



UNIVERSITÀ
DI PAVIA

FACOLTA' DI INGEGNERIA
DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

CORSO DI LAUREA MAGISTRALE IN BIOINGEGNERIA

TESI DI LAUREA

VERIFICHE AUTOMATICHE DI QUALITÀ DEI DATI PER UNO STUDIO CLINICO
MULTICENTRICO IN RIABILITAZIONE

Candidato Cigliutti Elisa

Relatore: Prof. Quaglini Silvana
Correlatore: Prof. Nicora Giovanna

A.A. 2025/2026

Indice

Indice	1
Capitolo 1 - Introduzione	4
Capitolo 2 - Controllo di qualità del dato: definizione e background	6
2.1 Background	6
2.1.1 La provenienza dei dati	7
2.1.2 Dati mancanti	8
2.1.3 La raccolta dati digitale	9
2.1.4 Strumenti di data quality assurance	11
2.2 Le dimensioni della qualità del dato	12
2.1.1 Completeness	13
2.1.2 Correctness	13
2.1.3 Concordance	14
2.1.4 Plausibility	14
2.1.5 Currency	15
2.1.6 Conformance	15
2.1.7 Bias	15
2.1.8 Tecniche di controllo della qualità dei dati	15
Capitolo 3 - Il caso di studio: il progetto Fit4MedRob e il trial clinico	
StrokeFit4	17
3.1 Il protocollo di StrokeFit4	19
3.1.1 Endpoint dello studio	19
3.1.2 Caratteristiche	20
3.1.3 - Valutazione delle ADL con mBI	23
3.1.4 - Disegno di trattamento e dati raccolti	24

Capitolo 4 - Strumenti di raccolta dati in StrokeFit4	27
4.1 REDCap	27
4.1.1 I controlli di qualità nativamente implementati in REDCap	32
4.1.1.1 Flag Required/Identifier	32
4.1.1.2 Field Validation	33
4.1.1.3 Branching logic	34
4.1.1.4 Action tags	34
4.1.1.6 Scheduling mode e modalità di alert	35
4.1.2 Controlli di qualità implementati ad hoc per lo studio: accorgimenti progettuali	35
4.1.2.1 Uso di calculated fields per il calcolo degli score	35
4.1.2.2 Gli eventi Status, Timings e Treatments	36
4.1.2.3 Campo “scala eseguita”	37
4.1.3 Monitoraggio dell’inserimento dati: sezione di data export, report e stats	37
4.2 Plat4MedRob	38
Capitolo 5 - La dashboard per la Data Quality	41
5.1 Controlli automatizzati	44
5.1.1 Missing values detection	44
5.1.2 Validity check	45
5.1.3 Outlier detection	46
5.1.4 Controlli automatizzati sulla Plat4MedRob	47
5.2 Workflow di monitoraggio della qualità dei dati	47
5.3 Risultati	49
5.4 Limitazioni nel processo di data quality	53

Capitolo 6 - Controllo di qualità knowledge-based: una proposta di ontologia per le scale cliniche	56
6.1 Le ontologie: definizione e applicazioni	56
6.2 StrokeFit4 ontology	57
6.3 Adattamento dell'ontologia al problema di qualità del dato	60
6.4 Rilevanza per l'analisi dei dati	66
Capitolo 7 - Conclusioni e sviluppi futuri	67
Bibliografia	69

Capitolo 1 - Introduzione

Grazie all'avanzamento della tecnologia e dell'informatica i dati sono entrati a far parte della vita quotidiana delle persone, influenzando il modo in cui pensiamo e le scelte che prendiamo. Diventa quindi importante chiedersi: quanto sono affidabili?

Se non gestiti adeguatamente, dati incorretti, mancanti o polarizzati possono portare a trarre conclusioni errate e, nel contesto di uno studio, potenzialmente a invalidare i risultati ottenuti. Questo assume una speciale rilevanza in campo medico, in cui le conseguenze sono particolarmente impattanti: una decisione mal informata può compromettere l'efficacia di un trattamento clinico; inoltre se i dati usati non sono sufficienti a descrivere la popolazione d'interesse si rischia di propagare *bias* nella pratica clinica. Se questi dati sono poi usati per l'allenamento di algoritmi di intelligenza artificiale, è possibile amplificare involontariamente le possibili disparità sociali presenti tra i pazienti.

Esaminare la qualità dei dati su cui si lavora è quindi un processo fondamentale della ricerca che richiede attenzione e cura. Eseguire controlli periodici durante la fase di raccolta dei dati e mantenere una comunicazione con gli operatori clinici che la effettuano può essere molto vantaggioso per permettere di effettuare correzioni in itinere e ridurre i possibili errori nel *dataset* finale.

Ho avuto la possibilità di sperimentare questa modalità operativa durante il lavoro svolto presso il Laboratorio di Informatica Biomedica Mario Stefanelli (BMI) dell'Università di Pavia. In questa tesi descriverò il problema della qualità dei dati portando esempi dalla letteratura (capitolo 2), quindi illustrando il caso di studio e in particolare il protocollo adottato (capitolo 3), gli strumenti di raccolta dati (capitolo 4) e i conseguenti controlli eseguiti (capitolo 5). Terminerò il quinto capitolo parlando di alcune limitazioni riscontrate, e nel capitolo 6 proporrò una possibile estensione che permetta di integrare nell'algoritmo di

monitoraggio dei dati anche conoscenze mediche mediante l'utilizzo di un'ontologia. Le conclusioni sono infine riportate nel capitolo 7.

Capitolo 2 - Controllo di qualità del dato: definizione e background

In un'era in cui il dato clinico sta alla base dello sviluppo della conoscenza medico-scientifica, è sempre più necessario porsi il problema di valutare la qualità dei dati stessi e dell'informazione che da essi si ricava. Dati di scarsa qualità portano a conclusioni incorrette [1], come evidenziato dall'espressione “*garbage in, garbage out*”: dati incompleti, errati o non rappresentativi possono avere conseguenze importanti sulle successive analisi, alterando la valutazione degli *endpoint*, compromettendo la riproducibilità degli studi e riducendo la validità di eventuali modelli predittivi, che sono così portati a prendere decisioni inappropriate.

Definire la qualità del dato però non è banale. Spesso si fa riferimento al concetto di *fitness for use* proposto da Juran [2], secondo cui un dato è di qualità sufficiente se è utile al raggiungimento di un determinato scopo definito da un utente, ed è quindi una proprietà che dipende fortemente dal contesto. E' importante sottolineare che per un dato medico la qualità non è limitata alla correttezza formale, ma riguarda anche la validità clinica. La valutazione della qualità deve quindi tenere conto, oltre che del singolo dato, del contesto clinico e dei processi che lo hanno generato, nonché delle finalità per cui viene usato nell'analisi.

2.1 Background

Il tema della qualità dei dati è ampiamente documentato nella letteratura medica, in quanto, come accennato, dati di scarsa qualità possono invalidare i risultati delle analisi.

2.1.1 La provenienza dei dati

Un primo problema, presente principalmente negli studi retrospettivi, riguarda la fonte dei dati usati. Non sempre i siti che raccolgono e mettono a disposizione dei ricercatori *dataset* impongono che siano riportate informazioni come la provenienza dei dati, su come sono stati raccolti e sui possibili limiti riscontrati nella raccolta. Ad esempio, se non spiegata e adeguatamente considerata, l'assenza di alcune fasce di età o di peso nel campione può portare a conclusioni che non sono applicabili a quella parte di popolazione per cui è mancata una rappresentatività. Modelli di intelligenza artificiale che vengono allenati su questi dati e pubblicati possono essere alla meglio di scarsa utilità e alla peggio, nel caso in cui vengano applicati alla pratica clinica, rappresentare un rischio per il paziente, portando a false scoperte e consigli medici inadeguati [3]. Un secondo aspetto legato alla provenienza incerta è il rischio di usare *dataset* in cui la non chiarezza del processo di raccolta potrebbe nascondere una non conformità agli standard etici, e potenzialmente mettere in dubbio l'espressione del consenso informato all'utilizzo dei dati [4].

L'assenza di una provenienza verificabile compromette l'integrità dei dati utilizzati e la validità degli studi derivati. Per favorire l'accertamento dell'origine dei dati sono stati sviluppati alcuni strumenti e metodologie di indagine. Degno di nota è il lavoro svolto dalla fondazione *GoFAIR* [5], che offre una linea guida al riutilizzo dei dati. Secondo i principi proposti, un buon *dataset* è *FAIR*, ovvero gode delle seguenti proprietà:

- *Findability*: dati e metadati sono facilmente reperibili e leggibili, sia da un utente umano sia da un computer.
- *Accessibility*: una volta trovati i dati, è importante che siano accessibili. Questo è particolarmente rilevante per *dataset* non pubblici, per i quali potrebbe essere necessaria un'autenticazione.
- *Interoperability*: i dati devono prestarsi all'integrazione con altri dati e con applicazioni che ne permettono lo *storage* e il *processing*.

- *Reusability*: i metadati (e dati) devono essere descritti bene, in modo che possano essere replicati o complementati con dati provenienti da *setting* diversi.

2.1.2 Dati mancanti

Un problema che invece è trasversale a tutti gli studi basati su dati è quello della mancanza dei dati stessi. La *missingness* in statistica è divisa in tre tipologie:

1. *Missing Completely At Random (MCAR)*: si ha quando la mancanza di un dato è un evento indipendente sia dai parametri osservabili che da quelli non osservabili. Siccome non è correlata alle variabili dello studio, l'assenza del dato non porta a distorsioni delle analisi. Nella pratica è raro che i dati siano *MCAR*.
2. *Missing At Random (MAR)*: la mancanza del dato non dipende dal valore del dato stesso, ma è spiegabile utilizzando le variabili osservate per le quali abbiamo disponibili informazioni complete. Se non trattati in modo adeguato questi potrebbero avere un'influenza polarizzante sulle analisi.
3. *Missing Not At Random (MNAR)*: si verifica quando la probabilità che un dato sia mancante dipende da valori non osservati, anche dopo aver tenuto conto delle variabili osservate. In particolare, la mancanza può dipendere proprio dal valore che il dato avrebbe assunto se fosse stato rilevato. Questo meccanismo viene spesso definito non ignorabile, poiché, se non adeguatamente considerato, può introdurre distorsioni significative nelle analisi.

Esempi di come queste casistiche si possono presentare all'interno di uno studio sono riportati nella Tabella 1.

Tabella 1: tipi di *missing data* ed esempi

<i>Missing</i>	Esempio
<i>MCAR</i>	Il paziente dimentica di compilare un campo in un questionario.
<i>MAR</i>	L'età avanzata può spiegare la mancata compilazione di alcune domande di un questionario, ma la mancata compilazione, a parità di età, è indipendente dalle risposte che il questionario avrebbe riportato se fosse stato compilato.
<i>MNAR</i>	Persone con redditi molto alti o molto bassi potrebbero essere meno propense a dichiarare il proprio reddito per motivi di <i>privacy</i> (la mancanza è legata al valore della variabile stessa, che non viene quindi osservata)

I dati mancanti, soprattutto di tipo *MNAR*, possono introdurre polarizzazioni e quindi nel corso del tempo sono stati creati diversi metodi per la loro gestione, come ignorare *record* o *feature* con molti dati mancanti o utilizzare tecniche di *imputing* per completare i dati. Per buona pratica i tassi di *missingness* e i metodi applicati vanno accuratamente riportati e documentati, in modo da assicurare la validità e la ripetibilità delle analisi svolte [6].

2.1.3 La raccolta dati digitale

Gli studi basati su trial clinici devono sicuramente tenere conto dei principi *FAIR* per produrre *dataset* riutilizzabili, ma generalmente non soffrono delle problematiche relative alla scarsa documentazione delle fonti in quanto la raccolta dei dati necessari è parte dello studio stesso. Possono comunque essere soggetti a problemi legati alla mancanza dei dati e a errori di *reporting*.

La raccolta dati in clinica avviene generalmente utilizzando strumenti digitali, gli *electronic Case Report Forms (eCRF)*. La pratica clinica tuttavia ha ancora un

forte legame con gli strumenti cartacei, e questo dà spazio a possibili errori di trascrizione dei dati [7]. Grazie all'aspetto digitale degli *eCRF* è possibile inserire controlli sintattici, che riguardano il tipo di dato e l'adesione a una formattazione predefinita. Sono poi possibili controlli semantici, che riguardano l'applicazione di conoscenze tipicamente mediche come ad esempio i limiti tra cui ci si attende di trovare i valori. Questi permettono di limitare gli errori di compilazione, ma non possono evitarli del tutto: ad esempio non è possibile intercettare l'inesattezza di un valore sbagliato che però rientra nei *range* fisiologici. Oltretutto durante il processo di raccolta può essere molto complicato andare a valutare errori dovuti alla correlazione tra le variabili, sia perché alcuni dei dati potrebbero essere mancanti, sia perché non è detto che le informazioni correlate vengano raccolte nello stesso istante di tempo.

Un altro tema relativo alla raccolta cartacea è che per la trascrizione delle informazioni sulla *eCRF* potrebbe essere incaricata una persona che non possiede necessariamente una formazione medica e potrebbe quindi non essere in grado di notare alcune incorrettezze.

Uno studio sulla completezza di un *database* condotto da Lavieri et al. [8] ha identificato 5 principali barriere all'utilizzo degli *eCRF*:

1. La *usability* degli strumenti forniti, ovvero quanto sia semplice utilizzarli in modo corretto
2. Il *commitment* degli operatori che si occupano dell'inserimento dei dati, che potrebbero non percepire la responsabilità legata al ruolo, ricordare male le informazioni o non essere determinati
3. Problemi logistici che sono esterni agli *eCRF*, come potrebbero essere la gestione di grandi numeri di pazienti o una difficoltà di accesso alla strumentazione elettronica
4. La formazione degli utenti all'utilizzo degli strumenti elettronici
5. Lacune nella specificazione dei campi che potrebbero generare ambiguità

2.1.4 Strumenti di data quality assurance

Nel corso del tempo hanno ottenuto sempre più rilevanza due operazioni fondamentali nella valutazione della qualità dei dati, ovvero la *data quality assurance* e il *data quality assessment*. L'*assessment* avviene generalmente alla fine della raccolta e descrive gli errori riscontrati, mentre l'*assurance* è un processo continuativo che avviene di pari passo alla raccolta dei dati e prevede interventi volti a migliorare la qualità del *dataset* finale. In questa sezione parlerò di tecniche che possono assistere quest'ultima procedura.

Un primo possibile aspetto riguarda la trascrizione nei *Clinical Report Forms* di dati che sono già contenuti all'interno della cartella clinica del paziente (*EHR - Electronic Health Record*). In uno studio condotto da Cheng et al. [9] si è mostrato che in alcuni casi potrebbe essere utile implementare una mappatura tra *eCRF* e *EHR* che permetta di automatizzare la compilazione di questi campi, risparmiando tempo agli operatori clinici e riducendo la probabilità di commettere errori di trascrizione. Questo richiede generalmente molto tempo e per studi multicentrici presenta complicazioni dovute al fatto che i diversi ospedali potrebbero non condividere gli stessi sistemi di *EHR*, quindi questa tecnica risulta vantaggiosa solo se il numero di pazienti è effettivamente molto grande; per pochi pazienti rimane comunque più efficiente la compilazione manuale.

Esistono poi *framework*, come quello proposto da INGENIO-DQ [10], e soluzioni *open source* [11] che possono aiutare a identificare i problemi nella raccolta dati, tenerne traccia e segnalarli al lato clinico per collaborare verso la raccolta di dati di qualità. Per poter agire in questo senso la qualità va definita in modo chiaro e misurabile. Nella seguente sezione riporterò la definizione di qualità proposta da Weiskopf e Weng.

2.2 Le dimensioni della qualità del dato

Secondo la *review* [12] la qualità di un dato può essere espressa tenendo conto di 5 dimensioni, riportate nella Tabella 2.1.

Tabella 2.1: le dimensioni della qualità del dato

Dimensione	Risponde alla domanda
<i>Completeness</i>	L'informazione è presente?
<i>Correctness</i>	L'informazione presente è vera?
<i>Concordance</i>	Informazioni diverse sono consistenti tra di loro / rispetto a un'altra fonte?
<i>Plausibility</i>	L'informazione è verosimile e sensata?
<i>Currency</i>	L'informazione è attuale / rilevante per la descrizione del paziente in un certo istante?

Una successiva revisione del *paper* proposta da Lewis et al. [13] ha aggiunto altre due dimensioni, riportate nella Tabella 2.2.

Tabella 2.2: le dimensioni della qualità del dato

Dimensione	Risponde alla domanda
<i>Conformance</i>	L'informazione è conforme con lo standard di rappresentazione per quel tipo di dato?
<i>Bias</i>	Sono presenti distorsioni sistematiche che influenzano la disponibilità o la rappresentazione del dato?

