DOKTORI (PhD) ÉRTEKEZÉS

Vámosiné Pacza Tünde

DEBRECENI EGYETEM

Debrecen, 2025

DEBRECENI EGYETEM

TÁPLÁLKOZÁS - ÉS ÉLELMISZERTUDOMÁNYI DOKTORI ISKOLA

Doktori Iskola vezető:
Prof. Dr. Szilvássy Zoltán
egyetemi tanár, az MTA doktora

Témavezető:
Dr. Baranyi József, PhD
tudományos tanácsadó

AZ ANYATEJ MOLEKULÁRIS ÖSSZETÉTELÉNEK DINAMIKUS MODELLEZÉSE A KÖRNYEZETI TÉNYEZŐK FÜGGVÉNYÉBEN

Készítette:

Vámosiné Pacza Tünde

doktorjelölt

Debrecen
2025

AZ ANYATEJ MOLEKULÁRIS ÖSSZETÉTELÉNEK

DINAMIKUS MODELLEZÉSE

A KÖRNYEZETI TÉNYEZŐK FÜGGVÉNYÉBEN

Értekezés a doktori (PhD) fokozat megszerzése érdekében
az élelmiszertudományok tudományágban

Írta: Vámosiné Pacza Tünde okleveles matematikus

Készült a Debreceni Egyetem
Táplálkozás- és Élelmiszertudományi Doktori Iskolája
(Élelmiszertudományi programja) keretében

Témavezető: Dr. Baranyi József, PhD

Az értekezés bírálói:

……

……

A bírálóbizottság:

elnök: ……

tagok: ……

……

……

……

Az értekezés védésének időpontja: 20…. …… … .

TARTALOMJEGYZÉK

1. BEVEZETÉS

A táplálkozás egészségre gyakorolt hatását már több ezer éve felismerték, azonban az eddig létrehozott tápanyag adatbázisok nem tartalmazzák és elemzik az élelmiszerekben található molekulák jelentős részét. Az élelmiszerek összetételéről jelenleg rendelkezésre álló átfogó adatbázisok csak az élelmiszereinkben található összes kémiai anyag töredékét fedik le, az egészségünk szempontjából lényeges tápanyag-összetevőkre összpontosítva, és több ezer más molekula, melyek közül pedig soknak jól dokumentált egészségügyi hatásai vannak, nem követhető nyomon *(Hooton és mtsai., 2020)*. Mindezen túl a már feltárt összetevőket eddig főként leíró módszerekkel jellemezték, és az olyan megfigyelések, mint például a fokhagymának a szív- és érrendszeri betegségek megelőzésére gyakorolt pozitív hatása, nélkülözték a mechanisztikus, biokémiai magyarázatokat *(Barabási és mtsai., 2020)*. A bizonytalanság fő forrásai (i) a többezer kémiai kölcsönhatás által okozott komplexitás; (ii) a mérések és megfigyelések inherens hibái; (iii) illetve számos egyéb részlet volt *(Barabási és mtsai., 2020)*. Ahhoz, hogy ezeket a bizonytalanságokat jobban megértsük, és mélyebben feltárhassuk táplálkozásunknak az egészségre gyakorolt hatását, elengedhetetlen az összetevők által meghatározott komplex rendszer átfogóbb megismerése, melyhez egy jól szervezett adatbázis óriási segítséget nyújthat.

Az egyik legfontosabb és legösszetettebb táplálék - mellyel először találkozunk emberi életünk során - az anyatej. Az anyatej nélkülözhetetlen az újszülött növekedéséhez és fejlődéséhez közvetlenül a születés után, és pótolhatatlan táplálékforrás a csecsemő túléléséhez *(Rossum és mtsai., 2005; Agostoni és mtsai., 2009)*. Egyedülálló forrása különféle bioaktív komponenseknek, amelyeket minden anya „személyre szabottan" állít össze, hogy kielégítse a fejlődő csecsemő szükségleteit *(Christian és mtsai., 2021; Samuel és mtsai., 2020)*. Emiatt a WHO és az ENSZ Gyermekalapja kizárólagos anyatejes táplálást ajánl a születés után legalább 6 hónapig, és az anyatejes táplálás folytatását javasolja legalább 2 éves korig *(WHO, 2003)*.

