g. discipline, impulsiveness, self-consciousness, vulnerability, competence)
* Lacasse et al. (2020) [10]	1. Civil status 2. Racial/ cultural group 3. Level of education 4. Household income, size, and composition 5. Household food insecurity 6. Ownership of the household 7. Sense of belonging to the local community 8. Occupation: type, working hours, self-employment 9. Amount of stress
* Smith and Koehoorn (2016) [23]	1. Responsibility for caring for children 2. Occupational segregation 3. Hours of work relative to partner/spouse 4. Education relative to partner/spouse.

Assessment of Applicability and Practicability

Social Gender, Gender Identity, and Pro- and Retrospective Scores

We distinguish between articles targeting gender identity and articles targeting social gender. We further classify them into pro- and retrospective development and application.

The four identified gender identity measures (see Table 1) are prospectively applicable and could be included in future studies for further data collection regarding gender identity. Participants need to be asked explicitly which gender they belong to, this information cannot be analyzed from existing data. Moreover, we found 18 gender scores to measure social gender in different ways and through varying approaches. Seven of them are prospective gender scores (see Table 2) that could provide guidance on how to apply (social) gender to a study from the outset. 11 of them are retrospective (see Table 3) and could be applied on already existing data.

Cohort type and size

The background of the cohorts is of importance for the practicability. A score contains different variables, whether it was developed on data from population studies or on patient data. The majority of datasets are based on population studies and thus on specific information that are typically collected in theses studies.

Only two gender scores were developed exclusively based on patient data. Teterina et al. [35] focused on patients that entered the emergency department with traumatic brain injuries and published an innovative approach to distinguishing gender by diagnostic codes. These diagnostic codes also describe the background of the accident or incident and were therefore able to draw conclusions about gender. Cipriani et al. [34] focused on the effects of gender on emergency patients

with psychological crisis. Therefore, they used variables assessing childhood trauma, education, sleep quality and living situation.

Moreover, one article by McGuire et al. [27] assesses gender identity during transition through the collection of patient and population data.

However, all of these studies are restricted by the specific patient cohort group and are focused on specific conditions. This hampers their applicability on the general public and their usefulness as broadly applied gender scores for clinical routine data.

Beside that, the population cohorts are neither comprehensive nor inclusive. Many studies used a cohort with a specific age group [11, 20, 29, 31, 36, 37], or with a specific background like being retired [29] or being a worker [10, 22]. This is a major weakness for generalizing results, as a gender score based on a limited cohort is not representative of society as a whole and therefore lacks comprehensiveness.

Furthermore, the majority of cohorts used for developing gender scores are based in the same geographic region, which are Europe and North America. Only one gender identity measure was used on a cohort in Oceania [12]. We could not find a single gender score publication with a cohort from Asia, South America or Africa. As a result, the developed scores are not representative for every cohort group and are not universally applicable.

With respect to the size of cohorts, which forms the basis for the statistical development of retrospective gender scores and influences the reliability of statistical models, we noted that some works used huge cohorts or datasets with $n > 100,000$ [24, 35, 38] and the outstanding cohort size of Smith and Koehoorn [23] with around $n = 700,000$. Other gender scores worked with smaller cohort sizes of $n < 1,000$ [21, 22, 33].

Usability

We examined the validity of scores, meaning how representative the work is regarding gender. Gender can be different in each society, so it is a major task to implement a gold standard. Some research groups used sex as dependent variable and therefore, gender as predictor for sex [a.o. 9, 10, 11, 20]. This approach can be problematic since gender should be considered beyond biological sex.

Statistical regression approaches can be systematically validated by using a control set to test the generalization of the regression model, e.g. by systematic evaluation on a dedicated validation set [20, 34, 35, 38]. A number of scores following the approach of Pelletier et al. [20] do not specify whether they performed a systematic validation of their developed scores [9, 24, 29]. The work of Nauman et al. [11] does not employ a validation and the work of Yuan et al. [37] reports a performance value without describing the used validation setting.

Exclusively theory driven gender scores, where variable scaling is determined by human experts were not evaluated since there is no gold standard for gender. [22, 23, 36].

Looking at usability, that is referring to how well established a score already is, some scores did not mention any practical application [31], whereas others are often cited, adapted, and reapplied [20, 21, 23, 33].

While the latter seem to be widely accepted and used, they might still not be well suited to represent gender. The most cited BSRI, which is also part of other gender scores [e.g. 20, 29] (see Figure 2), is often criticized [15].

Regarding practicability we considered the scope and complexity of questionnaires and how universally applicable variables are on existing data. Some variables are very specific for the used

cohort, whereas others frequently appear in similar studies.

Most gender identity measures are well-suited for clinical data collection, being concise and easy to understand, except for McGuire's score [27], which involves 23 questions and focuses on gender transition, making it less suitable.

BSRI [21] and PAQ [33] are lengthy prospective gender scores, whereas other scores containing them are extended as well. Tibubos [30] developed a shorter form of the PAQ with only 8 variables.