2.1.1 *Completeness*

La *completeness* è la dimensione maggiormente controllata quando si parla di qualità dei dati. Si tratta della presenza delle informazioni su un paziente, e ha quindi a che vedere con la disponibilità dei dati (o, viceversa, con la loro *missingness*). La completezza può essere misurata osservando la presenza o l'assenza degli elementi all'interno dei *form*, e in caso di assenza può essere gestita confrontando il *dataset* ottenuto con un'altra fonte di informazione rilevante per il paziente detta *gold standard* (un esempio può essere una cartella clinica diversa da quella che si sta osservando).

La mancanza del dato in un contesto clinico non è sempre associata a una mancanza informativa. Un dato potrebbe essere assente in quanto non ancora trascritto e in alcuni casi potrebbe non essere stato rilevato in modo intenzionale, ad esempio per indisposizione del paziente. Un controllo di completezza del *dataset* non deve quindi limitarsi a verificare la presenza del dato, ma deve tenere conto del motivo per cui un dato potrebbe mancare e delle implicazioni che questo può avere sulla validità dei risultati.

2.1.2 *Correctness*

La *correctness* è una misura di quanto gli elementi presenti siano corretti, ovvero “contengono informazioni vere”. Questo concetto è spesso legato ai concetti di accuratezza ed errore. Non si tratta quindi di una verifica di validità formale, ma di veridicità del dato.

La correttezza si può valutare utilizzando il *gold standard*, controllando la coerenza tra elementi corrispondenti o in alcuni casi effettuando *check* di validità, con cui si può ad esempio verificare se la misura riportata rientra nei range attesi di una certa variabile.

2.1.3 *Concordance*

La *concordance* è presente quando c'è compatibilità tra gli elementi raccolti. Ad esempio, si può valutare la coerenza tra due dati diversi che misurano la stessa (o quasi la stessa) informazione, oppure tra campi che riguardano informazioni diverse che hanno senso se considerate insieme.

Generalmente i controlli di concordanza si effettuano confrontando campi di uno stesso *record*, ma talvolta possono anche riguardare confronti con altre fonti di informazioni (come *record* cartacei o informazioni riportate dal paziente) o con distribuzioni delle stesse informazioni in *dataset* provenienti da pratiche mediche simili. Questi controlli sono particolarmente interessanti perché indagano relazioni semantiche tra i dati.

2.1.4 *Plausibility*

Un ulteriore controllo importante è quello di plausibilità, che si ha se c'è concordanza tra il dato misurato e la conoscenze pregresse dalla letteratura medica o da altre informazioni contestuali. La plausibilità è spesso collegata alla validità del dato raccolto e alla sua integrità.

I controlli possono includere la verifica che il dato stia dentro a range biologicamente plausibili o confronti con distribuzioni di altri *dataset*. Un particolare esempio può essere l'*outlier detection*, ponendo però particolare attenzione a non confondere la rarità del dato con la sua implausibilità.

Analisi di concordanza e plausibilità sono modi possibili di stimare la correttezza del dato quando non è possibile misurarla in modo diretto, e in alcuni casi queste due dimensioni sono considerate più come approcci metodologici che come effettive proprietà del dato.

2.1.5 *Currency*

La *currency* di un dato spesso è legata al fatto che sia rilevato per tempo o che sia recente. I dati sono ritenuti correnti se sono raccolti in un intervallo di tempo ragionevole per la descrizione del paziente, oppure se rappresentano adeguatamente un paziente in un certo periodo di interesse.

Un modo per valutare se un dato è corrente è riferirsi ai *log* di *data entry*.

2.1.6 *Conformance*

La *conformance* si ha in relazione all'aderenza del dato a una struttura relazionale predefinita. Questa dimensione è stata aggiunta per tenere conto del cambiamento metodologico portato dall'adozione di *framework* standardizzati, che permettono non solo di cogliere informazioni intrinseche e contestuali dei dati raccolti ma incentivano anche la comunicazione e collaborazione tra team di ricerca. Questo evidenzia l'importanza di avere dati strutturati.

2.1.7 *Bias*

La dimensione del *bias* è stata introdotta per valutare i dati che sono *Missing Not At Random*. In questi casi ignorare la mancanza del dato può portare a trarre conclusioni sbagliate dai dati, e talvolta a perpetuare all'interno del contesto medico pregiudizi e ineguaglianze sociali sistemiche.

La presenza di *bias* nei dati è anche un buon esempio di meccanismi nascosti presenti nei contesti di raccolta dati reali che hanno effetti sulla qualità. Comprendere questi meccanismi è necessario per andare a individuare i punti chiave su cui lavorare per migliorare la qualità del dato e il suo ciclo di vita.

2.1.8 Tecniche di controllo della qualità dei dati

Riporto nella Tabella 3 un riassunto dei metodi di *data quality assessment* più comunemente usati, e la loro rilevanza nel misurare ciascuna dimensione.

Tabella 3: metodi per la data quality

Metodo	Spiegazione
<i>Gold standard comparison</i>	Un <i>dataset</i> proveniente da un'altra fonte (o da fonti multiple) viene usato come <i>gold standard</i> per confrontare le informazioni raccolte.
<i>Data element agreement</i>	Due o più elementi di uno stesso <i>record</i> sono confrontati per vedere se riportano la stessa informazione o informazioni compatibili.
<i>Element presence</i>	Determinazione della presenza di un certo elemento.
<i>Data source agreement</i>	Confronto dei dati del <i>record</i> con i dati di un'altra fonte per vedere se sono concordi.
<i>Distribution comparison</i>	Confronto tra le distribuzioni o le statistiche riassuntive dei dati del <i>record</i> e quelle clinicamente attese per i concetti d'interesse.
<i>Validity check</i>	Insieme di tecniche per verificare che i dati presenti nel <i>record</i> abbiano senso.
<i>Log review</i>	Esame delle informazioni sulle pratiche di <i>data entry</i> (come date di inserimento o di modifica dei dati).
<i>Structural agreement</i>	Adesione della struttura del dato a strutture predefinite.

Capitolo 3 - Il caso di studio: il progetto Fit4MedRob e il trial clinico StrokeFit4

Fit For Medical Robotics, abbreviato in *Fit4MedRob*, è un'iniziativa italiana, finanziata dal Ministero dell'Università e della Ricerca, iniziata a Dicembre 2022 e con termine previsto a Novembre 2026, mirata alla promozione di processi di riabilitazione robotica rivolti a persone con compromesse o assenti capacità motorie, sensoriali o cognitive [14].

L'attuale quadro della pratica riabilitativa, in Italia così come in tanti altri Paesi, presenta alcune criticità, tra cui ad esempio una inadeguata personalizzazione del trattamento o una mancata continuità della cura, il che può portare a risultati che sono insoddisfacenti per il paziente a fronte di costi elevati per il Sistema Sanitario. La recente evoluzione delle tecnologie robotiche e bioingegneristiche potrebbe rappresentare una risposta a queste problematiche, portando a miglioramenti significativi nell'accessibilità e nell'efficacia della riabilitazione. Non ci sono tuttavia in letteratura prove scientifiche abbastanza forti che provano questa ipotesi. Infatti gli studi condotti finora presentano risultati contrastanti, e sono per lo più stati condotti su campioni di pazienti troppo poco numerosi e troppo eterogenei.

Il progetto *Fit4MedRob* vuole portare a un avanzamento della conoscenza scientifica su queste tematiche sotto tre principali aspetti, formulati come *missions*:

- *Mission 1*: confrontare, mediante *trial* pragmatici di grosse dimensioni, l'efficacia della riabilitazione tradizionale con quella della riabilitazione assistita da tecnologia robotica. A tale scopo i suddetti *trial* sono stati condotti (alcuni sono tuttora in corso) mediante tecnologie già presenti in commercio. Prima dell'avvio dei *trial*, *Mission 1* ha condotto una serie di *surveys* per capire quali sono i bisogni di pazienti e terapisti in questo ambito. Obiettivo secondario di *Mission 1* è stato quindi quello di sperimentare nuovi modelli organizzativi di riabilitazione che vadano

incontro ai bisogni degli utenti e allo stesso tempo migliorino l'efficienza dei reparti di riabilitazione.

- *Mission 2*: anche sulla base dei risultati delle *surveys* di *Mission 1*, progettare e sperimentare l'efficacia di nuove tecnologie, non ancora presenti sul mercato. Alcuni prototipi erano già presenti all'inizio di *Fit4MedRob*, altri sono stati ideati e sviluppati nel corso di questi 4 anni.
- *Mission 3*: questa mission riguarda la ricerca di base nell'ambito della robotica, come lo studio di nuovi materiali e algoritmi per avanzare le tecnologie di nuova generazione, mantenendo contemporaneamente obiettivi di sostenibilità e conformità alle linee guida etiche e legali.

Mi concentrerò nella prossima sezione su un particolare trial di *Mission 1* denominato *StrokeFit4*. Questo studio, registrato su *clinicaltrials.gov* [15] con l'id **NCT06547827**, propone un confronto tra la riabilitazione tradizionale e la riabilitazione assistita da *Robot and Allied Digital Technologies (RADT)* per quanto riguarda il recupero funzionale dell'indipendenza nelle attività della vita quotidiana in pazienti sopravvissuti a un ictus.

Nel corso della tesi triennale, ho avuto modo di partecipare a diverse riunioni con i medici e i terapisti coinvolti in *StrokeFit4* e questo mi ha permesso di acquisire una certa familiarità con i dati che dovevano essere raccolti e con i possibili problemi che sarebbero potuti sorgere nella raccolta stessa.

3.1 Il protocollo di *StrokeFit4*

Le *RADT* sembrano rappresentare un importante miglioramento dell'efficacia nella riabilitazione di pazienti con esiti di ictus, tuttavia, come già accennato in termini generali, la letteratura attualmente presente a riguardo non porta risultati conclusivi sul tema a causa di una serie di fattori, come l'eterogeneità in termini di protocolli terapeutici o un numero insufficiente di pazienti reclutati. Spesso inoltre gli studi pubblicati si focalizzano su un *device* o un distretto corporeo, perdendo la visione globale sia del paziente che dei costi.

StrokeFit4 [16] propone quindi un'analisi che confronti questo tipo di terapia con quella tradizionale osservando il recupero di tutte le funzioni compromesse, e analizzando contemporaneamente la sostenibilità economica del trattamento proposto.

3.1.1 *Endpoint* dello studio

L'*outcome* primario dello studio è dimostrare la non inferiorità della riabilitazione integrata con dispositivi robotici rispetto a quella tradizionale in pazienti nella fase subacuta dell'ictus per quanto riguarda il recupero delle abilità di vita quotidiana (*ADL - Activity of Daily Living*), misurato mediante il *modified Barthel Index (mBI)*.

L'ampia quantità di dati raccolta permetterà in seguito di valutare alcuni obiettivi secondari:

- Dimostrazione dell'eventuale superiorità del trattamento riabilitativo con *RADT* rispetto al trattamento tradizionale, qualora fosse verificata la non inferiorità
- Confronto tra i due gruppi rispetto ai vari domini target della riabilitazione, in accordo con all'*International Classification of Functioning, Disability and Health (ICF)*
- Analisi di parametri neurofisiologici e fattori coinvolti nella

neuroplasticità

- Confronto tra i due gruppi del tasso di recupero della destrezza manuale e delle performance nel cammino
- Valutazioni degli effetti del trattamento sulle ADL e sulla qualità di vita durante un *follow up* di 6 mesi
- Creazione di un modello predittivo in grado di prevedere l'efficacia del trattamento robotico e tecnologico in soggetti con esito di ictus
- Valutazione della sostenibilità economica del trattamento per il paziente, il pagatore e la società, con modelli di costo-efficacia, di costo-utilità e, per il Sistema Sanitario, dell'Analisi di Impatto sul Budget.

3.1.2 Caratteristiche

La dimensione campionaria è stata calcolata considerando di voler ottenere una potenza pari all'80% rispetto all'*outcome* primario (*mBI*) con un intervallo di confidenza bilaterale al 95% e una deviazione standard di 20 punti. La numerosità risultante è di 506 pazienti (253 per gruppo), aumentati a 596 per considerare un *rate* di *drop-out* dallo studio del 15%.

L'arruolamento è affidato a 27 centri clinici afferenti al progetto e ha una durata prevista di 28 mesi (agosto 2024 – novembre 2026). I pazienti vengono randomizzati da un ente centrale (Università di Pavia) in due bracci paralleli, uno relativo alla riabilitazione tradizionale e uno relativo a quella robotica. La randomizzazione viene stratificata per centro clinico, latenza dall'evento ($>$ o $\leq$ 3 mesi) e *setting* clinico (degenza / ambulatorio).

La Tabella 4 riporta i criteri di inclusione ed esclusione dello studio.

Tabella 4: criteri per l'arruolamento nel *trial StrokeFit4*

Criteri di inclusione	Criteri di esclusione
<ul style="list-style-type: none"> - Diagnosi di primo ictus, ischemico o emorragico, verificata mediante TAC o Risonanza Magnetica - Età superiore a 18 anni - Latenza dall'evento ≤ 6 mesi - Da lieve a grave compromissione dell'arto superiore (sezione motoria della scala Fugl-Meyer Assessment of Upper Extremity ≤ 58) e/o da lieve a grave compromissione dell'arto inferiore (punteggio della scala Functional Ambulation Categories ≤ 4) e/o lieve a grave compromissione dell'equilibrio (Berg Balance Scale ≤ 45); - Condizione clinica stabile tale da permettere il trasferimento in palestra e l'esecuzione dei trattamenti previsti. 	<ul style="list-style-type: none"> - Instabilità clinica - Disturbi comportamentali/cognitivi che impediscono un'adeguata compliance del paziente al trattamento riabilitativo sia convenzionale che robotico (deficit cognitivo severo, Montreal Cognitive Assessment <10) - Diagnosi concomitante di altre patologie neurologiche, psichiatriche o internistiche che, a giudizio dello sperimentatore clinico, possano interferire con le procedure di studio e/o impattare in modo significativo con l'outcome riabilitativo - Rigidità o ipertono (Modified Ashworth Scale > 3) nell'arto plegico/paretico; - Gravi disturbi visivi non correggibili mediante lenti che impediscono al paziente di eseguire il trattamento con dispositivi tecnologici e/o robotici; - Donne in gravidanza; - Rifiuto o impossibilità di firmare il consenso informato.

Una volta randomizzati, i pazienti vengono sottoposti a un protocollo riabilitativo che consiste in 25 sedute da 45 minuti, con frequenza di:

- 5 volte a settimana per 5 settimane per i pazienti ricoverati (*inpatients*)
- 3 volte a settimana per 8,3 settimane per i pazienti ambulatoriali (*outpatients*)

La riabilitazione, adattata a ciascun paziente, riguarderà tre principali aree, in accordo con quanto previsto dall'*ICF*:

1. *Funzioni sensori-motorie*: funzioni neuro-muscolo-scheletriche legate alla mobilità e alla stabilità, funzioni sensoriali e del dolore, funzioni di tolleranza all'esercizio, ...
2. *Funzioni mentali specifiche*: attenzione, memoria, funzioni percettive, funzioni cognitive di livello superiore, funzioni mentali del linguaggio, ...
3. *Attività e partecipazione*: cura di sé, attività della vita quotidiana

In seguito al termine del trattamento seguirà un *follow up* di 6 mesi incentrato sull'efficacia percepita dal paziente nel recupero delle *ADL* e nell'impatto sulla qualità della vita.

Per la natura stessa dello studio, esso non può essere condotto in cieco per quanto riguarda pazienti e operatori clinici, quindi a essere svolta in cieco sarà solo la valutazione dei risultati.

Lo studio è quindi multicentrico, multimodale, randomizzato, controllato, parallelo, a valutatori in cieco e di tipo pragmatico.

3.1.3 - Valutazione delle ADL con mBI

Il Barthel Index è una scala per la valutazione dell'indipendenza del paziente nello svolgimento di *task* relativi ad attività della vita quotidiana. Il *modified Barthel Index* [17], proposto da Shah, Vanclay e Cooper al fine di aumentare la sensibilità della scala, valuta il paziente in base a un punteggio in 100esimi che somma i punteggi ottenuti nei seguenti domini:

1. Alimentazione
2. Igiene personale
3. Capacità di farsi il bagno o la doccia
4. Abbigliamento
5. Trasferimenti carrozzina/letto
6. Uso del WC
7. Continenza urinaria
8. Continenza intestinale
9. Deambulazione / uso della carrozzina
10. Capacità di fare le scale

Il test è di facile somministrazione, alta affidabilità ed è ripetibile nel tempo e per questo è ampiamente diffuso nella pratica clinica, soprattutto per quanto riguarda gli studi sulla riabilitazione post-ictus. A un punteggio di 0 corrisponde un livello di disabilità totale e a un punteggio di 100 corrisponde la totale indipendenza. Per quanto riguarda gli score intermedi la minima differenza clinicamente rilevante (*MCID*, *Minimal Clinically Important Difference*) è di 9.25, anche se generalmente i punteggi vengono suddivisi in classi di disabilità [18] come descritto nella Tabella 5.