Számos tanulmány foglalkozik az anyatej táplálkozási összetevőinek meghatározásával *(Kim & Yi, 2020; Picciano, 2001)*, de annak egyéntől és időtől való függésére még nincs általánosan elfogadott elmélet. Többek között vizsgálták az anyatejben található makró- és mikró tápanyagokat *(Lönnerdal, 2016; Gidrewicz & Fenton, 2014)*, valamint tanulmányozták immunológiai összetevőit. Mindezen túl a

legújabb analitikai technikák (például új generációs szekvenálás) *(Nyquist és mtsai., 2022)* alkalmazásával a komponensek által kifejtett hatásokat is tanulmányozzák. Mindezek ellenére még mindig jelentős a tudáshiány, hiszen nem állnak rendelkezésre megbízható becslések sem a „referencia” és „standard” anyatej összetételre, sem a szoptatás során - az anyai és környezeti tényezők hatására - bekövetkező összetételbeli változásokra *(Casavale és mtsai., 2019)*.

1.1. CÉLKITŰZÉS

Jelen kutatás középpontjában az anyatej molekuláris szintű összetételének feltérképezése és időbeli változásainak matematikai modellezése állt. Ehhez elsődleges cél a tudományos kutatások publikált adataiból egy adatbázis létrehozása volt, majd a létrehozott adatbázis felhasználásával az anyatej összetétel időbeli változásának modellezése.

A disszertáció első része az anyatej biokémiai összetételét tartalmazó, MilkyBase *(Pacza és mtsai., 2022)* nevű adatbázist írja le. Az adatokat tudományos publikációkból részben gépi tanulással, részben manuálisan válogattuk és digitalizáltuk. Elsődleges célunk egy olyan ontológia definiálása volt, amely annak feltárását segíti, hogy az anyatej összetétele hogyan függ különböző tényezőktől. Másodlagos cél, hogy egy ilyen ontológia használható legyen másfajta élelmiszerek kutatásában is, ahol a felhasználók saját adataikat is hasonló formátumban tárolhatják, illetve publikálják.

A disszertáció második része a létrehozott MilkyBase *(Pacza és mtsai., 2022)* adatbázis adatainak felhasználásával bemutatja, hogy az anyatej kutatásban hogyan alkalmazhatóak az élelmiszer-mikrobiológiában már elterjedten használt prediktív modellezés módszerei.

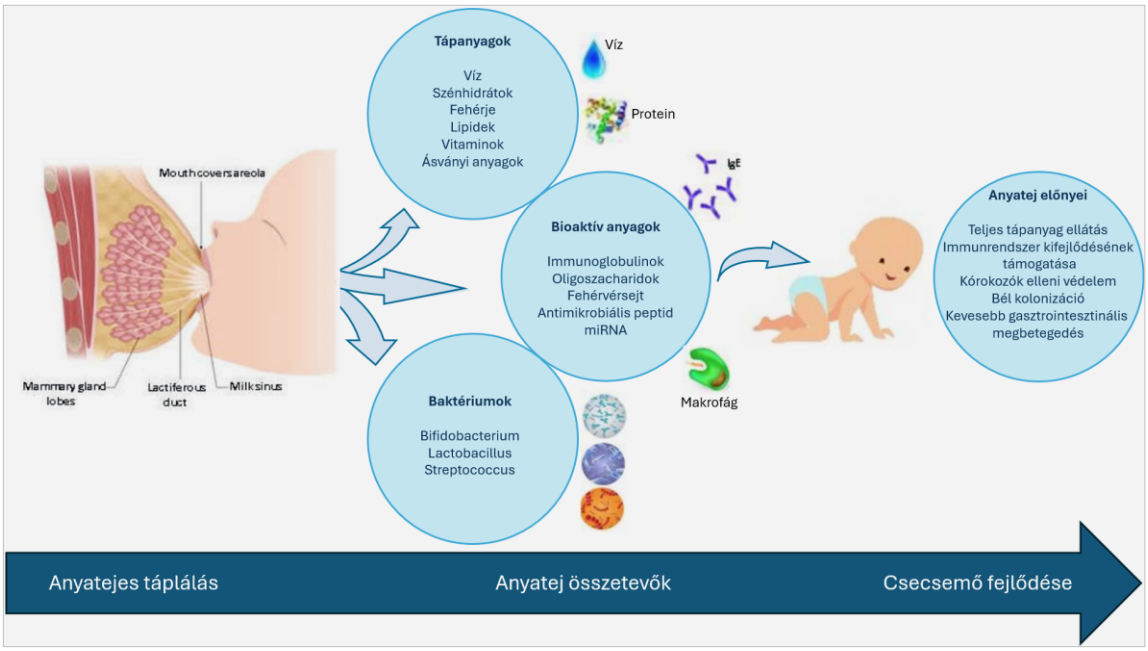
A cél az volt, hogy a modell segítségével pontosabb képet alkothassunk az anyatej komponenseinek időbeli, és egyéntől függő változásáról, valamint az azt leginkább befolyásoló tényezőkről, ezzel is segítve a további anyatej-összetevők megismerését célzó kísérletek megtervezését. A módszer bemutatásán túl a dolgozat utolsó részében rámutatok a "prediktív anyatejkutatás" lehetséges korlátaira és felhasználási lehetőségeire is.