We also listed different forms of data collection, from self-reporting questionnaires [e.g. 20, 29] to interviews [e.g. 13, 28, 37].

The length of forms could be especially relevant, if participants do a self-reporting questionnaire. Interviews tend to be more time-consuming for the conducting staff.

With retrospective scores, the length might not have such a big impact. But some scores stand out and are much shorter [23, 37] than the majority. A minimalistic score can be a trade-off between prediction quality and the number of necessary variables. This was, for instance, investigated by Ballering et al. [38], who compared the performance between gender scores including 9 and 85 predictive variables.

Implementation

Only three research groups made their analysis code publicly available (*Code published*) and provided reproducible gender scores [9, 32, 38].

Two studies published their data [32] or used publicly available data [37] (*Data published*), that is necessary to evaluate their conclusion. All other works did not publish data, which might be justified with data protection regulations. However, anonymized data could still be helpful for recreating a study.

Statistical parameters were published by most retrieved papers (*Statistical parameters published*), some more complete than others.

However, in the majority of cases, our assessment showed that reconstructing a gender score from provided parameters would be challenging.

We identified one paper that did not publish the gender score (*Not published*) [31]. They adapted the scores of Nauman et al. [11] and Pelletier et al. [20] to form a new gender score, but they do not provide any specific information regarding their methodology.

Overall, repeatability is made difficult by many research groups and the usage of existing scores is made complicated.

Interrelations

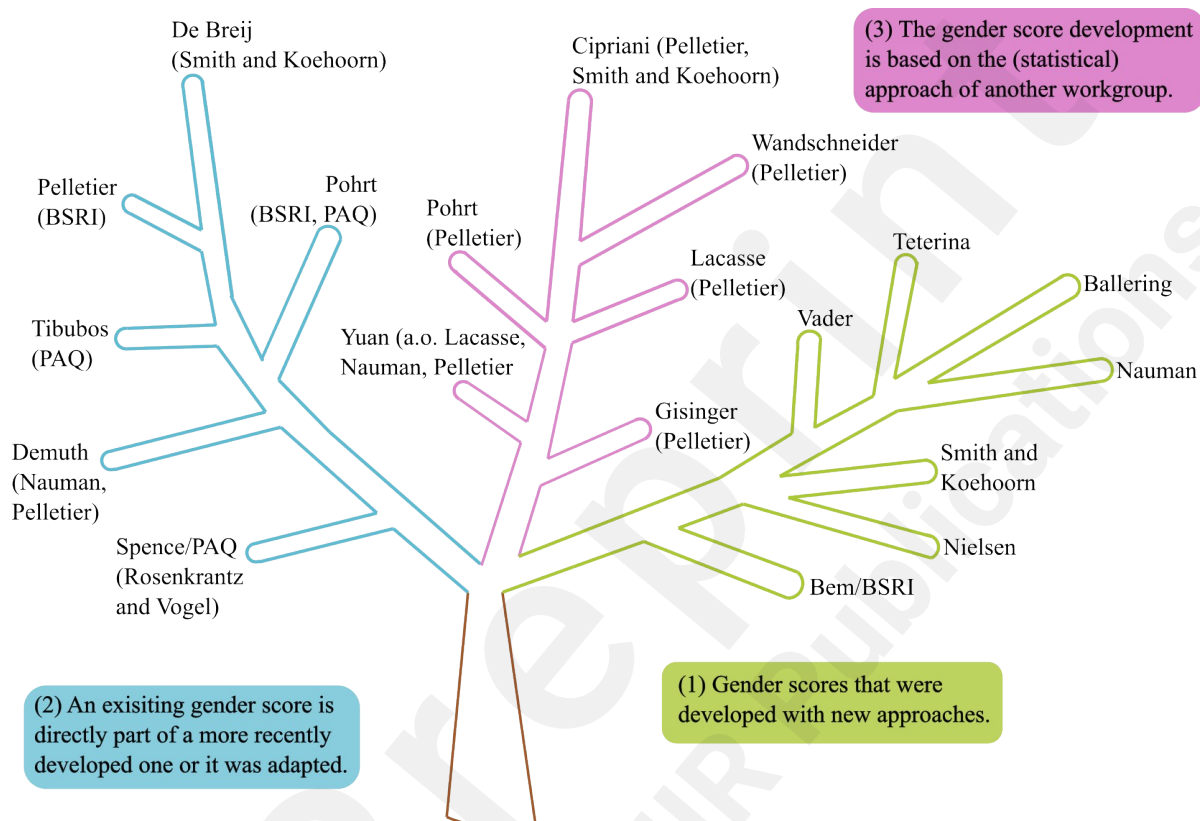
We registered different dependencies and relations between gender scores and visualized them (see Figure 2).

Seven of our retrieved articles were developed with totally new approaches or methods (see Figure 2, (1) *green branch*).