Tabella 5: *cutoff* comunemente utilizzati per la scala *mBI*

Cutoffs	Livello di disabilità
0 - 24	Disabilità totale
25 - 49	Disabilità grave
50 - 74	Disabilità moderata
75 - 89	Disabilità lieve
90 - 99	Disabilità minima
100	Nessuna disabilità

3.1.4 - Disegno di trattamento e dati raccolti

Durante il corso del trial viene raccolta una grande quantità di dati molto eterogenei. Nelle Tabelle 6.1 e 6.2 è riportato un riassunto estratto dal protocollo che mette in evidenza i dati e i tempi in cui vengono raccolti nel corso della riabilitazione. I tempi sono indicati nel seguente modo:

- *E*: arruolamento del paziente
- *T0*: prima seduta riabilitativa
- *Ts_x*: x-esima seduta
- *T1*: ultima seduta riabilitativa (25esima)
- *Tm_x*: valutazione di follow up effettuata x mesi dopo il *T1*

Tabella 6.1: tempistiche della raccolta dati nel *trial StrokeFit4*

		E	T0	Ts_5	Ts_10	Ts_15	Ts_20	T1	Tm_1	Tm_2	Tm_3	Tm_4	Tm_5	Tm_6
Screening	Consenso informato	X												
	Anagrafica	X												
	Dettagli sull'evento acuto	X												
	Fugl-Meyer Assessment - Arti superiori	X												
	Functional Ambulation Categories	X												
	Berg Balance Scale	X												
	Randomizzazione	X												
	Stile di vita pre-morbidità													
	Modified Rankin Scale		X											
Baseline Assessment	Riserva cognitiva (CRiQ)													
	Attività fisica (LTPAQ)													
	Quadro clinico													
	Gravità dell'ictus (NIHSS)													
	Thumb Localizing Test (TLT)		X											
	Test del monofilamento di Semmes - Weinstein													
	Fatigue Severity Scale													
	Precedenti setting riabilitativi		X											
	Questionari ad hoc													
	Comorbidità													
	Cumulative Illness Rating Scale (CIRS)		X											
	Situazione sociale													
	Blaylock Risk Assessment		X											
	Screening Score (BRASS)													
	Stato nutrizionale													
	Mini Nutritional Assessment - Short Form		X											
	Functional Oral Intake Score													
	SARC-F													
	Handgrip													
	Quadro clinico psicologico													
	Montreal Cognitive Assessment (MoCA)		X											
	Beck Depression Inventory													
	Hospital Anxiety and Depression Scale (HADS)													
	Intrinsic Motivation Inventory		X											
	Analisi genetica		X											
	Prelievo del sangue													
Outcome primario	Modified Barthel Index (mBI)		X					X						

Tabella 6.2: tempistiche della raccolta dati nel *trial StrokeFit4*

		E	T0	Ts_5	Ts_10	Ts_15	Ts_20	T1	Tm_1	Tm_2	Tm_3	Tm_4	Tm_5	Tm_6
Outcome secondari	Scale cliniche													
	Fugl-Meyer Assessment - Arti superiori (funzioni sensoriali e motorie)													
	Fugl-Meyer Assessment - Arti inferiori (funzioni sensoriali e motorie)													
	Berg Balance Scale													
	Symbol Digit Modalities Test													
	Modified Ashworth Scale - arto superiore e inferiore		X					X						
	Motricity index per arto superiore e inferiore													
	Trunk Control Test													
	Numerical Rating Scale per il dolore													
	Doleur Neuropathique 4													
	Action Research Arm test													
	2-Minute Walking Test													
	EQ-5D-5L													
	Neurofisiologia													
	Elettroencefalografia quantitativa (qEEG)		X					X						
Accettabilità e usabilità delle tecnologie	Analisi biochimiche													
	Prelievi del sangue per analisi del siero		X					X						
	Recovery Rate													
	Box and Block test		X	X	X	X	X	X						
	10-metre walking test		X	X	X	X	X	X						
	Follow-up													
	Modified Barthel Index (mBI)								X	X	X	X	X	X
	EQ-5D-5L													
	Questionari ad hoc													
	EQ-5D-5L		X			X		X						
Accettabilità e usabilità delle tecnologie	Technology Acceptance Model			X		X		X						
	Computer Skills		X											
	Soddisfazione (questionario ad hoc)		X					X						
	Complessità (questionario ad hoc)		X					X						
	Valutazione dei costi													
	Questionario ad hoc		X			X		X						
	Sessioni riabilitative							X						
	Questionario ad hoc													
	Valutazioni di sicurezza													
Accettabilità e usabilità delle tecnologie	Eventi avversi													

Capitolo 4 - Strumenti di raccolta dati in *StrokeFit4*

La raccolta dei dati per il progetto *StrokeFit4* è organizzata su due principali piattaforme:

- *Research Electronic Data Capture (REDCap)*[19], una piattaforma dedicata alla raccolta di dati strutturati per gli studi clinici;
- *Plat4MedRob*, una piattaforma per la raccolta dei dati non strutturati, come quelli derivati da *imaging* biomedico e altri segnali.

Le prossime sezioni descrivono il funzionamento delle piattaforme, in particolare considerando il modo in cui i dati vengono raccolti, le implicazioni a livello di verifica della qualità e l'integrazione tra le informazioni provenienti dai due sistemi.

4.1 *REDCap*

REDCap è una piattaforma online sviluppata dalla Vanderbilt University per rispondere alle esigenze di raccolta di dati negli studi clinici. Caratteristiche chiave che rendono questo tool ampiamente utilizzato nella pratica clinica sono l'intuitività, la flessibilità e l'orientamento al lavoro in team.

La raccolta dati è organizzata in *eCRF* chiamati *data collection instruments*, con i quali è possibile raccogliere un grande numero di tipi di dati. Gli strumenti vengono definiti e testati in fase di development e sono disponibili agli utenti che si occupano di *data entry* quando il progetto viene messo in produzione. E' consigliabile effettuare tutte le correzioni necessarie in fase di sviluppo, perché quando il progetto è in modalità *production* determinate modifiche potrebbero risultare distruttive e comportare quindi una perdita di dati.

Una funzionalità che è stata di particolare importanza durante questo studio è la modalità di studio longitudinale, che dà la possibilità di organizzare la struttura

della raccolta dati in eventi. Una volta definiti gli eventi è possibile stabilire quali *instruments* vanno raccolti in ogni evento e per quali di questi è possibile inserire più di una risposta. Questo rende questa modalità adattabile virtualmente a qualsiasi protocollo longitudinale.

Il processo di *data entry* si basa sulla creazione di un *record*. Il *record* generalmente rappresenta un paziente ed è composto da tutti gli strumenti previsti nel *setup* del progetto (tenendo conto della divisione in eventi). I *record* inseriti sono visualizzabili in una *dashboard*.

A coordinare gli utenti che collaborano a un progetto è un particolare utente detto *administrator*, che agisce sugli altri nei seguenti modi:

- Modificando gli *user rights* su due piani:
 - gestendo le modalità di accesso al sito, rendendo visibili o nascoste determinate funzionalità quali gli strumenti di modifica dei *form*, i log del progetto o i sistemi di *alert*.
 - gestendo i permessi di accesso ai *record* inseriti (*nessuno, read only, view & edit, delete*) e alla modalità di esportazione dei dati (*nessuno, de-identified, remove all identifier fields, full dataset*)
- Assegnando gli utenti a *Data Access Groups (DAGs)*, e limitando quindi i *record* visibili e modificabili a quelli creati da altri utenti presenti nello stesso *DAG*. Questo è utile ad esempio per dividere gli utenti dedicati al *data entry* sulla base del centro clinico per cui lavorano, il che serve la doppia funzione di associare immediatamente il paziente al centro di cura e di rendere inaccessibili i dati dei pazienti a utenti provenienti da altri ospedali.
- Accettando o rifiutando eventuali modifiche alla struttura degli *eCRF* che potrebbero rendersi necessarie nel corso del progetto.

Queste funzionalità e la facilità di utilizzo di *REDCap* lo hanno reso uno strumento fondamentale per la raccolta dei dati strutturati. Nel corso della mia

attività di tesi triennale ho contribuito a implementare i *data collection instruments* che sono attualmente in uso per la raccolta dei dati di *StrokeFit4*. La maggior parte degli strumenti rappresenta scale cliniche validate, brevemente descritte nelle Tabelle 7.1 e 7.2. Gli strumenti *REDCap* relativi alle scale che non sono protette da *copyright* sono disponibili su *GitHub* al seguente link: [bmi-labmedinfo/REDCap_Rehabilitation_Measures: This repo contains REDCap versions of widely used clinical outcomes scales in rehabilitation](https://github.com/bmi-labmedinfo/REDCap_Rehabilitation_Measures).

Tabella 7.1: Le scale cliniche utilizzate per la valutazione dei pazienti in
StrokeFit4

ICF Domain	Scale	Outcome	Outcome measure [range]	Self-admin	Copyright	REDCap entry option
Body Function	National Institutes of Health Stroke Scale (NIHSS) [31]	Stroke severity	Global numerical score [0–42]	No	No	15 fields, each with Multiple Choice – Drop-down List (Single Answer)
Body Function	Thumb Localizing Test [32]	Proprioception	Partial numerical scores [0–3]	No	No	Matrix of 12 fields – one choice per line, from 0 (normal) to 3 (severely impaired)
Body Function	Semmes Weinstein Filament test [32]	Sensory perception	Global numerical score [1–5]	No	No	1 Multiple Choice – Drop-down list (Single Answer). The choice represents the lighter filament felt by the patient
Activity, Participation	Fatigue Severity Scale [33]	Fatigue	Global numerical score [9–63]	No	No	Matrix of 9 fields – one choice per line, from 1 to 7
Body Function	Cumulative Illness Rating Scale (CIRS) [36]	Burden of chronic illness across multiple systems	2 Global numerical scores	No	No	Matrix of 14 fields – one choice per line, from 1 (no impairment/damage) to 5 (very severe)
Body Function	Fugl-Meyer Assessment (lower limb) [34]	Motor recovery, coordination and reflex activity of the LL	4 global numerical scores	No	No	7 matrix of fields – one choice per line
Body Function	2-minute Walk test [37]	Walking endurance	Numerical value (distance in meter)	No	No	Text box (Number, minimum number 0)
Body Function	10-metre walking test [38]	Walking speed	Numerical values (velocity)	No	No	Optional field: aid used by the patient
Body Function	Fugl-Meyer Assessment (upper limb) [34]	Motor recovery, coordination and reflex activity of the UL	7 global numerical scores	No	No	Text boxes (Number, minimum number 0)
Body Function	Action Research Arm Test (ARAT) [39]	Ability to perform movements	4 global numerical scores	No	Yes	Optional field: aid used by the patient
Activity	Box and Block Test [40]	Gross manual dexterity	2 global numerical scores	No	No	7 matrix of fields – one choice per line
Activity	Premorbid Modified Rankin Scale [41]	Functional Independence	Global numerical score [0–6]	No	No	7 matrix of fields – one choice per line
Activity	Modified Barthel Index (MBI) [42]	Functional Independence	Global numerical score [0–100]	No	Yes	12 Multiple Choice – Drop-down List (Single Answer)
Body Function	Cognitive Reserve (CRiq) [43]	Cognitive reserve	Global numerical score	No	Yes	Text boxes (Number, minimum number 0)
Activity, Participation	Lifetime Physical Activity (LTPAQ) [44]	Physical activities levels	-	No	No	Text boxes (Numbers), Multiple Choice – Radio buttons (Single and Multiple Answer)
Body Structure, Body Function, Participation	5-level EuroQol-5D version (EQ-5D-5 L) [45]	Health-related quality of life	Global numerical score [0–1]	No	Yes	Multiple Choice – Radio buttons (Single Answer), Slider / Visual Analog Scale
Body Function and Structure, Activities and Participation, Environmental factors	Blaylock Risk Assessment Screening Score (BRASS) [45]	Risk of prolonged hospital stay	Global numerical score [0–36]	No	No	Multiple Choice – Drop-down List (Single Answer) and "Multiple Choice – Drop-down List (Multiple Answers)"
Body Function	Mini-nutritional assessment [46]	Nutritional status	Global numerical score [0–14]	No	No	Multiple Choice – Radio Buttons (Single Answer)
Body Function	Functional Oral Intake score [47]	Nutritional status	Global numerical score [1–7]	No	No	Multiple Choice – Radio Buttons (Single Answer)

Tabella 7.2: Le scale cliniche utilizzate per la valutazione dei pazienti in
StrokeFit4

ICF Domain	Scale	Outcome	Outcome measure [range]	Self-admin	Copyright	REDCap entry option
Body Function, Activities and Participation	SARC-F [48]	Sarcopenia screening	Global numerical score [0–10]	Yes	No	Multiple Choice – Radio Buttons (Single Answer)
Body Function	Handgrip [49]	Muscle strength	Global numerical score	No	No	Multiple Text Box field (Number)
Body Structure and Function	Montreal Cognitive Assessment (MoCA) [50]	Detection of mild cognitive impairment and early dementia	Global numerical score [0–30]	No	No	Descriptive text with figures illustrating the exercise, and numeric text boxes
Body Function	Beck Depression Inventory (BDI) [51]	Depressive symptoms severity	Global numerical score [0–63]	Yes	Yes	Multiple Choice – Radio Buttons (Single Answer)
Body Function	Hospital Anxiety and Depression Scale [52]	Anxiety and depression symptoms	2 Global numerical score	Yes	Yes	Multiple Choice – Radio Buttons (Single Answer)
Body Function	Intrinsic Motivation Inventory (IMI) [53]	subjective experience related to a specific activity or task	4 partial numerical scores	Yes	Yes	Matrix of fields (one choice per line, from 1 to 7)
Body Function	Symbol Digit Modalities [54]	Processing speed, attention and visual motor coordination	Global numerical score [0–110]	No	Yes	Text boxes (Number)
Activity	Berg Balance Test (BBT) [35]	Balance and risk of failure	Global numerical score [0–56]	No	No	Multiple Choice – Radio Button (Single Answer)
Body Structure and Function	Modified Ashworth Scale [55]	Muscle spasticity	-	No	No	Matrix of Field

4.1.1 I controlli di qualità nativamente implementati in *REDCap*

La piattaforma *REDCap* è fornita di funzionalità che permettono di considerare la qualità dei dati già a livello di design e di monitorarla nel corso dell'inserimento dei dati. Nelle prossime sezioni descriverò le funzionalità utilizzate per il *trial StrokeFit4* e altri accorgimenti presi a livello di sviluppo per favorire la qualità dei dati.

4.1.1.1 *Flag Required/Identifier*

I campi possono essere identificati come *required* o *identifier*, il che rende obbligatoria la compilazione. L'effetto sulla compilazione è di far visualizzare all'utente un alert sulla mancata compilazione al momento del salvataggio del form. La differenza nell'uso del flag *identifier* riguarda invece principalmente l'esportazione dei dati, durante la quale è possibile escluderli.

Questa funzionalità è importante come notifica all'utente, che potrebbe aver saltato un campo per distrazione, ed è quindi stata inserita ovunque possibile. Bisogna però puntualizzare che all'utente viene comunque permesso di salvare i dati inseriti ignorando il *flag* su quelli mancanti, e addirittura di segnare il *form* come completo. Questo vincolo quindi è utile a chi fa *data entry*, ma non è sufficiente per affermare la completezza del *dataset* ottenuto.

In *REDCap* pertanto un campo va dichiarato “obbligatorio” quando lo studio sottostante lo ritiene veramente necessario ai fini delle analisi dati previste, ma non per questo la sua mancata compilazione impedisce la memorizzazione della scheda. Questa scelta è stata fatta perché impedire la memorizzazione di una scheda potrebbe portare alla perdita dei dati già inseriti e questo costituirebbe una perdita di tempo per i *data entry*, spesso impegnati anche in altre attività cliniche.

4.1.1.2 *Field Validation*

Come anticipato *REDCap* è in grado di raccogliere una grande quantità di tipi di dati e modificare la visualizzazione del campo a seconda del tipo raccolto. Un esempio può essere la visualizzazione di *radio button* o tendine per domande a scelta multipla, o *checkbox* quando è possibile esprimere più di una scelta.