2. IRODALMI ÁTTEKINTÉS

2.1. ANYATEJ

A csecsemők optimális növekedése és fejlődése érdekében kizárólagos szoptatás ajánlott az élet első hat hónapjában *(WHO, 2003)*. Az anyatej az egyetlen olyan táplálék, amely kielégíti a csecsemők összes táplálkozási igényét, optimális alkalmazkodást, normális növekedést, érést és fejlődést biztosít, és mindezeken túl különleges kötődést alakít ki az anya és a csecsemő között, az élet korai szakaszában *(Agostoni és mtsai., 2009; Eidelman és mtsai., 2012; Yi & Kim, 2021; Kramer, 2010)*. Az újszülött egészségére gyakorolt pozitív hatása is széleskörűen bizonyított. Rövid távon az anyatej csökkenti a gyomor- és bélrendszeri problémák *(Victora és mtsai., 1987)*, valamint az alsó légúti fertőzések okozta megbetegedések és a halálozás kockázatát, míg hosszú távon például a magasabb IQ-val és a metabolikus szindrómák megelőzésével hozható összefüggésbe *(Horta, 2019)*. A tápanyagok (szénhidrátok, lipidek, fehérjék, vitaminok és ásványi anyagok) mellett bioaktív komponenseket (hormonok, citokinek, növekedési faktorok, antimikrobiális anyagok, sejtek stb.) is tartalmaz *(Zivkovic és mtsai., 2011; Newburg, 2001)*, amelyek fontos szerepet játszanak a központi idegrendszer, az anyagcsere, az immunrendszer és a mikrobiom fejlődésében *(Gertosio és mtsai., 2016; Carr és mtsai., 2021; Boix-Amorós és mtsai., 2019; Victora és mtsai., 2016; Patro-Gołąb és mtsai., 2016)* (1.ábra.).

A szoptatást pozitív egészségügyi hatásokkal hozták összefüggésbe, beleértve a megnövekedett intelligenciát, a fertőző és a nem fertőző betegségek (elhízás, atópiás betegségek, cukorbetegség, gyulladásos bélbetegségek) kockázatának csökkenését *(Carr és mtsai., 2021; Boix-Amorós és mtsai., 2019)*. Az anyatejnek a csecsemőkori táplálkozásban betöltött döntő szerepe, a hosszú távú egészségre gyakorolt hatása miatt széles körű klinikai *(Gertosio és mtsai., 2016; Carr és mtsai., 2021; Boix-Amorós és mtsai., 2019; WHO, 2003)*, társadalmi és gazdasági érdeklődésre tart számot *(WHO, 2003; Rollins és mtsai., 2016)*.



1.ábra. Az anyatej összetevőinek jótékony hatása a csecsemő fejlődésére

Forrás: *(Lyons és mtsai., 2020)*

Az anyatej a táplálkozási és bioaktív komponensek komplex biológiai rendszere, ahol az összetevők állandó kölcsönhatásban vannak egymással *(Christian és mtsai., 2021)*. Az elmúlt évtizedekben széles körű kutatási erőfeszítések történtek az anyatej összetevőit és az azok jelenlétét, illetve az összetevők mennyiségét befolyásoló tényezők megismerésére *(Ballard & Morrow, 2013; Perrella és mtsai., 2021; Sánchez és mtsai., 2021; Carr és mtsai., 2021; De Weerth és mtsai., 2022; Samuel és mtsai., 2020)*. Ezen erőfeszítések ellenére még mindig hiányzik annak a dinamikus rendszernek a mechanisztikus megértése, amelyet az anya és a gyermek az anyatej révén alkot *(Christian és mtsai., 2021; Shenhav & Azad, 2022)*. Jelenlegi ismereteink az anyatej összetételéről és az összetétel változásáról nagyrészt az összetevőket értékelő olyan vizsgálatokon alapulnak, melyek jellemzően az összetevők változékonyságát és a változás dinamikáját külön-külön elemzik *(Christian és mtsai., 2021; Samuel és mtsai., 2020; Boix-Amorós és mtsai., 2019)*.