Other scores adapted or included content-related parts of already existing scores (see Figure 2, (2) *blue branch*). For example, Pelletier et al. [20] included the BSRI masculinity and femininity score [21]. Spence et al. [33] were inspired by Rosenkrantz and Vogel [39] for the development of the PAQ in 1975 and thereupon Tibubos [30] developed a short-form of the PAQ in 2022. It is notable that the BSRI [21] and PAQ [33] each built the basis of two other scores.

Sometimes research groups implement the statistical or methodological approach of another team (see Figure 2, (3) *pink branch*). For instance, the statistical approach of Pelletier et al. [20] was frequently used as basis for current gender scores, i.e. by Wandschneider, Lacasse, Gisinger, Pohrt, Cipriani, and Yuan [9, 10, 24, 29, 34, 37].

Figure 2. Interrelations between gender scores. The first one or two author names identify the work and the developed gender score (see Table 2 and 3). The green branch (1) shows gender scores that were developed with completely new approaches and methods. We used two abbreviations of gender scores, because they are widely known under these names. BSRI: Bem Sex Role Inventory. PAQ: Personal Attributes Questionnaire. In blue (2) existing gender scores are displayed with the included or adapted gender scores in brackets. In pink (3) gender scores are displayed, that were inspired in their development by previous methods and (statistical) approaches (here in brackets).



Applicability on Clinical Routine Data

We tested all retrieved scores and their published variables for whether they are available in the MII CDS. Initially, we planned to only test retrospective scores, developed on clinical data. However, we found that this would limit the number of possible scores too much.

The evaluation of the applicability on clinical routine data required a careful and considerate approach as variable names are often ambiguous and it is usually not possible to directly match them. This can, for instance, be seen in variables related to care work, which include “responsibility for caring for other people in the household” [29], “dinner is always prepared by someone else” [38], “caregiver strain” [32], “taking care of sick people” [36], “informal caregiving” [22], “responsibility for caring for children” [23]. Even though, some gender scores might be well described and could be used for other datasets, it showed to be complex to transfer variables appropriately to another dataset.

Applicability of Gender Scores on the MII CDS

In the comparison between the variables contained in gender scores and within the MII CDS we

employed a two-step approach: we first tried to match the exact gender score variables against the MII CDS; and, in case this remained unsuccessful, we extended the search to similar or generic terms.

We found that no variable of any gender score is covered in the MII CDS. This finding is surprising since a high number of variables covered in gender scores, such as occupation, level of education, civil status, children, household composition and size, and social support, are part of social history which is clinical routine information.

Therefore, our principle finding regarding this question is that no gender score can be applied upon the MII CDS. One exception is the score of Teterina et al. [35], which includes diagnostic codes and disease specific data, which could be matched against the MII CDS. This score, however, is highly specific, and not usable as a generalized approach.

Currently, the MII CDS only encodes the administrative gender as male, female, undifferentiated, or diverse [26]. Administrative gender refers to the gender that is recorded or recognized in official document, e.g. the ID card. It may differ from the BAS.

Overall, most scores are applied on population studies or included in studies from the outset. These gender scores frequently include personality traits, and psychological assessments (e.g. stress, loneliness scales) which are not part of clinical routine and, therefore, are also not part of the MII CDS.

To summarize, there are no fitting variables for the MII CDS, and it is impossible to apply an existing gender score to the dataset in order to include gender-specific analysis on clinical routine data.

Discussion

Principal Findings

This work identifies and reviews state-of-the-art gender scores and gender identity measures. These were systematically assessed regarding their applicability and practicability in health research. Lastly, we tested their applicability on clinical routine data, using the MII CDS as a reference dataset.

We found that no gender score is applicable on the MII CDS, because the variables required for the gender scores are not part of the CDS.

However, many variables that are commonly included in gender scores are also assessed during a patient's clinical stay. For instance, variables regarding the personal background (e.g. are you living with your partner or children, which professional field or self-employment) are often asked during a clinic stay to verify if specific therapies or treatments are realizable. Physicians need to know if patients have responsibilities at home, if they have support, and if they have a stable social environment at home. Therefore, a person's background is gender-dependent and does not only influence clinical routine, but also research data and therapy guidelines (e.g. risk factors depend on the lifestyle).

We argue that the current lack of applicability of gender scores on clinical routine data is mainly an issue resulting from missing data collection in or extraction from clinical information systems. Since the MII CDS is a state-of-the-art effort to standardize research data from clinical routines, we recommend that the Person Module of the MII CDS should be updated and expanded, to create a sufficient knowledge base for gender health research. Specifically, we formulated four action steps, that enable the extension of the MII CDS Person Module and allow gender-specific analyses (see

Figure 1):

- Analyze MII CDS social history text data with artificial intelligence.
- Collect social history data in a machine-readable form.
- Integrate a Two Step gender identity measure during patient admission.
- Develop custom gender scores approaches based on available MII CDS variables.