Tutti i tipi di dato esprimibili come testo (stringhe libere, stringhe con una certa formattazione, numeri interi, numeri reali, date, ...) condividono la stessa rappresentazione a livello di campo, ovvero la *text box*: questo potrebbe portare a difficoltà notevoli nel processo di analisi, in quanto utenti diversi potrebbero rappresentare le informazioni richieste utilizzando formati diversi. L'esempio più comune di questa problematica è la rappresentazione della data (gg/mm/aaaa, gg/mm/aa, gg-mm-aaaa, mm-gg-aaaa, aaaa-mm-gg, gg/mm/aaaa hh:mm:ss, ...).

Per questo motivo il *Designer*, sezione utilizzata per creare i *form*, permette di applicare al campo di testo una validazione. Esistono validazioni predefinite, ad esempio per data, *e-mail*, numero di telefono, ma anche *integer* e *numeric* per standardizzare la rappresentazione dei valori numerici. Per i campi numerici e le date è possibile inoltre specificare un range di validità del valore inserito. Il campo *Field note* può essere usato per esplicitare il tipo e il formato di validazione.

Questa opzione rende impossibile salvare dati fuori *range* o che non rispettano la validazione specificata, e questo è stato molto utile ad esempio per ottenere una rappresentazione uniforme dei valori decimali, forzando tutti gli utenti a definirli con il punto.

4.1.1.3 *Branching logic*

Talvolta la compilazione dei campi può essere soggetta a determinate logiche: per esempio se una scala non va applicata a un determinato tipo di paziente, la presenza di dati può essere come minimo fonte di confusione in fase di analisi, oltre che essere un affaticamento evitabile per chi inserisce i dati.

REDCap prevede questo tipo di struttura e la implementa mediante *branching logic*. Grazie a questa opzione è possibile rendere la visualizzazione di uno o più campi condizionale rispetto ai valori assunti da campi diversi, precedentemente inseriti. E' inoltre possibile applicare queste logiche direttamente agli strumenti, rendendo non compilabili interi *form* in assenza di determinate condizioni prestabilite.

Questa funzionalità non si è rivelata particolarmente necessaria nell'implementazione degli strumenti di questo studio, in quanto per la maggior parte gli strumenti rappresentano scale cliniche validate che non presentano questo tipo di logica compilativa.

4.1.1.4 *Action tags*

Gli *action tags* sono molteplici e variegati, e permettono di personalizzare l'esperienza di compilazione dei *form*. L'utilizzo che ne abbiamo fatto all'interno dello studio è stato principalmente basato sul tag *@Default*, che permette di precompilare i campi se si ha già un'idea del valore che verrà inserito (per considerazioni sulla frequenza o perché deducibile da campi compilati in precedenza), al fine di velocizzare la compilazione e di prevenire dati mancanti. Ad esempio, inizializzando tutte le comorbidità del paziente al valore "no" si riduce il carico di lavoro richiesto durante l'inserimento dei dati. Naturalmente questa funzionalità va applicata con cautela quando il dato mancante e il "no" vanno tenuti distinti in fase di analisi.

4.1.1.6 *Scheduling mode e modalità di alert*

La *scheduling mode* è una particolare funzionalità della modalità longitudinale. Permette di definire distanze temporali tra gli eventi e notificare via *mail* agli utenti che bisogna compilare i *form*. Viste le brevi tempistiche del trattamento, questa modalità sarebbe stata più un sovraccarico informativo nei confronti degli utenti che un *reminder* utile, quindi abbiamo scelto di non adottarla.

Esiste poi una sezione che permette di definire *alert* sulla base di un certo *trigger*, che può essere il salvataggio di un modulo o il verificarsi di una certa condizione logica. Anche questa funzionalità non si adattava bene alle modalità dello studio.

4.1.2 Controlli di qualità implementati *ad hoc* per lo studio: accorgimenti progettuali

4.1.2.1 *Uso di calculated fields per il calcolo degli score*

Le scale cliniche sono alla base dell'informazione raccolta da questo studio. Spesso queste scale si presentano come questionari a risposta multipla, in cui alla risposta è legato un valore numerico, generalmente intero e positivo, che contribuisce allo score totalizzato nella scala (solitamente una somma).

L'esistenza di regole definite con cui associare il punteggio totale alle risposte fornite per le singole domande ha permesso di rappresentare la maggior parte degli score mediante campi calcolati in modo automatico da *REDCap*. Questo ha una serie di implicazioni sulla qualità dei dati inseriti. Prima di tutto, all'utente che inserisce i dati non è richiesto né di effettuare il calcolo né di riportare il punteggio, il che rappresenta un risparmio di tempo e di fatica e riduce il rischio di riportare valori errati. Inoltre, implementando una logica "*if*" nel calcolo richiesto a *REDCap* è possibile impedire che nel campo vengano salvate informazioni fino a quando non sono compilati tutti i quesiti della scala. Questo

impedisce per *design* che vengano salvati punteggi parziali, che sarebbero dati errati e difficili da verificare.

4.1.2.2 *Gli eventi Status, Timings e Treatments*

Gli eventi presenti nel progetto di *REDCap* sono corrispondenti ai *timing* previsti dal protocollo, fatta eccezione per tre particolari eventi che sono stati inseriti per aggiungere informazioni aggiuntive rilevanti.

L'evento *Status* comprende uno strumento ripetuto chiamato "*Stato paziente e note*", con il quale è possibile inserire al bisogno variazioni dello stato del paziente, data e motivazioni collegate. Una variazione potrebbe essere il passaggio dal trattamento al *follow up*, oppure il *dropout* dallo studio. Avere queste informazioni immediatamente disponibili facilita l'analisi dei dati permettendo ad esempio la corretta valutazione dei dati mancanti di alcuni *record*. Inoltre, permetterà al termine dello studio di indagare eventuali problemi sistematici che portano a un tasso di *dropout* maggiore in alcuni centri clinici rispetto ad altri.

L'evento *Timings* comprende unicamente la scheda "*Timing*" e permette di raccogliere in un unico strumento le date in cui viene effettuato un determinato evento dello studio. Avere queste informazioni nello stesso posto permette anche a vista di osservare eventuali ritardi nel trattamento o nella compilazione dei form e contattare i centri interessati.

L'evento *Treatments* comprende anch'esso uno strumento ripetuto chiamato "*Diario dei Trattamenti*", che raccoglie alcune informazioni contestuali rispetto al trattamento eseguito (come la data, la durata, il tempo di utilizzo dei robot e il numero di pazienti e terapisti presenti nella seduta). Questo è particolarmente utile per valutare situazioni in cui non viene effettuato il numero di sedute previsto dallo studio ma vengono compilati lo stesso i dati di *T1* o di *Follow up*.

Questi tre eventi e i relativi strumenti sono quindi stati di particolare utilità nel monitoraggio dell'inserimento dei dati.

4.1.2.3 Campo “scala eseguita”

In una delle versioni iniziali del progetto *REDCap* era stato inserito, per ogni scala, un campo che indicasse se la scala era stata eseguita, se sì in che data e se no per quale motivazione. Lo scopo di questo campo era quello di permettere di tracciare correttamente situazioni critiche tipiche della pratica clinica, come il rifiuto da parte del paziente di partecipare al trattamento, ed eventuali difficoltà nell'esecuzione della scala (es. test del cammino per 10 metri inapplicabile a un paziente che non deambula).

Un confronto con la controparte medica ha in seguito portato a eliminare questo campo al fine di snellire i *form* e ridurre il carico di lavoro associato alla compilazione. Tuttavia problemi di questo tipo si sono verificati nel corso del progetto, e questo dimostra che sarebbe stato utile mantenere il campo anche nella versione definitiva.

4.1.3 Monitoraggio dell'inserimento dati: sezione di data export, report e stats

La piattaforma stessa di *REDCap* prevede una sezione di statistiche sui dati inseriti, che può mostrare a colpo d'occhio informazioni riassuntive sulla missingness e sulle tendenze centrali dei dati inseriti. Tuttavia, con l'avanzamento dell'arruolamento, i dati inseriti sono diventati troppo numerosi e le statistiche del sito non potevano rispondere ai bisogni del monitoraggio dell'inserimento dati sia per il lungo tempo di esecuzione e sia perché parziali e poco estensibili a *check* più complessi.

Altre funzionalità di questa sezione sono invece state notevolmente utili, in particolare quelle legate all'*export* dei dati. L'esportazione può riguardare l'intero *dataset*, ma può anche basarsi su report user-defined che filtrano i dati sulla base delle condizioni logiche espresse. Grazie a questa possibilità è possibile selezionare i pazienti e le feature di interesse e ridurre notevolmente il numero di dati da elaborare.

4.2 Plat4MedRob

A differenza di *REDCap*, che consente la gestione dei dati mediante schede elettroniche configurabili, *Plat4MedRob* è una piattaforma sviluppata da Eustema (azienda partner del progetto *Fit4MedRob*) appositamente per la raccolta e analisi di dati non strutturati generati durante *StrokeFit4*. La piattaforma è stata disegnata per ricevere e conservare dati da un'ampia gamma di fonti, come sensori *wearable*, macchine robotiche usate in riabilitazione, *software* di monitoraggio e dispositivi mobili. Inoltre essa può importare tutti i dati strutturati memorizzati in *REDCap*. Essa fornisce pertanto l'integrazione di tutti i dati del *trial*, provenienti da tutte le possibili sorgenti, permettendo di avere una visione ampia e completa delle attività riabilitative.

Per la maggior parte dei dati, *Plat4MedRob* sfrutta i formati *HL7-FHIR* e *DICOM*, favorendo l'interoperabilità, la comunicazione e la collaborazione tra vari professionisti sanitari.

La *dashboard* del sito presenta tre diverse sezioni: *Data acquisition*, *Data visualization* e *Comparative analysis*.

Tramite la sezione *Data acquisition* è possibile registrare i dati direttamente dal device selezionato e monitorare il loro processing. In alternativa è possibile caricare direttamente file locali. Un'altra opzione è quella di acquisire i dati

presenti su *REDCap*, andando così ad aggiornare, al bisogno, le sezioni descritte in seguito.

I dati sono poi visualizzabili attraverso la sezione *Data visualization*, che permette di cercare i dati di interesse all'interno della propria categoria utilizzando una serie di filtri specifici per ogni categoria (Tabella 8).

Tabella 8: categorie di ricerca presenti nella sezione *Data visualization*

Categoria	Dato visualizzato
<i>Patients</i>	Informazioni generali su anagrafica, anamnesi e setting riabilitativo del paziente; informazioni sull'evento acuto di ictus
<i>Encounters</i>	Informazioni specifiche raccolte durante un evento, che può essere un <i>timing</i> (in questo caso sono riportati i risultati delle scale somministrate) o una seduta riabilitativa (in questo caso sono riportati parametri specifici del tipo di riabilitazione eseguita)
<i>Observation</i>	Informazioni relative a una determinata osservazione clinica, come ad esempio la localizzazione dell'ictus
<i>EEG (Elettroencefalogramma)</i>	Permette la visualizzazione dell'andamento del segnale elettroencefalografico nel tempo, con possibilità di selezionare il canale di acquisizione
<i>IMU/EMG</i> <i>(Inertial Measurement Unit)</i>	Come sopra, ma per la visualizzazione dei dati delle <i>IMU</i> e dell'elettromiografia

<i>/Elettromiografia)</i>	
<i>Images</i>	<p>Permette di visualizzare i risultati di imaging biomedico, presentati in formato <i>DICOM</i>.</p> <p>Generalmente questa sezione si riferisce a esiti di TAC o risonanza magnetica.</p>

Ogni sezione permette di scaricare i dati presentati nel formato adeguato.

La sezione *comparative analysis* permette infine di definire raggruppamenti di pazienti sulla base dei valori delle diverse variabili e confrontare i risultati ottenuti in determinate scale per determinati *encounter*.

Le sezioni *Patients*, *Encounters* e *Observation* si aggiornano automaticamente sulla base dei dati presenti su *REDCap*.

Capitolo 5 - La dashboard per la Data Quality

Questo capitolo rappresenta una parte molto importante della tesi, poiché descrive la mia attività relativa allo sviluppo e all'utilizzo della *dashboard* per la *data quality*. Accanto alla realizzazione del progetto *REDCap* per la raccolta dei dati, tale attività rappresenta il mio apporto personale al progetto *Fit4MedRob*.

La natura multicentrica del *trial* porta con sé una serie di complicazioni per quanto riguarda la standardizzazione del *data entry*. Un primo problema può essere quello della formazione degli operatori per quanto riguarda la corretta raccolta dei dati e somministrazione delle scale cliniche: per questa ragione, prima dell'inizio dell'arruolamento dei pazienti, sono stati organizzati incontri con professionisti delle diverse discipline coinvolte nel trial al fine di definire al meglio linee guida chiare che facilitassero il corretto utilizzo degli strumenti a disposizione. E' stato quindi predisposto un *repository* che raccoglie i manuali e i protocolli per la somministrazione delle scale di valutazione, nonché video dimostrativi realizzati da fisioterapisti esperti, con l'obiettivo di illustrare le corrette modalità di somministrazione e favorire l'uniformità delle procedure.

Ciononostante, rimane una variabilità tra i centri per quanto riguarda i protocolli di inserimento dati e l'interpretazione del significato dei campi, soprattutto per quanto riguarda quelli a testo libero. Questo può portare all'introduzione di *bias* o alla registrazione di dati non compatibili tra di loro: un esempio che abbiamo riscontrato durante il trial è stata la diversa interpretazione del risultato del test *Ten Meters Walking Test*, in cui alcuni centri registravano la velocità con cui il paziente percorreva i dieci metri e altri il tempo impiegato.

Infine, esiste una componente di *human error* dovuta al tempo limitato che si può dedicare al *data entry* all'interno di una giornata lavorativa, alla poca confidenza

con l'uso della piattaforma e alla stanchezza. Questi fattori possono portare ad errori di battitura o omissioni che vanno a compromettere l'integrità dei dati.

I controlli possibili con *REDCap* non sempre riescono a cogliere tutte queste sfaccettature, quindi per estenderli e per facilitare la visualizzazione dei dati, aspetto necessario per la valutazione manuale, abbiamo creato una *Data Quality Dashboard (DQD)* [20] semplice e interattiva. Il codice è scritto in *Python* e presentato con la libreria *Streamlit*, che permette la creazione di *web applications* eseguibili localmente.

L'utilizzo della *dashboard* permette di intercettare mensilmente irregolarità nei dati inseriti, specialmente riguardo valori mancanti o anonimi, e riferirli al centro clinico interessato.

La Figura 1 illustra l'architettura funzionale della *dashboard* realizzata e il suo *workflow* di utilizzo. In sintesi, i dati raccolti in *REDCap* vengono esportati verso la *dashboard*; vengono quindi effettuati i controlli che individuano i valori mancanti e i dati che meritano attenzione (es. valori fuori *range*); queste segnalazioni vengono raccolte in una serie di file excel, uno per ogni centro clinico. Ogni file viene inviato tramite mail ai *data entry* in ogni centro, i quali dovrebbero provvedere alla correzione dei dati su *REDCap*. Il processo viene iterato ogni mese. Le sezioni successive dettagliano ogni passaggio.