Az anyatej-összetétel heterogenitásának okai sokfélék lehetnek. Az összetételre hatással lehet az anyai genetikai állománya és táplálkozása *(Golan és mtsai., 2017; Moltó-Puigmartí és mtsai., 2010; Samuel és mtsai., 2019)* (amelyet a földrajzi elhelyezkedés és az életmód nagymértékben befolyásol), valamint a gyógyszerek vagy

táplálékkiegészítők használata is *(Fischer és mtsai., 2010)*. Továbbá a paritás, a szüléskori terhességi kor *(Léké és mtsai., 2019)*, a szülés és szoptatás módja *(Silva és mtsai., 2021)* és szakaszai, valamint a csecsemő neme és egészségi állapota szintén jelentős faktorok *(Galante és mtsai., 2018)*. Ezen túlmenően a rendelkezésre álló adatokat a tej mintavételével, kezelésével *(Rodríguez-Cruz és mtsai., 2020)*, tárolásával és elemzésével kapcsolatos módszertani tényezők is befolyásolják *(Christian és mtsai., 2021; Samuel és mtsai., 2020)*.

Az anyatej összetétellel kapcsolatos adatok előállítása és gyűjtése, a statisztikai és modellezési módszerek, valamint az eredmények értelmezése ritkán összpontosít az „anya-tej-csecsemő" rendszer dinamikus jellegére. Erre a hiányosságra mutatott rá Shenhav és Azad is alapvető publikációjukban *(Shenhav & Azad, 2022)*, ahol ezt a hiányosságot három fő okra vezették vissza:

(i) Jellemzően csak az egyes elemeket vagy komponenseket vizsgálják, így a köztük lévő kölcsönhatások kimaradnak.

(ii) A vizsgálatok túlnyomórészt keresztmetszeti jellegűek, azaz a tej összetételének pillanatfelvételét mutatják be, nagyrészt figyelmen kívül hagyva az összetevők időbeli változását a laktáció során.

(iii) Hiányoznak a számítási módszerek az anya-tej-gyermek triász és környezete összetett ökoszisztémájának ábrázolására és megfejtésére.

Azonban más komplex rendszerekhez hasonlóan, az anyatej dinamikája sem jósolható meg egyszerűen csak az egyes összetevők kinetikájából *(Christian és mtsai., 2021; Samuel és mtsai., 2020)*. Különféle környezeti tényezők (mint például az anya étrendje, a terhességi kor, a földrajzi elhelyezkedés stb.) okozhatják az anyatej összetételének variabilitását és határozhatják meg az anya-csecsemő dinamikát *(Eidelman és mtsai., 2012)*.

Következésképpen, ha az anyatej egészségre gyakorolt hatását alaposabban meg akarjuk érteni, nem elég az egészségi állapotot közvetlenül az egyes összetevők függvényében tanulmányozni, hanem - a szóban forgó tényezők közötti kölcsönhatások módosító hatásai miatt - azok dinamikáját is vizsgálni kell. *(Christian és mtsai., 2021; Samuel és mtsai., 2020; Boix-Amorós és mtsai., 2019)*