As mentioned above, social history is part of the clinical routine. However, the social history data is not usable in the MII CDS. One solution could be to analyze the social history full-text data with artificial intelligence, such as Large Language Models, to collate and make the data concise and usable for research. Another approach could be to already collect social history data, such as occupation, education, civil status, children, household composition and size, social support, in a machine-readable form. Instead of writing or dictating the social history as full-text for the physician's report, individual fields for each variable to click on or fill in, could be established in the clinical IT systems. This would require less effort for processing and collating the data for further usage.

Moreover, to include as many gender dimensions as possible it would be beneficial to collect more gender-specific data. A promising approach would be to implement the Two-Step gender identity measure [13] with open-ended answers during patient admission. Implementing a prospective approach, with minimal effort, would provide valuable insights into gender identity in everyday clinical practice

Additionally, researchers should think beyond the common gender score approaches and develop a gender score that is suitable to clinical data and not only population studies.

An innovative approach was carried out by Teterina et al. [35], which used diagnostic codes of patients with traumatic brain injuries in Canadian emergency departments.

Researchers should try similar approaches on general— non-disease-specific — clinical data with diagnostic or medication variables.

Furthermore, developing a minimalistic retrospective score could prove advantageous in balancing prediction quality with the number of required variables, making the scores applicable even when limited information is available, e.g.[38].

Even though gender is gaining importance in health research, it is still far from being implemented and our investigation shows that a gold standard is missing that accounts for variations across social groups, e.g. including geographical regions, age, time.

Due to the lack of a gold standard, some research groups use gender as a predictor for sex when developing gender scores, which contradicts the purpose of introducing gender in research — to move beyond biological sex and gather more detailed information.

Solely theory driven approaches, where experts determine the used variables and their scaling, cannot be validated since there is no gold standard. However, the missing statistical component makes those scores easier to apply.

Lastly, repeatability is a major challenge when aiming at implementing existing gender scores. Due to inconsistent research group practices, the existing gender scores are often described only theoretically without providing detailed variables, questions, or complete information on variable loadings. We therefore urge authors to provide complete documentation and executable models when developing gender scores.

To achieve the necessary progress, every inclusion of gender into health research is beneficial, even

if it might not be comprehensive or sufficient to cover all aspects. Regarding the future implementation of gender analysis, we suggest four steps (see Figure 1) for the systematic standardization of gender information.

Limitations

Our collection of gender scores is based on a scoping review, conducted to provide a broad overview of the topic. One limitation of this approach is that only one reviewer carried out the whole-text screening, and the categorization of the retrieved articles. However, during data extraction more than 30% (7/22) of the publications were initially double-checked by a second domain expert [16]. The reviewers were in full agreement on all reviewed articles, thus a single reviewer continued the data extraction. Moreover, our review does not claim to provide a complete list of all existing gender scores, but does provide general statements on the current state of gender scores and their applicability in health research.

Moreover, the MII CDS might not be fully representative to test against population-based gender scores. Most gender scores were developed on population or specific patient data, while the MII CDS contains clinical routine data. However, it is important to not only include gender in population, but also in clinical research. This is why we selected the MII CDS as a representative dataset for clinical routine data to test the practicability in this setting. As a result, we were able to prove that gender scores are not yet a feasible instrument for clinical routine data and are able to make recommendations on how to update and extend the MII CDS in the future.

Comparison with Prior Work

Ballering et al. [19] highlight, that gender scores should be the minimum effort in epidemiological studies, but the community should go further than that. As gender scores imply a lack of gender consideration from the outset of the study, e.g. the study design. Our findings found several shortcomings with currently available retrospective gender scores. Therefore, our results suggest that self-reported gender assessments are more precise and better suited to assess gender, confirming the theory of Ballering et al. [19].

Moreover, we can confirm the finding of Ballering et al. [19], that the most common variables in gender scores are related to occupation, income, education, civil status, care-giving and household responsibilities, and ways of spending (leisure) time. Furthermore, our results showed that even though several of these variables are collected in everyday clinic work, but they are not processed and stored in machine-readable way.

Conclusions

Considering sex and gender enhances equity and research quality. Despite the high importance of the topic it remains challenging to include gender into health research. To contribute to the implementation of a standard for collecting and analyzing gender-specific data we identified, categorized and systematically assessed state-of-the-art gender scores. We evaluated their applicability and practicability on German research data from clinical routine, specifically the MII CDS. We do not claim to provide a complete list of all existing gender scores but provide insight on the current state of gender scores and their applicability in health research.

Our major finding is, despite different approaches of assessing gender, no standardized and validated

gender score exists, that could be used retrospectively in health research. Moreover, the MII CDS does not contain necessary variables for gender scores, even though several of the frequently used variables are part of clinical routine in Germany. This shows, that it remains challenging to implement a new standard. Therefore, we see an urgent need in including gender-relevant information to the MII CDS, to enable gender-specific analysis of clinical routine data. Furthermore, more gender-specific data should be routinely collected in clinics, for example, during patient admission.