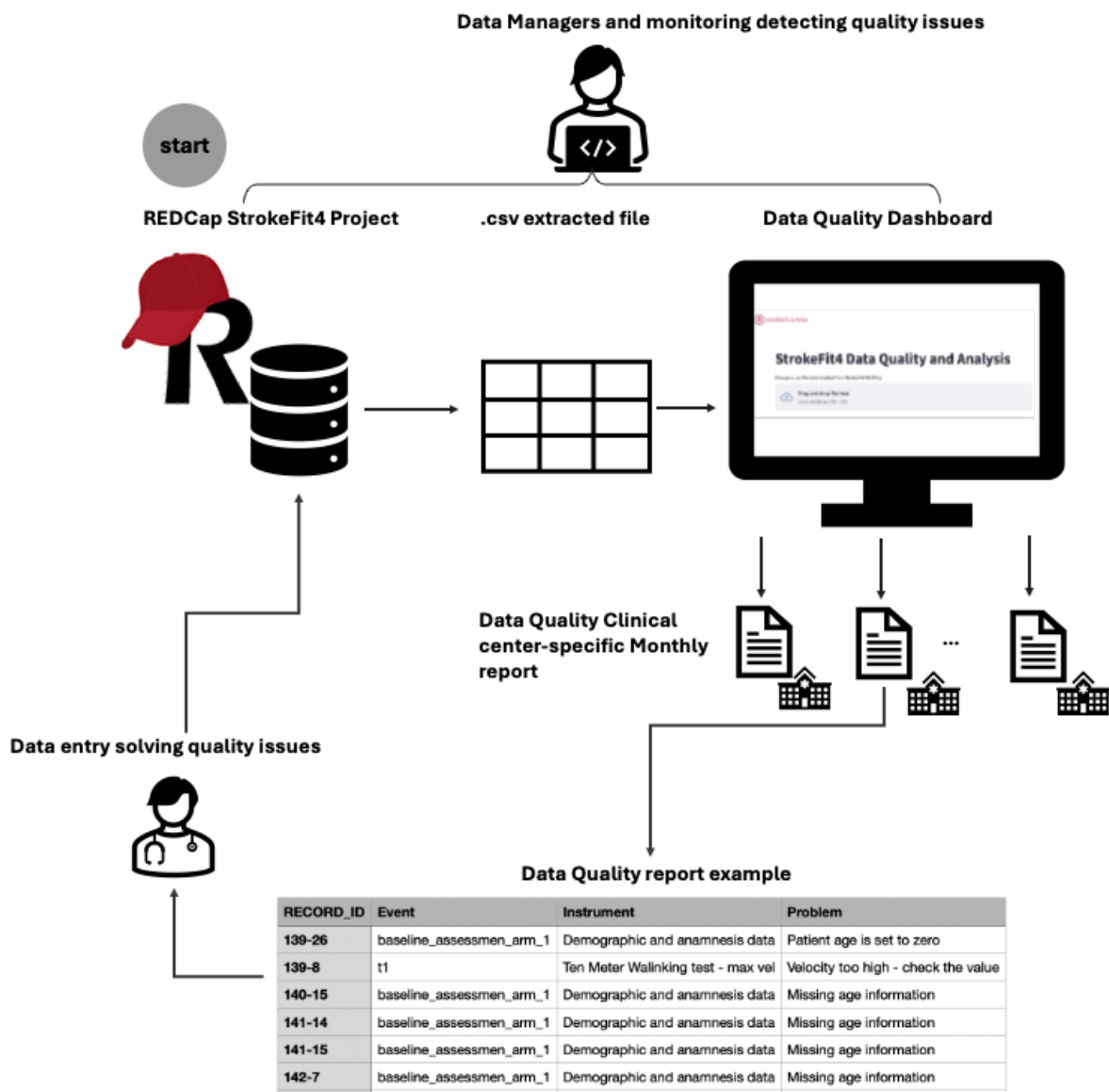


Figura 1: Architettura funzionale e workflow di utilizzo della dashboard per la data quality

5.1 Controlli automatizzati

Come già accennato, il processo di utilizzo della *dashboard* parte da *REDCap* con l'esportazione dei dati in formato .csv.

La presenza di eventi e di strumenti ripetuti complica il processo di *data export*: per poter descrivere correttamente ogni sezione compilata, la piattaforma genera un file molto ridondante e apparentemente pieno di valori mancanti. Questo è dovuto al fatto che per esportare un unico file la tabella generata ha come chiave l'insieme di 4 variabili: il *record id* (rappresentativo del paziente), l'evento, lo strumento ripetuto e l'istanza dello strumento ripetuto. Questi ultimi campi ovviamente risultano compilati solo per gli strumenti. Le colonne della tabella sono invece tutte le feature presenti nel progetto *REDCap*.

Data una riga quindi, le colonne che risultano compilate sono solo quelle relative agli strumenti raccolti per il paziente i-esimo durante l'evento j-esimo. Per gli strumenti ripetuti, grazie alle altre due specificazioni, ogni riga si riferisce alla k-esima compilazione dello strumento specificato.

Tutto il codice sviluppato per i controlli di qualità dei dati tiene quindi conto della struttura di esportazione dei dati. Attualmente il codice è disponibile su richiesta.

5.1.1 Missing values detection

Il primo controllo da eseguire sui dati è quello della presenza. L'analisi della *missingness* però non si può limitare al conteggio dei dati assenti: deve tenere conto di una ragionevole aspettativa di compilazione.

I motivi per cui un dato può mancare sono molteplici e non tutti collegati a un errore da parte del *data entry*. Ad esempio, un paziente che ha appena iniziato il

trattamento sicuramente non avrà i dati di *follow up*, perché temporalmente non è ancora avvenuto: non è ragionevole aspettarsi che i dati siano presenti, quindi non sarebbe corretto contarli come *missing*. Se questa osservazione può essere considerata banale, tutt'altro che banale è la programmazione del controllo dei vincoli temporali in un processo, come quello clinico, in cui gli eventi non cadono tutti esattamente nei tempi previsti.

Attraverso *Python* sono state formalizzate distanze temporali basate sulle tempistiche previste dal protocollo entro le quali è accettabile che il dato sia mancante, considerando sia il tempo necessario alla raccolta del dato sia quello necessario alla sua digitalizzazione, che potrebbe avvenire in un secondo momento.

Un altro motivo per cui il dato potrebbe mancare è dovuto all'inapplicabilità del quesito al paziente. Questo problema riguarda principalmente le scale, per cui in quelle che potevano esserne interessate è stato inserito un campo per indicare il motivo della mancata compilazione, in modo da non includerle nel conteggio dei dati mancanti.

5.1.2 Validity check

La maggior parte dei *check* di validità dei dati è stata implementata in *REDCap* utilizzando i metodi presentati nel paragrafo 4.1.1. Questo, insieme al fatto che la maggior parte dei punteggi è calcolata in modo automatico, ha reso fondamentalmente superflui i controlli di *range* sulla maggior parte degli strumenti.

Tuttavia, un controllo costante dei dati inseriti ha fatto emergere la necessità di effettuare controlli di validità non solo nei confronti del tipo di dato e del range attendibile, ma anche per quanto riguarda il protocollo di studio stesso. Questo tipo di controlli ha riguardato principalmente l'espressione del consenso

informato e il rispetto dei criteri di inclusione ed esclusione. Se tali criteri non sono osservati, di fatto avviene un errato arruolamento di un paziente che in realtà sarebbe non eligibile. Questo è successo ad esempio per le soglie definite per la valutazione cognitiva.

Infine, è stato necessario aggiungere un controllo sul tipo di dato per un campo particolare, ovvero l'esito del test del *Ten Meters Walking Test*. Per questo campo per un errore di progettazione era assente la validazione in *REDCap* e questo ha permesso agli utenti di inserire valori non validi che hanno complicato il processo di analisi. Alcuni utenti hanno inserito del testo libero, immediatamente identificabile come errato, mentre altri hanno usato la virgola come separatore dei decimali. A causa di questo *Python* interpretava l'intera colonna come stringa. I valori errati sono stati corretti a mano.

5.1.3 Outlier detection

Per ogni campo abbiamo poi inserito un controllo di detezione degli *outlier*, considerati come valori esterni all'intervallo definito dal 5-esimo e dal 95-esimo percentile dei dati.

Grazie a questo controllo è stato possibile trovare e segnalare alcuni errori legati all'interpretazione dei campi da parte dei clinici. Un esempio particolare riguarda il campo della data di nascita, che alcuni hanno confuso con il campo della data di arruolamento, portando ad avere nel *database* età registrate sotto all'anno di vita (improbabile per la patologia in esame e comunque incompatibile con i criteri di inclusione dello studio). Un altro esempio degno di nota è stato quello di valori esemplarmente alti nel test del cammino in 10 metri, dovuti alla registrazione del tempo impiegato invece che della velocità.

Complementarmente all'inserimento di questo controllo è stato importante mantenere una comunicazione con i centri clinici in modo da chiarire dubbi sul

significato del termine e dell'operazione a loro richiesta, ovvero la verifica della correttezza del dato. Abbiamo mantenuto un file apposito per conservare i dati confermati, in modo da evitare che contribuiscano al conteggio degli errori.

5.1.4 Controlli automatizzati sulla Plat4MedRob

I controlli di validità sui dati presenti in *Plat4MedRob* sono stati aggiunti recentemente, in quanto la piattaforma è stata resa disponibile all'utilizzo intorno all'aprile di quest'anno e l'inserimento effettivo dei dati sulla piattaforma è iniziato da pochi mesi. Per questo motivo e per la natura non strutturata dei dati sono attualmente limitati al controllo della presenza e hanno un *workflow* separato da quello di normale utilizzo della *dashboard*.

Dalla piattaforma vengono scaricati csv riguardanti le sezioni *Patients*, EEG e IMU/EMG, il cui contenuto è descritto nella sezione 4.2. Da *Patients* si evincono tutti i pazienti attualmente presenti su *REDCap*, e si controlla che ciascuno di loro abbia dei record corrispondenti nelle altre sezioni sia al T0 che al T1. Non è attualmente possibile automatizzare il controllo di presenza delle bioimmagini in quanto la sezione Images non permette di scaricare i dati in formato aggregato, e scaricare individualmente ogni immagine sarebbe più dispendioso che controllare a mano.

5.2 *Workflow* di monitoraggio della qualità dei dati

La presenza di un'unità di controllo centralizzata all'interno del *trial StrokeFit4* permette una possibilità unica, ovvero quella di influenzare la qualità dei dati direttamente durante la loro raccolta. perché questo funzioni in modo virtuoso è necessario prioritizzare la comunicazione e la collaborazione tra il lato informatico e il lato clinico. In accordo con i centri partecipanti si è deciso di mantenere una comunicazione sotto forma di report mensili, in modo da evitare di sovraccaricare gli operatori che si occupano di *data entry*.

In modo da facilitare il più possibile la lettura degli errori/dati mancanti (da qui in poi parleremo di errori intendo errori/dati mancanti) e la correzione degli stessi, il *report* è organizzato in 4 colonne che riportano l'id del paziente, l'evento e lo strumento per i quali è rilevato l'errore e una breve descrizione dell'errore stesso. Vi è poi una colonna dedicata ai commenti, in modo che una volta fatte le modifiche richieste l'operatore possa specificare di aver visto ed effettuato la correzione, oppure possa esprimere il motivo per cui la correzione non è possibile (ad esempio, un dato che è stato individuato come fuori range potrebbe essere un outlier clinicamente corretto, o un dato mancante che risale a un tempo troppo distante e non è più misurabile). La Figura 2 riporta un esempio di report: le prime 4 colonne rappresentano il report inviato, mentre la quinta colonna riporta la risposta dell'operatore del centro.

Record id	Event	Instrument	Problem	Commenti Data Entry
145-20	timing_arm_1	Timing	Missing date information for t6_m	rifiuta il colloquio
145-21	timing_arm_1	Timing	Missing date information for t6_m	si sta cercando di avere riscontro
145-22	T1	TAM - intentions score	Outlier	Giusto
145-22	T1	TAM - compatibility score	Outlier	Giusto
145-22	T1	Symbol Digit Modality Test	Missing Symbol Digit Modality Test	sistemato, è non eseguibile
145-23	timing_arm_1	Timing	Missing date information for t3_m	rifiuta il colloquio
145-23	T_5S	Box and Block Test - Not affected side	Outlier	Giusto
145-25	timing_arm_1	Timing	Missing date information for t2_m	rifiuta il colloquio
145-25	timing_arm_1	Timing	Missing date information for t1_m	rifiuta il colloquio
145-25	T1	DN4 Index	Outlier	Giusto
145-25	treatments_ar	Diario dei trattamenti	Missing treatments (24/25 inseriti)	il paziente ha rifiutato l'ultimo trattamento, lo status è stato aggiornato.
145-25	t0_arm_1	Modified Barthel Index	MBI al T0 maggiore di 95	c'è stato un refusto, abbiamo sistemato.
145-26	T0	Computer Skills B score	Outlier	Giusto
145-26	T1	Symbol Digit Modality Test	Outlier	Giusto
145-26	T0	Symbol Digit Modality Test	Outlier	Giusto
145-26	treatments_ar	Diario dei trattamenti	Missing treatments (21/25 inseriti)	dimissione anticipata, status aggiornato
145-26	timing_arm_1	Timing	Missing date information for t1_m	si sta cercando di avere riscontro

Figura 2: *data quality report* con risposta da parte del centro clinico

Il flusso che porta alla realizzazione mensile dei report parte dall'esportazione di tutti i dati presenti nel *REDCap* e nel caricamento di questo file nella *Data Quality Dashboard*. In questa fase è necessario dedicare un momento all'ispezione manuale, sia tramite *dashboard* che tramite *REDCap*: osservare le distribuzioni dei dati permette di notare criticità di cui il controllo non è ancora stato automatizzato. La *dashboard* mostra statistiche generali sulle variabili demografiche, sia tra i vari centri clinici sia selezionando un centro clinico alla

volta. A margine di ognuna di queste statistiche è possibile aggiungere manualmente un problema rilevato da aggiungere al *report* specificando il paziente a cui l'errore è riferito. Al termine della *dashboard* è anche presente una sezione dalla quale aggiungere errori generali, specificando sempre paziente evento e strumento a cui si riferiscono.

Al termine di questa operazione è possibile esportare il *report*, sia per tutti i centri che per un centro alla volta. Generalmente esportiamo il *report* generale e poi lo suddividiamo per centri. Per facilitare la lettura, le righe relative allo stesso paziente vengono evidenziate con lo stesso colore, come mostrato nella Figura 2.

A questo punto i centri clinici hanno un mese di tempo, ovvero fino all'invio del prossimo *report*, per chiedere chiarimenti e rispondere alla nostra sollecitazione commentando opportunamente le righe del documento. Le risposte che confermano la risoluzione o non presenza degli errori vengono salvate in un documento in modo da evitare di ripetere la notifica di falsi errori (questo ha principalmente luogo nelle correzioni degli *outlier*).

5.3 Risultati

Grazie all'implementazione dei metodi sopra descritti e al controllo manuale siamo riusciti a controllare molte delle dimensioni espresse nel Paragrafo 2.2. Quanto fatto in particolare rientra nelle definizioni di controlli di *element presence*, *validity check*, *log review* e *structural agreement*.

Alcune delle metodologie proposte risultavano inapplicabili allo studio, come l'utilizzo dei *gold standard*, e altre non sono state possibili in modo approfondito con gli strumenti usati, come il *data element agreement*. Quest'ultimo verrà ripreso nel Paragrafo 5.4 come limitazione dei controlli eseguiti e nel Capitolo 6 ne verrà proposta un'implementazione.

Nel complesso possiamo vedere un riassunto delle dimensioni controllate nella Tabella 9.

Tabella 9: metodi di controllo applicati alle dimensioni di qualità

Dimensione	Metodi applicati
Completeness	Element presence
Correctness	Validity check, Log review
Concordance	Nessuno
Plausibility	Validity check
Currency	Log review
Bias	Element presence, Log review
Conformance	Element presence, Structural agreement

Definendo l'errore come una segnalazione qualsiasi generata dai metodi descritti nel Paragrafo 5.1, abbiamo ottenuto una misura del trend dell'errore. A causa del fatto che il *trial* non è ancora concluso non è possibile mostrare risultati numerici precisi in valore assoluto; come mostrato in Figura 3 (dove volutamente sono stati tolti i valori sulle ordinate), possiamo però riferire che, esprimendo l'errore in percentuale rispetto al numero di controlli effettuati, abbiamo notato, dall'inizio dell'implementazione del processo di *data quality*, una diminuzione di circa il 2%. Anche se, come meglio illustrato nella sezione successiva, non possiamo affermare con certezza che questo risultato sia esclusivamente dovuto al nostro intervento, i tipi di errori trovati e la successiva interazione con il personale dei vari centri ci ha dato la percezione dell'utilità del processo stesso. Qui di seguito mostriamo alcuni risultati in termini relativi, che possiamo quindi riportare anche prima della fine del trial.

Nella Figura 4 possiamo osservare la composizione mensile dei report in termini di tipologia di errore.

Un fattore degno di nota è che nel tempo gli errori totali sono diminuiti nonostante l'aumento del numero di controlli, e questo è indice sia della correzione degli errori precedenti che possibilmente di un aumento sia della familiarità con gli strumenti che della chiarezza del processo di *data entry*.