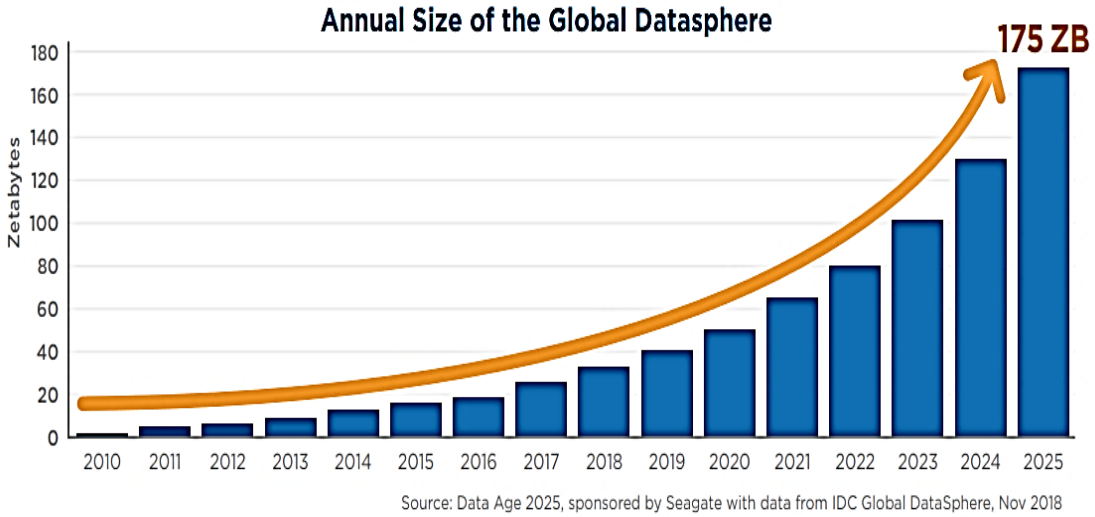
Legjobb tudásunk szerint még nem készült számszerű, statisztikai értékelés az anyatej-összetétel heterogenitásának okairól és azok rangsorolásáról. Ennek egyik oka az anyatej komponensekre vonatkozó longitudinális adatok hiánya *(Christian és mtsai., 2021; Shenhav & Azad, 2022)*, különösen az egyéni anyákra vonatkoztatva. Ezért célunk volt, hogy kutatásunk rávilágítson arra, hogyan lehet ezt a tudáshiányt betölteni, egy adatbázis (MilkyBase adatbázis) létrehozásával, mely ajánlott sablonként szolgál dinamikus modellezésre alkalmas ontológiával és adatokkal.

2.2. MATEMATIKAI FELDOLGOZÁS ÉS SZÁMÍTÁSTUDOMÁNYI MÓDSZEREK

2.2.1. Big Data

A világban az adatok mennyisége robbanásszerűén növekszik, mivel egyre többen ismerik fel az adatokban rejlő lehetőségeket és azok különböző területeken, például a tudományban, az iparban és a kormányzatban való felhasználásának az előnyeit *(Forum, 2012; Cukier és mtsai., 2014)*. A McKinsey Global Institute kutatása szerint az adatok jelentős értéket teremthetnek a világgazdaság számára, használatuk forradalmasíthatja a gazdaság egészét az összes ágazatban, növelve a vállalatok és a közszféra termelékenységét és versenyképességét, valamint jelentős gazdasági többletet teremtve a fogyasztók számára *(Manyika és mtsai., 2011)*. Az adat lehet korunk legértékesebb erőforrása. Az állításnak, hogy a világgazdaság működtetésében még az olajnál is értékesebb, az alapja az, hogy az adatok életünk szinte minden területét áthatják *(Qureshi, 2020)*.

Az International Data Corporation 2018-as előrejelzése szerint 2025-re az évenként generált adatok mennyisége, az úgynevezett globális adattér, 175 zettabájtra (ZB) nő (2. ábra), ami a 2016-ban keletkezett 16,1 zettabájt adatmennyiség több mint tízszerese *(Reinsel és mtsai., 2018)* .



2. ábra. A globális adattér változása 2010-től 2025-ig

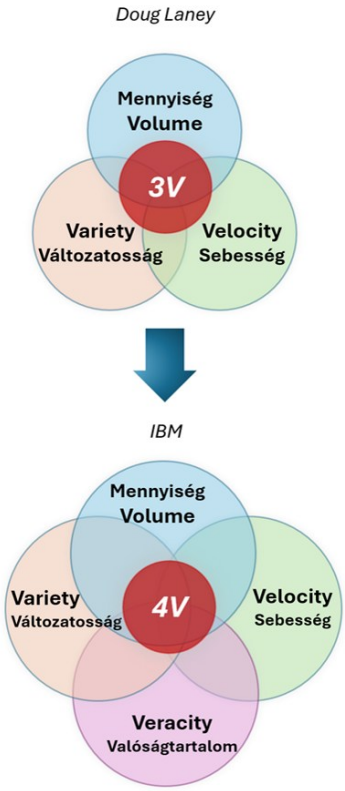
Forrás: *(Reinsel és mtsai., 2018)*

Ez a keletkezett hatalmas adatmennyiség óriási lehetőségeket kínál. Kutatók, adatszakértők, elemzők és üzleti felhasználók használhatják ezeket az adatforrásokat magasabb szintű analitikához, amelyek mélyebb elemzési lehetőségeket nyújtanak, és innovatív Big Data alkalmazásokat működtetnek.