Acknowledgments

This work was funded by the German Federal Ministry of Education and Research as part of the project “Inclusive Excellence in Medicine” (01FP23G10A, 01FP23G10B) and the “Network University Medicine” (01KX2121).

The authors declare that they have no competing interests. We did not collect any personal data, so ethical approval was not needed.

Author contributions:

LS did the conceptualization, methodology, formal analysis, investigation, data curation, visualization, and writing of the original draft

HB conducted the validation, and review and editing of the writing.

SS and EK conducted the conceptualization, methodology, validation, supervision, funding acquisition, review and editing of the writing.

DW carried out the conceptualization, methodology, validation, supervision, funding acquisition, writing, reviewing and editing of the original draft.

Conflicts of Interest

None declared.

Abbreviations

BSRI: Bem Sex Role Inventory

MI: Medical Informatics Initiative

CDS: core dataset

PAQ: Personal Attributes Questionnaire

Data Availability

Further data underlying our publication are available as supplementary material.

References

1. Tannenbaum C, Ellis RP, Eyssel F, Zou J, Schiebinger L. Sex and gender analysis improves science and engineering. *Nature*. 2019;575(7781):137-146. doi:10.1038/s41586-019-1657-6
2. Arcand M, Bilodeau-Houle A, Juster RP, Marin MF. Sex and gender role differences on stress, depression, and anxiety symptoms in response to the COVID-19 pandemic over time. *Front Psychol*. 2023;14. doi:10.3389/fpsyg.2023.1166154
3. German Ethics Council. In: Opinion: Intersexuality. German Ethics Council; 2012:26-32. Accessed January 19, 2024.

<https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/englisch/opinion-intersexuality.pdf>

4. WHO. Gender and health. 2024. Accessed September 13, 2024. <https://www.who.int/health-topics/gender>
5. Oertelt-Prigione S. Chapter 32 - The operationalization of gender in medicine. In: Legato MJ, ed. *Principles of Gender-Specific Medicine (Fourth Edition)*. Academic Press; 2023:503-512. doi:10.1016/B978-0-323-88534-8.00017-1
6. Geller SE, Koch AR, Roesch P, Filut A, Hallgren E, Carnes M. The More Things Change, the More They Stay the Same: A Study to Evaluate Compliance With Inclusion and Assessment of Women and Minorities in Randomized Controlled Trials. *Acad Med*. 2018;93(4):630-635. doi:10.1097/ACM.0000000000002027
7. Potluri T, Engle K, Fink AL, vom Steeg LG, Klein SL. Sex Reporting in Preclinical Microbiological and Immunological Research. *mBio*. 2017;8(6):e01868-17. doi:10.1128/mBio.01868-17
8. Shah K, McCormack CE, Bradbury NA. Do you know the sex of your cells? *Am J Physiol Cell Physiol*. 2014;306(1):C3-C18. doi:10.1152/ajpcell.00281.2013
9. Wandschneider L, Sauzet O, Razum O, Miani C. Development of a gender score in a representative German population sample and its association with diverse social positions. *Front Epidemiol*. 2022;2. doi:10.3389/fepid.2022.914819
10. Lacasse A, Pagé MG, Choinière M, et al. Conducting gender-based analysis of existing databases when self-reported gender data are unavailable: the GENDER Index in a working population. *Can J Public Health*. 2020;111(2):155-168. doi:10.17269/s41997-019-00277-2
11. Nauman AT, Behlouli H, Alexander N, et al. Gender score development in the Berlin Aging Study II: a retrospective approach. *Biology of Sex Differences*. 2021;12(1):15. doi:10.1186/s13293-020-00351-2
12. Fraser G, Bulbulia J, Greaves LM, Wilson MS, Sibley CG. Coding Responses to an Open-ended Gender Measure in a New Zealand National Sample. *J Sex Res*. 2020;57(8):979-986. doi:10.1080/00224499.2019.1687640
13. Lagos D, Compton D. Evaluating the Use of a Two-Step Gender Identity Measure in the 2018 General Social Survey. *Demography*. 2021;58(2):763-772. doi:10.1215/00703370-8976151
14. Ammon D, Kurscheidt M, Buckow K, et al. Arbeitsgruppe Interoperabilität: Kerndatensatz und Informationssysteme für Integration und Austausch von Daten in der Medizininformatik-Initiative. *Bundesgesundheitsbl*. 2024;67(6):656-667. doi:10.1007/s00103-024-03888-4
15. Donnelly K, Twenge JM. Masculine and Feminine Traits on the Bem Sex-Role Inventory, 1993–2012: a Cross-Temporal Meta-Analysis. *Sex Roles*. 2016;76(9):556-565. doi:10.1007/s11199-016-0625-y
16. Schindler L, Beelich H, Röhl S, Katsari E, Stracke S, Waltemath D. Applicability of