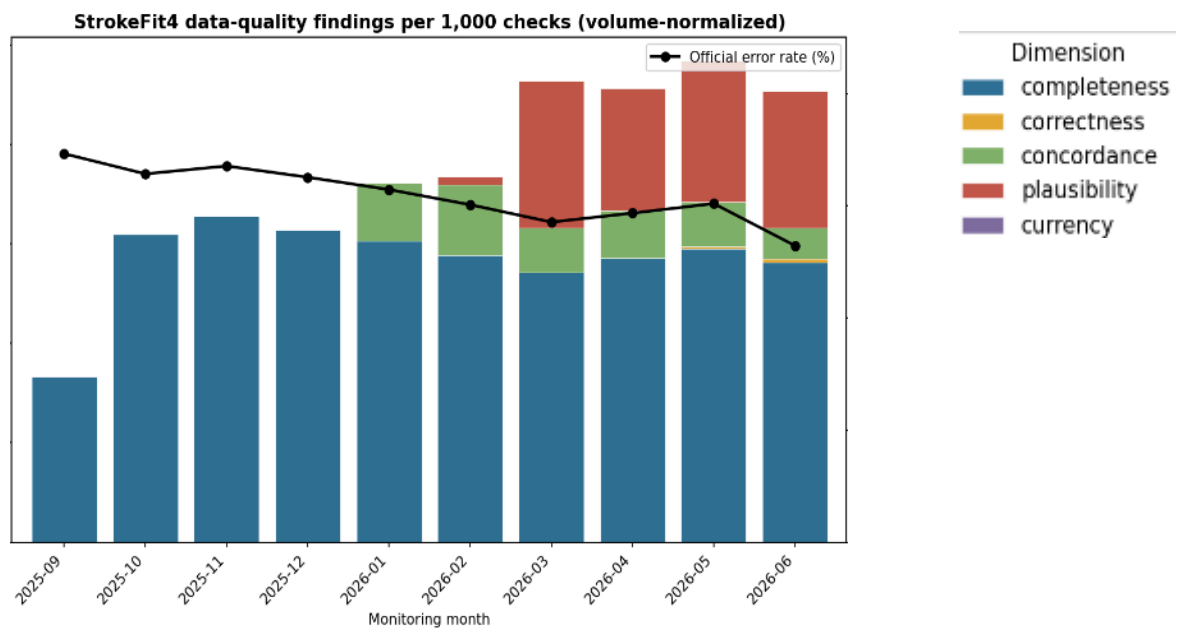


Figura 3: trend degli errori per mese, divisi per tipo

La *concordance* riportata nella figura si riferisce a un unico tipo di controllo, ovvero alla presenza di dati in scale segnalate come non effettuate.

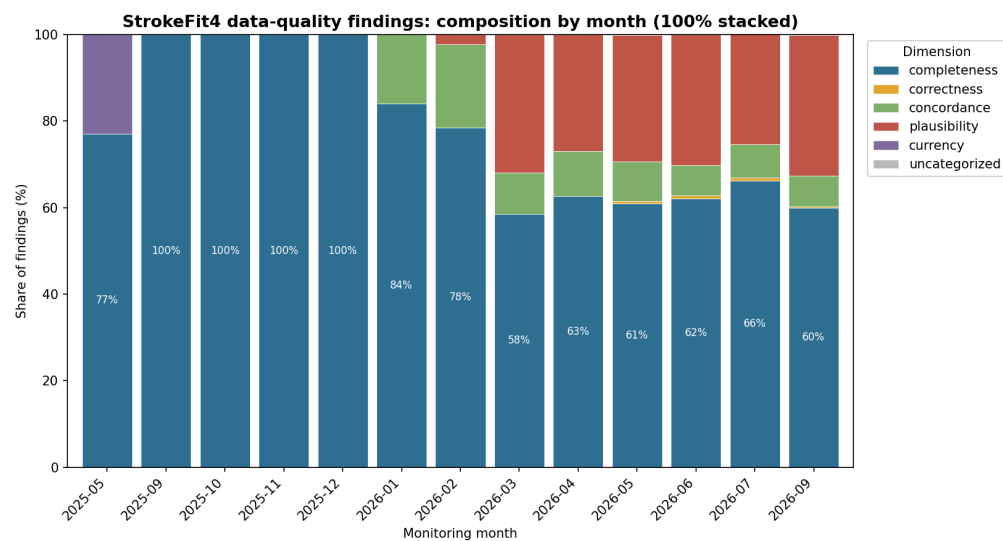


Figura 4: composizione percentuale degli errori mese per mese

Nella Figura 5 invece vediamo, in riferimento all’ultimo mese per cui è stato inviato un report (settembre 2026), la composizione percentuale degli errori trovati con i vari metodi per ognuno dei centri, che sono riportati sulle righe.

Tali composizioni mostrano diversità anche molto evidenti, ad esempio il decimo centro presenta molti probabili errori dovuti a valori fuori range, mentre il centro 21 mostra una grossa percentuale di errori di “data element agreement”.

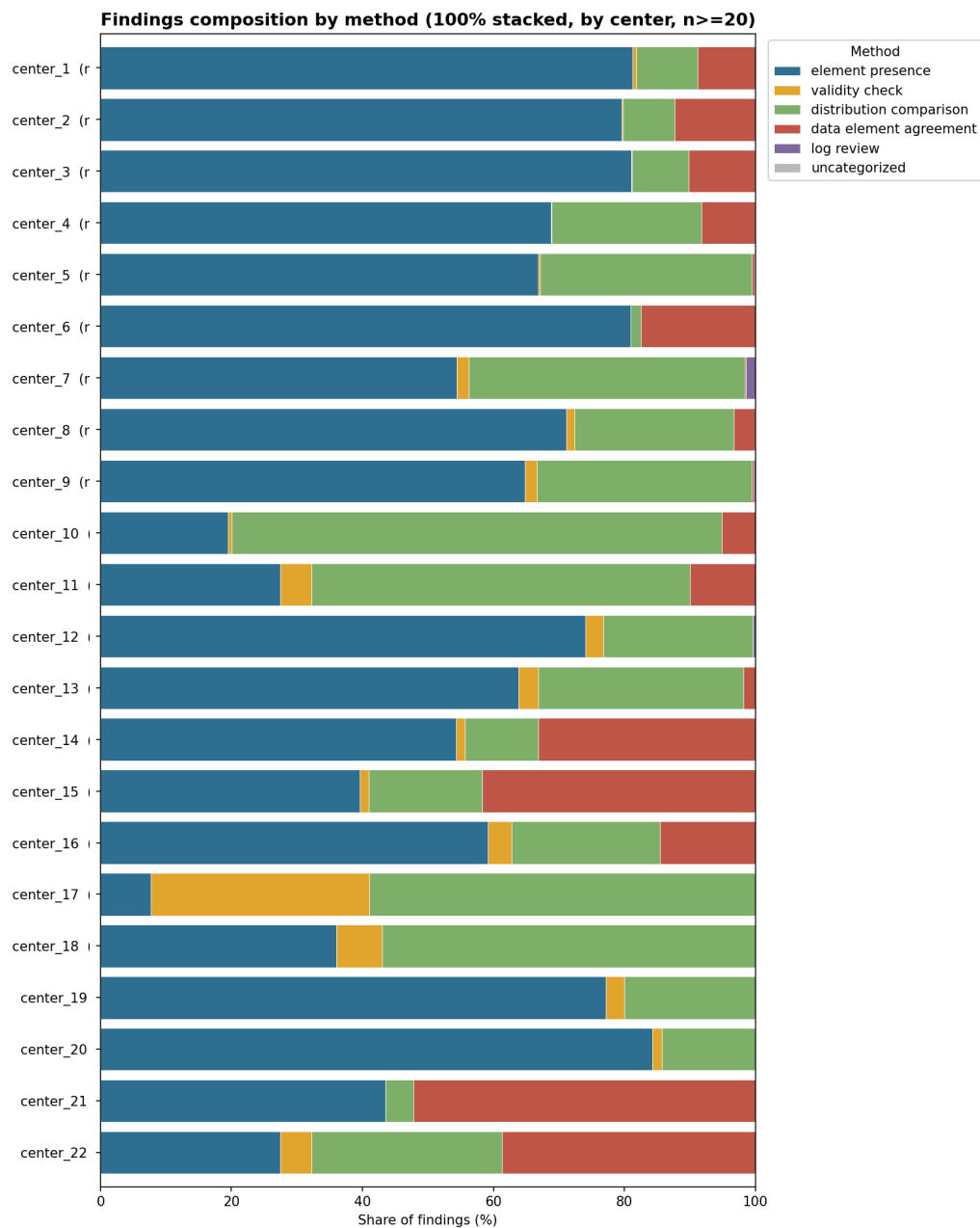


Figura 5: composizione degli errori per centro (una riga per ogni centro)

5.4 Limitazioni nel processo di *data quality*

La buona riuscita del monitoraggio della *data quality* è profondamente legata all'interazione con i *data entry*. La disponibilità richiesta è bidirezionale: come è importante che il gruppo di *data management* si renda disponibile a fornire chiarimenti e assistenza al bisogno, la controparte clinica dovrebbe rendersi

disponibile a verificare le correzioni richieste e fornire un *feedback* che può essere spunto di riflessione per un ampliamento o una migliore definizione dei controlli.

Purtroppo non tutti i centri hanno risposto come ci aspettavamo. All'ultimo *report* mandato (settembre 2026) 6 centri (35% dei centri contattati) hanno risposto restituendo le correzioni effettuate, uno (6%) ha solo confermato la ricezione e i restanti 10 (59%) non hanno mandato nessuna risposta. Alcuni centri non hanno pazienti registrati in *REDCap* e per loro i *report* non vengono generati.

Le motivazioni sono ancora da indagare. Potrebbe esserci, da parte di qualche *data entry*, una bassa percezione dell'importanza della qualità dei dati, ma potrebbero anche esserci cause di tipo organizzativo, ad esempio una carenza di personale dedicato al *data entry*. Inoltre, se tale attività viene delegata esclusivamente al personale clinico, è probabile che il tempo a disposizione per svolgere un compito giudicato "accessorio" sia insufficiente a garantire una buona qualità del dato. E' importante in fase di impostazione del *trial* riconoscere una certa autorevolezza al *team* che si occupa di gestire i dati e definire, lato ospedaliero, dei responsabili dell'inserimento dati con cui mantenere una comunicazione attiva. Nel caso di mancata risposta, va definito un canale di *escalation* per assicurarsi la massima collaborazione.

D'altro canto, questo aspetto ha sottolineato l'importanza di adottare un approccio di *quality by design*, quindi di prevedere nel momento dello sviluppo i possibili errori, limitarli il più possibile ed effettuare molti test prima di mettere il progetto in produzione.

Un'altra limitazione che abbiamo incontrato è stato il fatto che abbiamo iniziato a mandare sistematicamente i report da settembre 2025, quindi per il primo anno di arruolamento il corretto e completo inserimento dei dati era lasciato alla buona volontà degli operatori sanitari e a controlli manuali sporadici. A causa di questo

ci sono alcuni dati mancanti, in particolare relativi a scale non somministrate o misure non effettuate, che non sono più recuperabili. In futuro potrebbe essere più efficace monitorare l'inserimento dei dati direttamente dall'inizio dell'arruolamento.

Vi è poi un aspetto di possibile approfondimento, che verrà meglio descritto nel capitolo 6 e riguarda l'implementazione di controlli di *concordance*. Esistono relazioni tra le variabili che non sono immediatamente note osservando i dati raccolti, ma che si possono evincere dai metadati: l'esempio più semplice è quello dell'utilizzo delle scale cliniche per valutare lo stato del paziente. Scale con punteggi e significati diversi potrebbero essere usate per descrivere lo stesso distretto funzionale, e ci si può quindi aspettare che diano esiti coerenti. Questo aspetto non è facilmente considerabile usando approcci statistici che si basano su controlli dove i dati sono considerati indipendentemente uno dall'altro, ma potrebbe essere affrontato applicando ai dati un'ontologia derivante dalle informazioni presenti in letteratura e formando una *knowledge base*, su cui si potrebbe far ragionare un modello di intelligenza artificiale.

Un'ultima limitazione riguarda la dimostrazione di efficacia della *dashboard* di *data quality*. Grazie al nostro lavoro abbiamo potuto osservare qualitativamente un miglioramento nelle operazioni di registrazione dei dati. Quanto fatto però non è sufficiente a dimostrare l'efficacia dell'approccio adottato per una serie di motivi, primo tra tutti l'assenza di un gruppo di controllo. Occorrerebbe effettuare uno studio in cui ogni centro viene randomizzato a ricevere o a non ricevere il *report* di *data quality* e quindi verificare se il gruppo di intervento produce un *dataset* di qualità migliore. Inoltre è difficile valutare l'andamento dell'errore presente in termini assoluti, perché nel tempo sono aumentati i controlli effettuati, ma anche i pazienti su cui i controlli venivano fatti (a causa dell'avanzamento dell'arruolamento) e il numero centri partecipanti (alcuni centri si sono aggiunti allo studio grazie a Bandi A Cascata).

Capitolo 6 - Controllo di qualità knowledge-based: una proposta di ontologia per le scale cliniche

Le scale previste dal protocollo, come visto, sono molto numerose e talora ridondanti, nel senso che in alcuni casi misurano gli stessi domini. Questo offre una possibilità di effettuare un'ulteriore analisi della qualità del dato: il controllo che i risultati delle scale analoghe siano concordi. In generale l'automatizzazione di questo problema non è scontata:

- i punteggi ottenuti sono molto eterogenei tra loro, spesso non sono espressi sulla stessa scala (es. *range* diversi come 0-10, 0-100), unità di misura e talvolta neanche nello stesso tipo di dato;
- i *cutoff* utilizzati nella pratica clinica per classificare la gravità della patologia rilevata non suddividono sempre lo score nello stesso numero di classi, portando alcune scale ad avere una granularità maggiore di altre.

Queste problematiche rendono particolarmente difficile un confronto diretto tra le scale, ma potrebbero essere superate utilizzando un'ontologia.

6.1 Le ontologie: definizione e applicazioni

Un'ontologia è una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di interesse [21].

Le ontologie vengono comunemente utilizzate per risolvere problemi relativi alla comunicazione tra gli individui e all'interoperabilità tra i sistemi: lo stesso concetto potrebbe essere rappresentato in modi diversi a seconda del contesto (ad esempio per motivi di gergo), e questo porta a una perdita di informazione.

Gli elementi fondamentali che compongono un'ontologia sono:

- *Classe*: rappresentazione del concetto (es. Paziente)
 - *Individuo*: istanza concreta della classe, che ne eredita tutte le proprietà e gli attributi (es. il sig. Mario Rossi)

- *Proprietà*: caratteristiche intrinseche della classe (es. nome, cognome, ...)
- *Restrizioni*: definiscono e delimitano il tipo di dato che la proprietà può assumere (es. nome → stringa di testo; peso (Kg) → numero [0-500])
- *Relazioni*: definiscono le relazioni tra le classi (es. relazione “is a”)

Un'ontologia, incluse le istanze delle sue classi, forma una base di conoscenza, o *knowledge base*. La rappresentazione in una knowledge base è orientata non tanto alla singola istanza, ma alla descrizione di entità astratte e delle regole e dei vincoli che le legano. Questo cambio di paradigma rispetto alla tradizionale rappresentazione in *database* permette di sfruttare *software* detti *reasoner* per fare inferenza logica e dedurre nuova conoscenza da quella inizialmente formalizzata.

6.2 StrokeFit4 ontology

Una prima rappresentazione ontologica della struttura delle scale cliniche è quella proposta da Sacchi e Nicora [22], dove si utilizza come caso di studio proprio il progetto *StrokeFit4*.

L'obiettivo di questa ontologia è la rappresentazione di scale cliniche validate e delle loro caratteristiche al fine di sostenere una scelta informata delle scale da somministrare nel contesto di un protocollo riabilitativo, con particolare attenzione sulla completezza nella descrizione dello stato del paziente, sulla correttezza della somministrazione delle scale e sulla standardizzazione degli outcome misurati in diversi setting riabilitativi.

Per la rappresentazione dell'ontologia è stato usato Protégé, che sfrutta il linguaggio owl [23].

La struttura dell'ontologia, mostrata in Figura 6, si basa su sette tassonomie, ciascuna delle quali si coniuga in classi più specifiche:

1. *Measurement Goal*: rappresenta l'obiettivo clinico che viene valutato attraverso la scala. L'obiettivo può essere ad esempio la valutazione delle funzioni motorie, così come dello stato psicologico, della qualità di vita o del rischio;
2. *Information Need*: rappresenta lo specifico quesito clinico per cui viene applicata la scala. Alcuni esempi sono "Ansia", "Forza muscolare", "Dolore" o "Disabilità". L'*information need* è l'ambito misurato all'interno del contesto definito dal *measurement goal*;
3. *Measure*: rappresenta la scala clinica utilizzata, aggiungendo anche meta-informazioni come la lingua di somministrazione o il tempo di esecuzione previsto;
4. *Person*: specifica chi somministra una determinata scala. Generalmente questo viene fatto da un operatore sanitario, ma alcune scale potrebbero essere somministrate da un caregiver o compilate direttamente dal paziente (cosiddette "scale di autosomministrazione");
5. *Device*: esprime gli strumenti che servono per somministrare la scala.
6. *Medical Condition*: descrive la patologia target per la quale viene applicata una certa scala;
7. *Measured Value*: identifica il tipo di dato che risulta dalla scala, che può essere in generale numerico, categorico o ordinale.