Retrospective and Prospective Gender Scores for Clinical and Health Data: Protocol for a Scoping Review. *JMIR Research Protocols*. 2025;14(1):e57669. doi:10.2196/57669

17. Horstmann S, Schmechel C, Palm K, Oertelt-Prigione S, Bolte G. The Operationalisation of Sex and Gender in Quantitative Health-Related Research: A Scoping Review. *International Journal of Environmental Research and Public Health*. 2022;19(12):7493. doi:10.3390/ijerph19127493

18. Miani C, Wandschneider L, Niemann J, Batram-Zantvoort S, Razum O. Measurement of gender as a social determinant of health in epidemiology—A scoping review. *PLOS ONE*. 2021;16(11):e0259223. doi:10.1371/journal.pone.0259223

19. Ballering AV, Hartman TCO, Rosmalen JGM. Gender scores in epidemiological research: methods, advantages and implications. *The Lancet Regional Health – Europe*. 2024;43. doi:10.1016/j.lanepe.2024.100962

20. Pelletier R, Ditto B, Pilote L. A Composite Measure of Gender and Its Association With Risk Factors in Patients With Premature Acute Coronary Syndrome. *Psychosomatic Medicine*. 2015;77(5):517. doi:10.1097/PSY.0000000000000186

21. Bem SL. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*. 1974;42(2):155-162. doi:10.1037/h0036215

22. De Breij S, Huisman M, Boot CRL, Deeg DJH. Sex and gender differences in depressive symptoms in older workers: the role of working conditions. *BMC Public Health*. 2022;22:1023. doi:10.1186/s12889-022-13416-1

23. Smith PM, Koehoorn M. Measuring gender when you don't have a gender measure: constructing a gender index using survey data. *International Journal for Equity in Health*. 2016;15(1):82. doi:10.1186/s12939-016-0370-4

24. Gisinger T, Azizi Z, Alipour P, et al. Sex and gender aspects in diabetes mellitus: Focus on access to health care and cardiovascular outcomes. *Front Public Health*. 2023;11:1090541. doi:10.3389/fpubh.2023.1090541

25. The Medical Informatics Initiative's core data set | Medical Informatics Initiative. Accessed March 3, 2025. <https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set>

26. MII Core Data Set - Datasets. Accessed March 3, 2025. <https://art-decor.org/art-decor/decor-datasets--mide-#>

27. McGuire JK, Beek TF, Catalpa JM, Steensma TD. The Genderqueer Identity (GQI) Scale: Measurement and validation of four distinct subscales with trans and LGBTQ clinical and community samples in two countries. *International Journal of Transgenderism*. 2019;20(2-3):289-304. doi:10.1080/15532739.2018.1460735

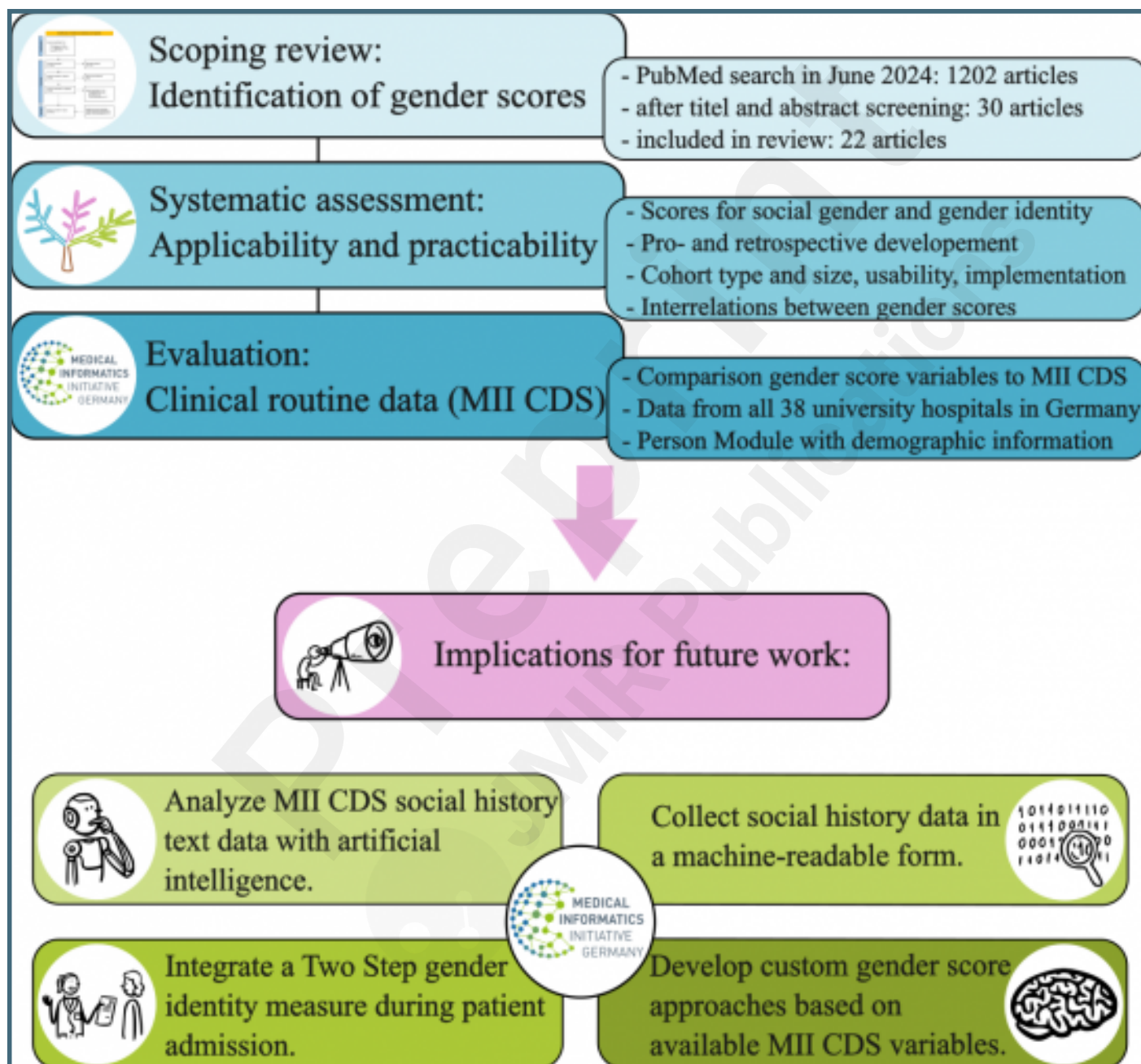
28. Lombardi E, Banik S. The Utility of the Two-Step Gender Measure Within Trans and Cis

- Populations. *Sex Res Social Policy*. 2016;13(3):288-296. doi:10.1007/s13178-016-0220-6
29. Pohrt A, Kendel F, Demuth I, et al. Differentiating Sex and Gender Among Older Men and Women. *Psychosomatic Medicine*. 2022;84(3):339. doi:10.1097/PSY.0000000000001056
30. Tibubos AN, Otten D, Beutel ME, Brähler E. Validation of the Personal Attributes Questionnaire-8: Gender Expression and Mental Distress in the German Population in 2006 and 2018. *Int J Public Health*. 2022;67:1604510. doi:10.3389/ijph.2022.1604510
31. Demuth I, Banszerus V, Drewelies J, et al. Cohort profile: follow-up of a Berlin Aging Study II (BASE-II) subsample as part of the GendAge study. *BMJ Open*. 2021;11(6):e045576. doi:10.1136/bmjopen-2020-045576
32. Nielsen MW, Stefanick ML, Peragine D, et al. Gender-related variables for health research. *Biology of Sex Differences*. 2021;12(1):23. doi:10.1186/s13293-021-00366-3
33. Spence JT, Helmreich R, Stapp J. Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology*. 1975;32(1):29-39. doi:10.1037/h0076857
34. Cipriani E, Samson-Daoust E, Giguère CE, Kerr P, Lepage C, Juster RP. A step-by-step and data-driven guide to index gender in psychiatry. *PLoS One*. 2024;19(1):e0296880. doi:10.1371/journal.pone.0296880
35. Teterina A, Zulbayar S, Mollayeva T, Chan V, Colantonio A, Escobar M. Gender versus sex in predicting outcomes of traumatic brain injury: a cohort study utilizing large administrative databases. *Sci Rep*. 2023;13(1):18453. doi:10.1038/s41598-023-45683-2
36. Vader SS, Lewis SM, Verdonk P, Verschuren WMM, Picavet HSJ. Masculine gender affects sex differences in the prevalence of chronic health problems - The Doetinchem Cohort Study. *Prev Med Rep*. 2023;33:102202. doi:10.1016/j.pmedr.2023.102202
37. Yuan J, Sang S, Pham J, Kong WJ. Gender Modifies the Association of Cognition With Age-Related Hearing Impairment in the Health and Retirement Study. *Front Public Health*. 2021;9:751828. doi:10.3389/fpubh.2021.751828
38. Ballering AV, Bonvanie IJ, Olde Hartman TC, Monden R, Rosmalen JGM. Gender and sex independently associate with common somatic symptoms and lifetime prevalence of chronic disease. *Social Science & Medicine*. 2020;253:112968. doi:10.1016/j.socscimed.2020.112968
39. Rosenkrantz P, Bee H, Vogel S, Broverman I. Sex-role stereotypes and self-concepts in college students. *J Consult Clin Psychol*. 1968;32(3):287-295. doi:10.1037/h0025909