Mediante le *Object Properties* è possibile esprimere importanti relazioni tra le classi descritte. Alcuni esempi sono i seguenti:

- *Measure targets MedicalCondition*: lega la scala alla patologia (o alle patologie) per cui viene usata.
- *Measure quantifies InformationNeed*: lega la scala al (o agli) *Information Need* di interesse (un *information need* potrebbe ad es., comprendere se il paziente è autonomo nelle ADL).
- *MeasurementGoal generates_need InformationNeed*: lega il contesto di valutazione alle caratteristiche indagate.

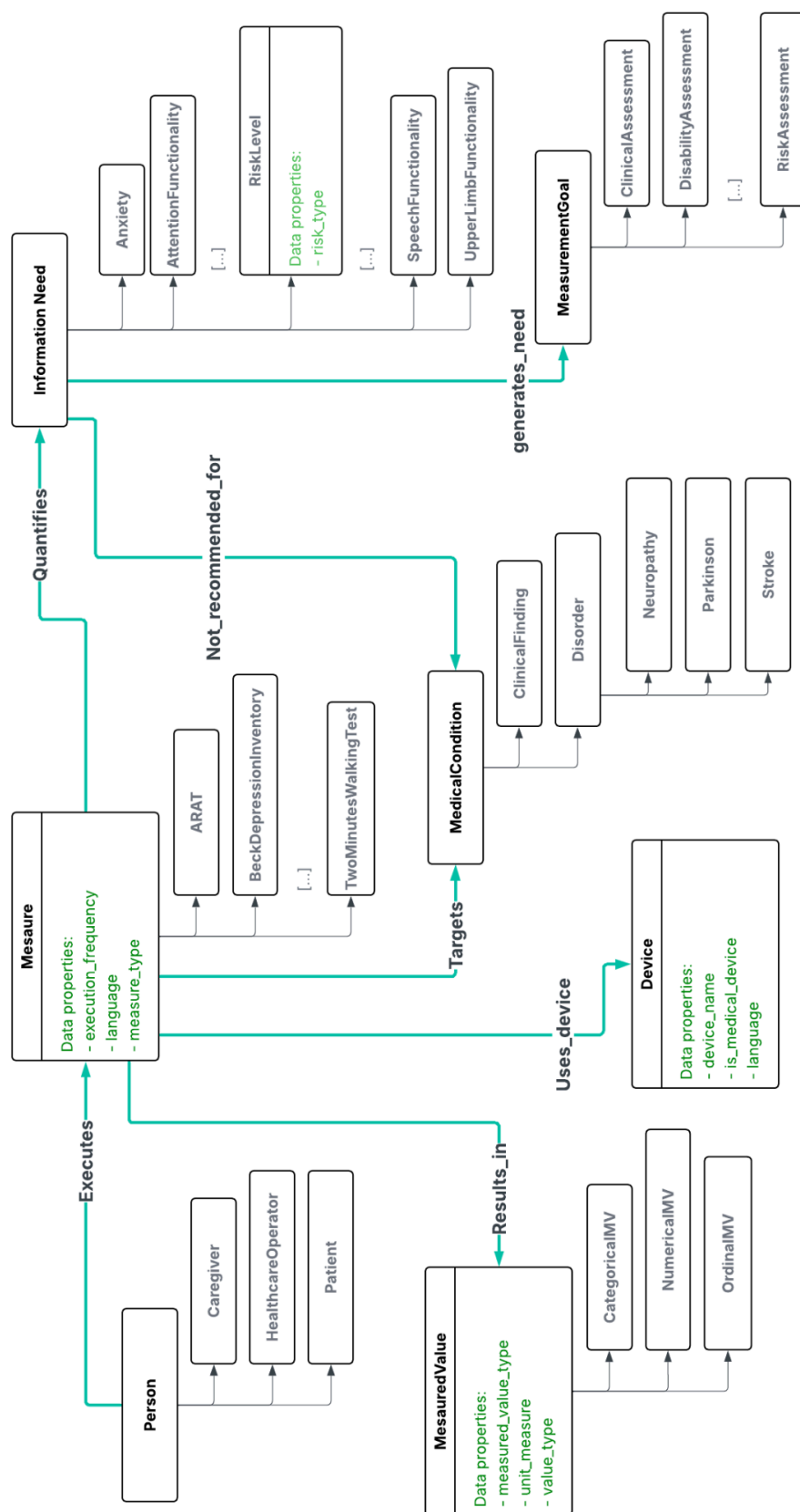


Figura 6: La struttura dell'ontologia delle scale di misura

Questa formalizzazione evidenzia la centralità della scala come strumento, descritta finemente attraverso le sue caratteristiche specifiche, i suoi obiettivi clinici e i suoi contesti di utilizzo. Tuttavia, se ci fermassimo a questo livello di rappresentazione, non sarebbe possibile utilizzare l'ontologia per valutare la coerenza tra le misure ottenute, in quanto un'istanza di tipo *scala* non rappresenta l'applicazione di una determinata scala a un paziente ma il concetto astratto della scala stessa, e risulta quindi utile nella definizione di un protocollo riabilitativo ma di difficile applicazione a un contesto di *data quality assessment*.

6.3 Adattamento dell'ontologia al problema di qualità del dato

Basandomi sulla struttura precedentemente descritta ho sviluppato un'ontologia con il fine di confrontare i risultati ottenuti per un dato paziente nelle diverse scale, ponendo al centro della modellizzazione il concetto di punteggio.

Rispetto alla versione precedente ho mantenuto le classi *Information Need*, *Measure* e *Person*, e ho aggiunto le seguenti classi:

- *Score*: un raffinamento della classe *Measured Value*, che va per ogni scala a rappresentare esattamente punteggi e sottopunteggi definendo il tipo di dato e il range di valori ammissibili.
- *Timestamp*: questa classe serve per descrivere il protocollo riabilitativo, associando a ciascun tempo di rilevazione le scale raccolte (classe *Measure*).
- *Severity*: con questa classe vengono rappresentati dei livelli di gravità descritti su una scala ordinale (dal meno grave al più grave). L'obiettivo di questa classe è uniformare i punteggi di ciascuna scala in un esito confrontabile con le altre.

La classe *MeasurementGoal* è stata mantenuta a livello di contenuto informativo facendola diventare una proprietà di *InformationNeed*.

Per capire meglio il funzionamento possiamo basarci sull'inserimento di istanze e seguire il percorso di un generico paziente che possiamo chiamare Mario Rossi. *MarioRossi* è un individuo che fa parte della classe *Person*, e in particolare della sottoclasse *Patient*. In quanto elemento di questa classe, egli eredita la relazione “*is_subject_to*”, che lo lega a una o più istanze della classe *Timestamp*.

Un elemento della classe *Timestamp* è caratterizzato da una data e da una relazione chiamata “*collects*”, che viene meglio specificata nelle sottoclassi e collega ogni timing del protocollo alle scale rilevate. Possiamo creare l'individuo *Baseline_MR*, che andrà a collegare le informazioni del paziente e della data alle scale effettuate. Una rappresentazione grafica di questo è raffigurata nella Figura 7.

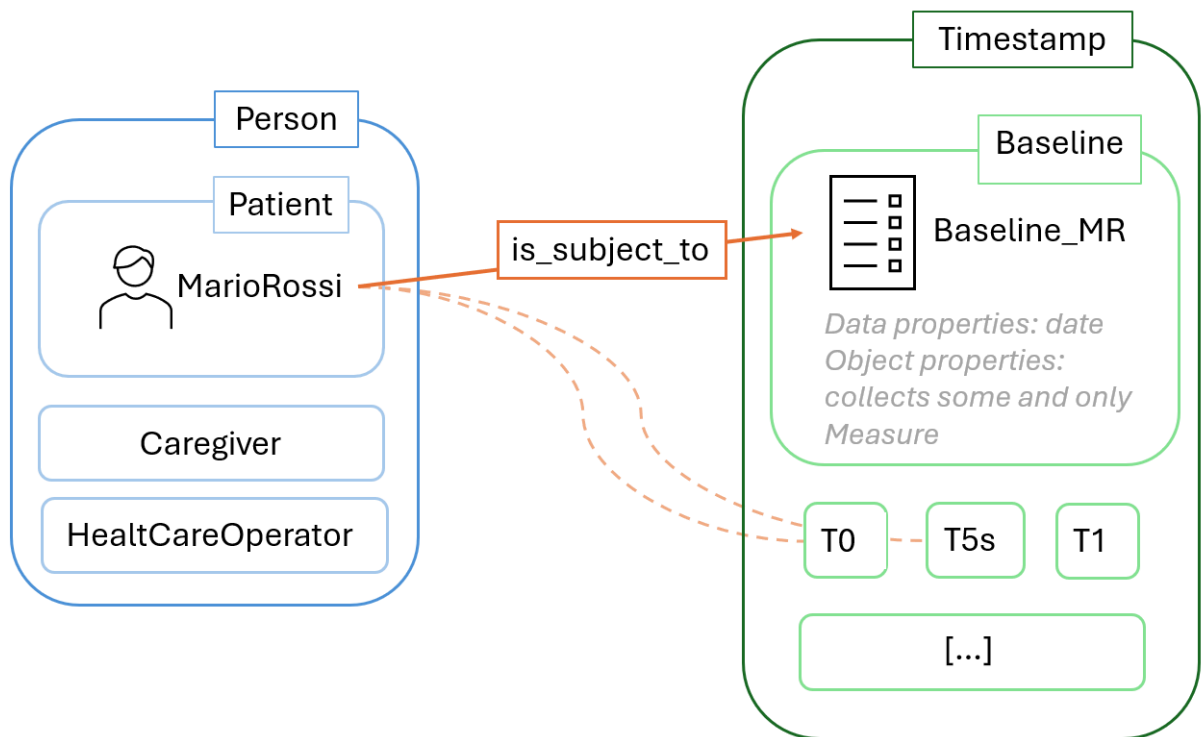


Figura 7: esempio di inserimento dell'individuo *MarioRossi* (istanza della classe *Patient*, che è sottoclasse di *Person*) che ha una relazione di tipo “*is_subject_to*” verso l'individuo *Baseline_MR* (istanza della classe *Baseline*, che è sottoclasse di *Timestamp*)

Essendo *Baseline_MR* individuo della sottoclasse *Baseline*, esso eredita una proprietà *collects* verso ognuna delle scale che vengono raccolte durante il baseline assessment, più chiaramente esplicitate nel capitolo 3.1.4.

La classe *Measure* è caratterizzata da due principali proprietà:

- Quantifica, con la relazione *quantifies*, un qualche *InformationNeed*
- Ha esito, mediante la relazione *results_in*, in un qualche *Score*

Lo *Score* ha come proprietà un qualche punteggio, che può essere numerico, categorico o ordinale, e una classificazione della gravità mediante la relazione *classified_as Severity*. La classe *Score* si dirama in tante sottoclassi quante sono le scale valutate: ciascuna sottoclasse, specifica per la propria scala, va a definire esattamente quale tipo di punteggio viene raccolto (o quali, nel caso di sottoscale). Osserviamo l'esempio della scala *HADS* (*Hospital Anxiety and Depression Scale*) nella Figura 8.

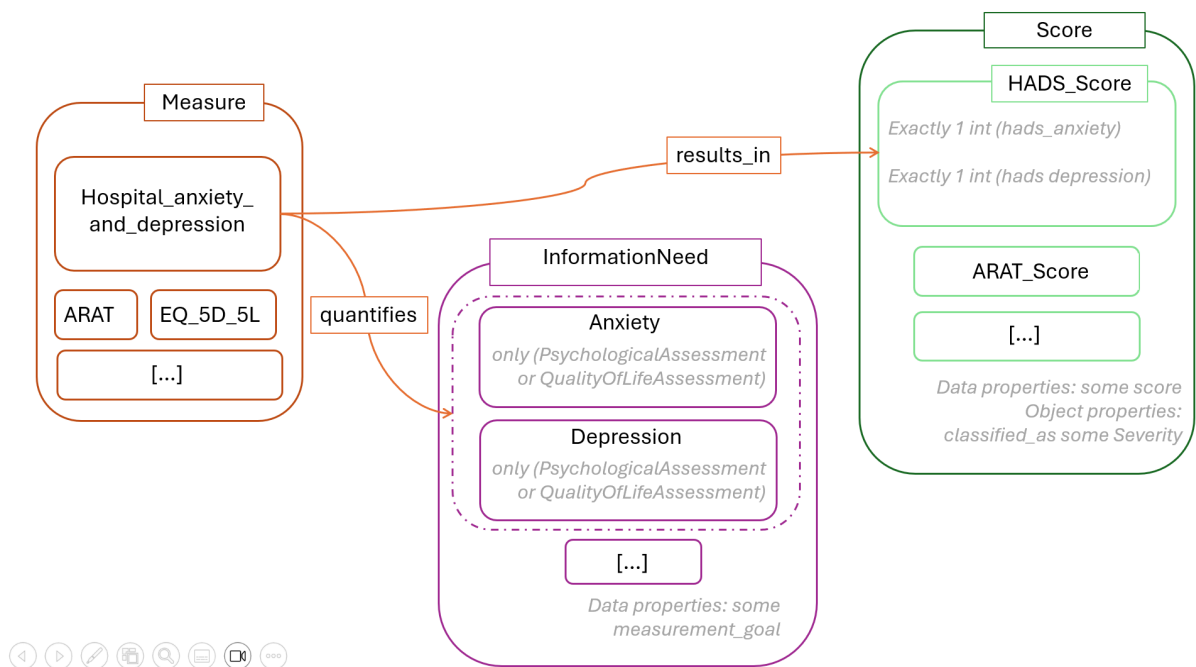


Figura 8: Rappresentazione delle relazioni associate a una possibile istanza della classe *Hospital_anxiety_and_depression* nei confronti di altre classi

Ogni *score* è collegato a una (o più, nel caso di sottopunteggi) *data property* specifica per la scala che definisce il *range* di valori che il dato può assumere.

Il vero passo in avanti sta nello scomporre ogni *score* in una serie di sottoclassi definite basate sui *cutoff* presenti in letteratura: in questo modo, una volta assegnato un valore al punteggio della scala, il *reasoner* sarà in grado di inferire da sé il livello di gravità associato.

Un esempio di questo è stato fatto per lo *score* del *modified Barthel Index* ed è riportato in Figura 9.

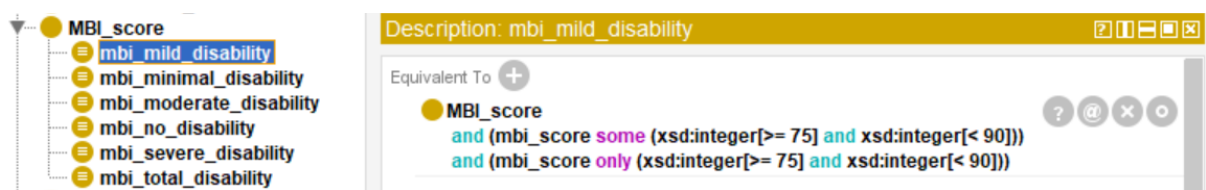


Figura 9: estratto dell'interfaccia di Protégé relativo alle sottoclassi della classe MBI_score

Grazie a quanto fatto finora è stato possibile arrivare a determinare per un certo quesito clinico (rappresentato dall'*Information Need*), valutato in un certo paziente in un certo istante di tempo, un livello categorico di rischio che viene associato direttamente dalla macchina in risposta a un certo punteggio totalizzato nella scala. I rischi di scale diverse tuttavia non sono ancora confrontabili tra di loro, il che rende necessario mapparli in una classe *Severity*.

All'interno della classe *Severity* ci saranno diverse fasce di gravità associate a un valore intero ordinale che permetta di interpretare l'esito ottenuto in relazione a una sorta di *ranking*. La definizione specifica di questa classe non è un problema facile, in quanto non solo alcune scale rilevano la gravità più finemente di altre (es. rischio alto/basso vs rischio grave/moderato/lieve/assente), ma non è detto che livelli di gravità associati alla stessa fascia abbiano significati clinicamente corrispondenti (es. il punteggio massimo del *modified Barthel Index* è ottenuto quando si ha totale disabilità, mentre il punteggio massimo della *modified Rankin*

Scale, anch'essa spesso usata per la valutazione del grado di disabilità, corrisponde al decesso del paziente). Per definire completamente questa classe è necessario l'affiancamento di un professionista sanitario formato sulle scale di interesse.