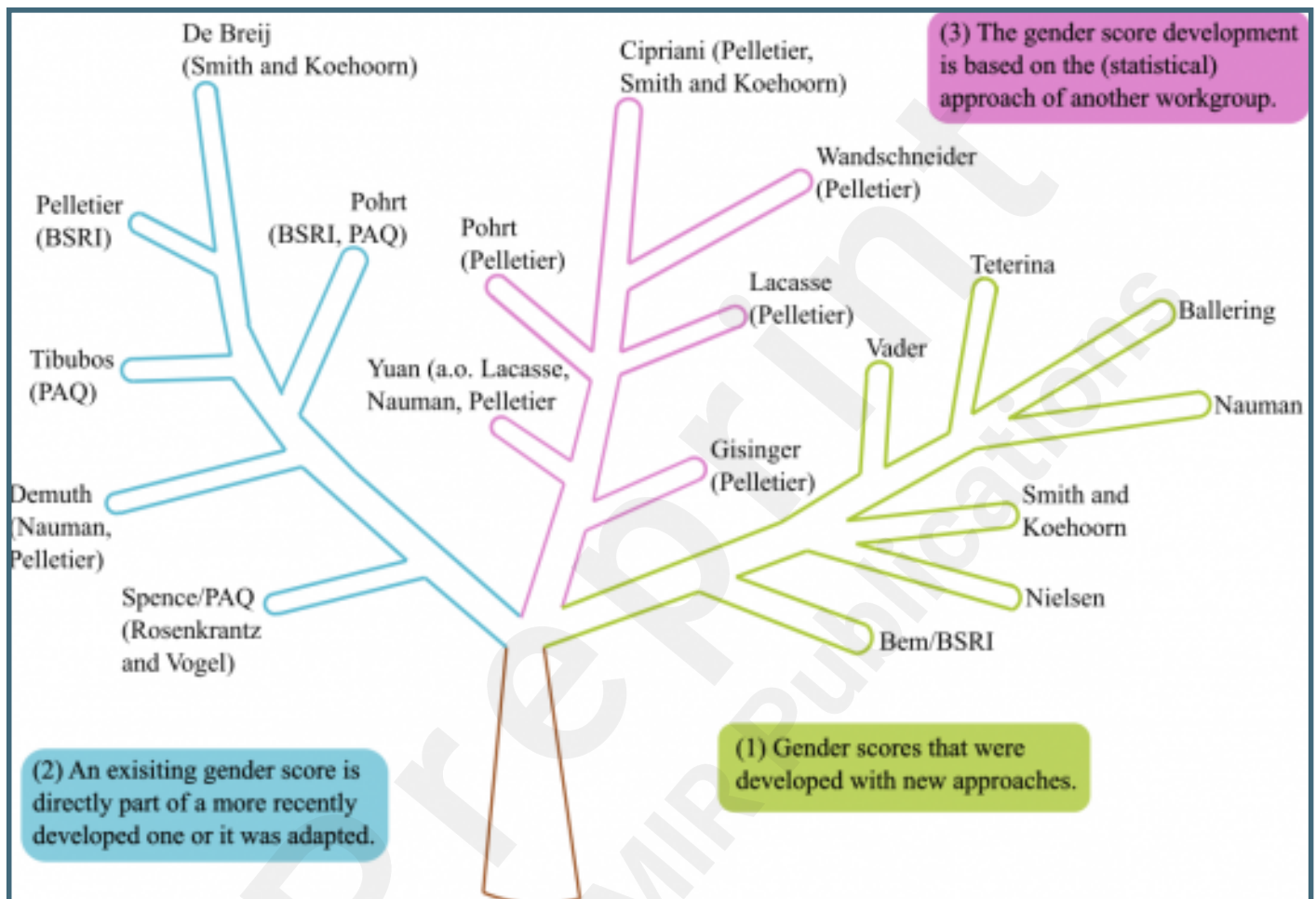
Supplementary Files

Figures

Overview of the implemented methods. (1) Identification of recently used gender scores through a scoping review (2) Systematic assessment of the retrieved gender scores with regard to various aspects of their applicability and practicability, and (3) Evaluation of the applicability of gender scores variables on clinical routine data, specifically the MII CDS. As implications for future work, we propose four action steps to update the MII CDS and enable gender-specific analyses. The goal is to improve gender-specific data collection in or extraction from clinical information systems, and are further described in the Discussion. MII CDS: Medical Informatics Initiative core dataset.



Interrelations between gender scores. The first one or two author names identify the work and the developed gender score (see Table 2 and 3). The green branch (1) shows gender scores that were developed with completely new approaches and methods. We used two abbreviations of gender scores, because they are widely known under these names. BSRI: Bem Sex Role Inventory. PAQ: Personal Attributes Questionnaire. In blue (2) existing gender scores are displayed with the included or adapted gender scores in brackets. In pink (3) gender scores are displayed, that were inspired in their development by previous methods and (statistical) approaches (here in brackets).



Multimedia Appendixes

Table providing detailed information from the assessment of the retrieved gender scores.

URL: <http://asset.jmir.pub/assets/ddb33b258d2138fd0effdee86bb8aafe.xlsx>