Un'iniziale prova del funzionamento di questa logica è stata fatta in relazione allo score del *modified Barthel Index*, ed è emerso che grazie alle regole logiche imposte il *reasoner* è in grado, dato uno score ottenuto in una valutazione, di dedurre sia la fascia di rischio relativa alla scala sia il punteggio di *Severity* corrispondente. La struttura finale dell'ontologia è rappresentata nella Figura 10.

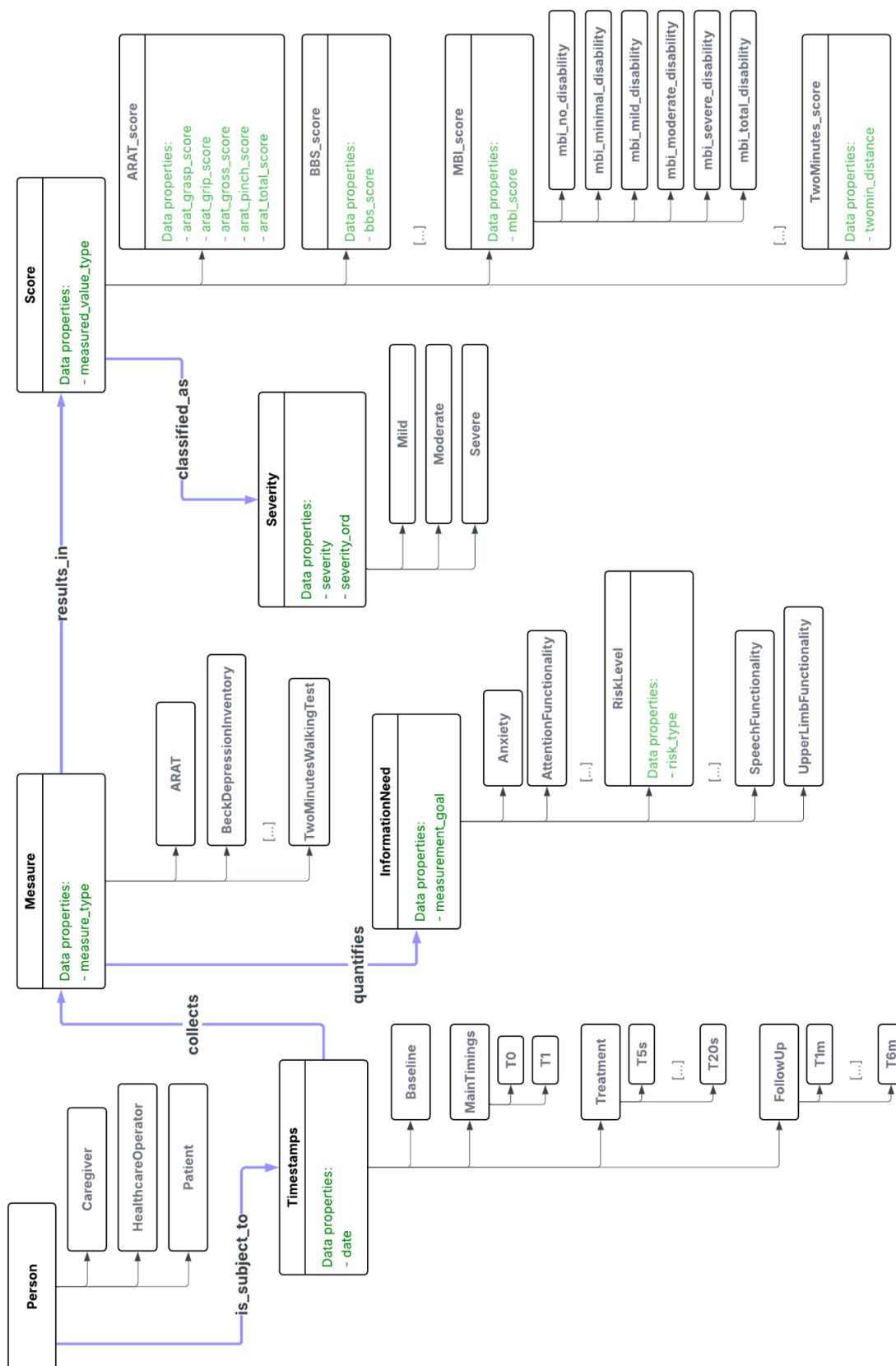


Figura 10: rappresentazione dell'ontologia proposta

6.4 Rilevanza per l'analisi dei dati

L'utilizzo di un'ontologia nell'analisi dei dati clinici permette di integrare al semplice dato conoscenze relative al contesto, come il dominio di applicazione di una scala, e alla letteratura medica di riferimento, come la gravità del rischio misurato. Il *reasoner* sarebbe in grado in modo automatico di inferire informazioni aggiuntive sulla base dei dati forniti in ingresso ed evidenziare situazioni degne di nota, come una non corrispondenza tra gli esiti delle scale.

La corretta definizione della classe *Severity* permetterebbe non solo di mappare i vari rischi rilevati sulla stessa scala, ma di avere un'informazione sulla distanza tra due risultati discordanti: distanze di una sola classe (es tra “*mild*” e “*moderate*”) sono sicuramente meno allarmanti rispetto a distanze di varie classi (es. “*mild*” e “*severe*”).

Una volta completa, questa ontologia potrà applicarsi a diversi tipi di analisi:

- *Qualità dei dati*: confronto tra scale relative allo stesso *InformationNeed* per lo stesso paziente nello stesso *timestamp* per valutare la coerenza dei risultati di scale diverse.
- *Monitoraggio di un paziente*: confronto dei risultati di una stessa scala per uno stesso paziente nel tempo per valutare l'efficacia del trattamento nel recupero di una specifica funzione.
- *Confronto tra pazienti*: arricchendo la descrizione dei pazienti con informazioni demografiche, valutare se esistono *cluster* di pazienti per i quali la riabilitazione risulta più o meno efficace.

Inoltre, il lavoro di formalizzazione delle scale svolto non sarebbe limitato al trial *StrokeFit4*, ma estendibile ad altri studi (andando ad aggiungere eventuali nuove scale e a definire dei *timestamp* specifici del protocollo).

Capitolo 7 - Conclusioni e sviluppi futuri

L'adozione di controlli periodici di qualità dei dati richiede uno sforzo che è assolutamente ragionevole e può avere conseguenze molto positive sui dati raccolti, da una parte assistendo il personale di *data entry* nel cogliere errori commessi e dall'altra riducendo le operazioni di pulizia dei dati necessarie nelle fasi preliminari dell'analisi. Non è però scontato che le varie figure coinvolte nei processi di raccolta dati colgano l'importanza di queste operazioni, quindi serve che la *data quality* sia un'intenzione principale e condivisa già a partire dal protocollo dello studio.

Gli aspetti fondamentali da considerare sono sicuramente: la qualità come obiettivo di design, con principi chiari e comuni a tutte le persone che collaborano allo sviluppo degli strumenti di raccolta dati; la formazione, sia sul significato dei dati medici per il lato informatico sia sull'importanza dei controlli per il lato clinico; e infine l'impegno ad una costante comunicazione tra le varie parti. I processi di controllo di qualità dei dati non devono essere percepiti come un aumento del carico lavorativo, ma come un mutuo supporto nel raggiungimento degli *endpoint* dello studio.

Un modo efficace di implementarli in uno studio è rappresentato da un *workflow* di tipo *prevent - detect - notify - correct - verify*. La prevenzione è da considerarsi in ambito di sviluppo, ed è l'attenzione nell'evitare ambiguità e migliorare l'usabilità degli strumenti, rendendo chiaro il contenuto richiesto e impedendo il più possibile gli errori legati all'interpretazione del campo o alla battitura. *Detect, notify, correct, verify* è riguarda invece trovare criticità, notificarle ai centri in modo che le possano correggere e verificare le risposte ottenute. E' importante che questo ciclo avvenga in un tempo ragionevole: se il tempo è troppo lungo si rischia che alcuni dati diventino irrecuperabili, mentre se è troppo breve può risultare pesante per gli operatori clinici

L'applicazione di strumenti come le ontologie può assistere nella ricerca delle anomalie e rendere più semplice i controlli trasversali a molte variabili, permettendo di considerare anche il significato clinico e semantico dei dati e l'esperienza medica come parte integrante dell'informazione.

In futuro l'efficacia delle pratiche finora descritte potrà essere valutata con uno studio ad hoc, confrontando il modello proposto in questa tesi con altri possibili modelli e con un gruppo di controllo che non riceverà correzioni. Questo permetterà di stabilire un modello lavorativo ideale che permetta alle varie parti coinvolte di lavorare in sinergia verso il miglioramento dei trial clinici.

Infine, sono lieta di riferire che il nostro lavoro sulla *data quality* [24] nel trial *StrokeFit4* è stato accettato ad un *workshop* della *Royal Statistical Society* di Londra ("*Data Quality and Statistics: Perfect together!*") organizzato da Rosemary Tate, ricercatrice di grande esperienza nel settore *Data Quality* [25][26][27].

Bibliografia

- [1] L. Houston, Y. Probst, P. Yu, e A. Martin, «Exploring Data Quality Management within Clinical Trials», *Appl. Clin. Inform.*, vol. 9, fasc. 1, pp. 72–81, gen. 2018, doi: 10.1055/s-0037-1621702.
- [2] J. M. Juran e F. M. Gryna, *Juran's Quality Control Handbook*. McGraw-Hill, 1988.
- [3] A. D. Gibson, N. M. White, G. S. Collins, e A. G. Barnett, «Evidence of unreliable data and poor data provenance in clinical prediction model research and clinical practice», *BMC Med.*, vol. 24, fasc. 1, p. 386, giu. 2026, doi: 10.1186/s12916-026-04981-y.
- [4] «Springer Nature retracts, removes nearly 40 publications», The Transmitter: Neuroscience News and Perspectives. Consultato: 14 ottobre 2026. [Online]. Disponibile su: <https://www.thetransmitter.org/retraction/exclusive-springer-nature-retracts-removes-nearly-40-publications-that-trained-neural-networks-on-bonkers-dataset/>
- [5] «FAIR», GO FAIR Foundation. Consultato: 15 ottobre 2026. [Online]. Disponibile su: <https://www.gofair.foundation/fair-principles>
- [6] Y. Ren *et al.*, «Missing data were poorly reported and handled in randomized controlled trials with repeatedly measured continuous outcomes: a cross-sectional survey», *J. Clin. Epidemiol.*, vol. 148, pp. 27–38, ago. 2022, doi: 10.1016/j.jclinepi.2022.04.019.
- [7] G. Lanzola, E. Parimbelli, G. Micieli, A. Cavallini, e S. Quaglini, «Data quality and completeness in a web stroke registry as the basis for data and process mining», *J. Healthc. Eng.*, vol. 5, fasc. 2, pp. 163–184, 2014, doi: 10.1260/2040-2295.5.2.163.
- [8] C. Lavieri *et al.*, «Database completeness and evaluation of potential barriers and facilitators of data collection in the national pediatric cancer registry in Ethiopia», *J. Cancer Policy*, vol. 48, p. 100729, giu. 2026, doi: 10.1016/j.jcpo.2026.100729.
- [9] A. C. Cheng *et al.*, «Multisite evaluation of automated electronic case report form data entry from electronic health records», *J. Biomed. Inform.*, vol. 180, p. 105060, ago. 2026, doi: 10.1016/j.jbi.2026.105060.
- [10] <https://www.lifescience.net>, «INGENIO-DQ: Data Quality Assessment for a Precision Medicine Data Model Using an International Common Framework.», Life Science Network. Consultato: 15 ottobre 2026. [Online]. Disponibile su: <https://www.lifescience.net>
- [11] B. Ansari *et al.*, «Data Quality Assurance Tool for the Acute to Chronic Pain Signatures Study (A2CPS): An Interactive R Shiny Application», *medRxiv*, p. 2026.01.07.26343620, gen. 2026, doi: 10.64898/2026.01.07.26343620.
- [12] N. G. Weiskopf e C. Weng, «Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research», *J. Am.*

- Med. Inform. Assoc.*, vol. 20, fasc. 1, pp. 144–151, gen. 2013, doi: 10.1136/amiajnl-2011-000681.
- [13] «Electronic health record data quality assessment and tools: a systematic review | Journal of the American Medical Informatics Association | Oxford Academic». Consultato: 12 ottobre 2026. [Online]. Disponibile su: <https://academic.oup.com/jamia/article/30/10/1730/7216383?login=true>
 - [14] «Vision». Consultato: 12 ottobre 2026. [Online]. Disponibile su: <https://www.fit4medrob.it/about/vision>
 - [15] Fondazione Don Carlo Gnocchi ETS, «Rehabilitation With and Without Robot and Allied Digital Technologies (RADTs) in Stroke Patients: a Pragmatic Multicenter Randomised Controlled Trial on the Effectiveness, Acceptability, Usability, and Economic-organizational Sustainability of RADTs From Subacute to Chronic Phase», clinicaltrials.gov, Clinical trial registration NCT06547827, lug. 2026. Consultato: 12 ottobre 2026. [Online]. Disponibile su: <https://clinicaltrials.gov/study/NCT06547827>
 - [16] I. G. Aprile *et al.*, «Rehabilitation with and Without Robot and Allied Digital Technologies (RADTs) in Stroke Patients: A Study Protocol for a Multicentre Randomised Controlled Trial on the Effectiveness, Acceptability, Usability, and Economic-Organisational Sustainability of RADTs from Subacute to Chronic Phase (STROKEFIT4)», *J. Clin. Med.*, vol. 14, fasc. 8, p. 2692, apr. 2025, doi: 10.3390/jcm14082692.
 - [17] S. Shah, F. Vanclay, e B. Cooper, «Improving the sensitivity of the Barthel Index for stroke rehabilitation», *J. Clin. Epidemiol.*, vol. 42, fasc. 8, pp. 703–709, 1989, doi: 10.1016/0895-4356(89)90065-6.
 - [18] S. Campagnini *et al.*, «Cross-validation of predictive models for functional recovery after post-stroke rehabilitation», *J. NeuroEngineering Rehabil.*, vol. 19, p. 96, set. 2022, doi: 10.1186/s12984-022-01075-7.
 - [19] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, e J. G. Conde, «Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support», *J. Biomed. Inform.*, vol. 42, fasc. 2, pp. 377–381, apr. 2009, doi: 10.1016/j.jbi.2008.08.010.
 - [20] Nicora G, Sacchi L, Tibollo V, Cigliutti E, Caretti M, Falchini F, Mauro MC, Fasano A, Siotto M, Pavan A, Cortellini L, Lattanzi S, Bellazzi R, Aprile IG, Germanotta M, Quaglini S., «Digitizing rehabilitation outcomes and assessing data quality in clinical trials: implementing validated scales in REDCap for a stroke RCT», *BMC Med. Res. Methodol.*, vol. 26, fasc. 1, p. 81, mar. 2026, doi: 10.1186/s12874-026-02801-8.
 - [21] T. R. Gruber, «A translation approach to portable ontology specifications», *Knowl. Acquis.*, vol. 5, fasc. 2, pp. 199–220, giu. 1993, doi: 10.1006/knac.1993.1008.
 - [22] L. Sacchi, G. Nicora, I. G. Aprile, e S. Quaglini, «Clinical Outcome Measurement Scales: A Domain Ontology and a Use Case for Stroke Rehabilitation», in *Artificial Intelligence in Medicine*, R. Bellazzi, J. M.

- Juarez Herrero, L. Sacchi, e B. Zupan, A c. di, Cham: Springer Nature Switzerland, 2025, pp. 334–338. doi: 10.1007/978-3-031-95841-0_62.
- [23] M. A. Musen, «The protégé project: a look back and a look forward», *AI Matters*, vol. 1, fasc. 4, pp. 4–12, giu. 2015, doi: 10.1145/2757001.2757003.
- [24] Nicora G., Aprile I, Fasano A et al. Prospective Data Quality Assurance in Multicenter Clinical Trials: Longitudinal Evidence from a Rehabilitation RCT, Data Quality and Statistics: Perfect together!, Royal Statistical Society, Thursday 22 October 2026, Bayes Centre, Edinburgh
- [25] Tate, A. Rosemary, Harnessing Pattern Recognition Techniques for Data Quality Detection, Studies in Health Technology and Informatics Conference, Paper Open Access2026 DOI: 10.3233/SHTI260246
- [26] Tate AR, Dungey S, Glew S, Beloff N, Williams R, Williams T. Quality of recording of diabetes in the UK: how does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study using the CPRD database. *BMJ Open*. 2017 Jan 25;7(1):e012905
- [27] Dungey S, Glew S, Heyes B, Macleod J, Tate AR. Exploring practical approaches to maximising data quality in electronic healthcare records in the primary care setting and associated benefits. Report of panel-led discussion held at SAPC in July 2014. *Prim Health Care Res Dev*. 2016 Sep;17(5):448-52