A „Big Data" fogalmát Doug Laney írta le először *(Laney, 2001)* a „V"-s szavak gyűjteményével, utalva az

1) az adatok növekvő mennyiségére (Volume),
2) az adatok előállításának és elemzésének növekvő sebességére (Velocity) és
3) a források, formátumok és ábrázolások növekvő körére (változatosság-Varierty).

Ehhez később az IBM hozzáadta a valóságtartalmat (Veracity) (3.ábra) is, az adatok lefedettségében, pontosságában és időszerűségében mutatkozó jelentős különbségek, azaz az adatforrások nagyon eltérő minőségének jellemzése érdekében. *(Wu és mtsai., 2016)*



3.ábra. A Big Data 4V definíciója

Forrás: *(Wu és mtsai., 2016)*

A definícióktól függetlenül a "Big Data" kifejezést a legegyszerűbben és a legjobban az adatok puszta mennyisége határozza meg *(Mayer-Schönberger & Cukier, 2013)*, olyan adathalmazokra utal, amelyek mérete meghaladja a digitális adat-robbanás (kb.2003) előtti tipikus adatbázisok méreteit *(Manyika és mtsai., 2011)*. Azt a hatalmas adatmennyiséget értjük alatta, amely lehetőséget nyújt egy olyan átfogó kép kialakítására, melyet kisebb adatkészletek elemzésével nem egyszerű megalkotni. Ez a meghatározás szándékosan szubjektív, és magában foglal egy mozgó definíciót arra vonatkozóan, hogy mekkora kell, hogy legyen egy adathalmaz ahhoz, hogy Big Data-nak minősüljön. Nevezetesen, ahogy a technológia fejlődik az idő haladtával, úgy az adatállományok mérete, a Big Data-nak minősülő adathalmazok mérete is növekszik *(Manyika és mtsai., 2011)*. A Big Data definíciója területenként is változhat, attól függően, hogy az adott területen milyen a tipikus adathalmazok mérete.

Az egydimenziós adatok sokszor nem teszik lehetővé rejtett tendenciák, korrelációk és kapcsolatok feltárását. A „Big Data" egyik küldetése ennek a hiánynak a betöltése a sokrétegű adathalmazok elemzésével. A keletkezett nagy adathalmazok olyan új (nem feltétlenül relációs) adatbázis struktúrák kifejlesztését követelték meg, amelyek képesek strukturálatlan adatok kezelésére, valamint olyan számítási algoritmusok fejlesztését, amelyek hatékonyan tudják kezelni a Big Data összes dimenzióját. A nagy adathalmazok kezelésére, elemzésére és vizualizálására kialakított új technikák és technológiák különböző területekről származnak, többek között a statisztikából, az informatikából, az alkalmazott matematikából és a közgazdaságtanból *(Manyika és mtsai., 2011)*. A gyakran használt technikák közé tartozik az adatbányászat, a szövegelemzés, a prediktív elemzés, az adatmegjelenítés, a mesterséges intelligencia, a gépi tanulás, a statisztika és a természetes nyelvi feldolgozás. A Big Data értékeinek kiaknázásához elengedhetetlen egyfajta rugalmas, multidiszciplináris megközelítési mód.

A benne rejlő lehetőségek ellenére, a nagy adathalmazok továbbra is ki vannak téve a hagyományos adatelemzési módszerekből eredő kihívásoknak, mint például a mintavételi hibák, torzítás, szignifikanciaszintek értelmezése stb. Mindazonáltal az élelmiszer-biztonság terén a nagy adathalmazokat használó eszközök kifejlesztése jelentős potenciállal rendelkezik a mikrobiális élelmiszer-biztonság és -minőség javítására. Mindezen túl a kifejezetten az élelmiszer-biztonsági alkalmazásokhoz létrehozott nagy adatkészletek mellett a szakemberek egyre inkább felismerik azon

nagyobb adatkészletek használatának értékét, amelyek nem kifejezetten az élelmiszer-biztonsági alkalmazásokhoz készültek *(Marvin és mtsai., 2017)*.

2.2.2. Big data az élelmiszertudományban: élelmiszer összetevő adatbázisok

Az élelmiszer-összetételre vonatkozó adatok adatbázisokba való rendezése már a 19. század közepén elkezdődött *(Church, 2006)*, és azóta több ország és szervezet létrehozta már a sajátját, az adatbázis céljától és kívánt felbontásától függően különböző ontológiákat követve *(Yeung, 2023)*. Az élelmiszerösszetevő adatbázisok olyan fontos eszközök, amelyek a - főként a feldolgozatlan - élelmiszerek tápanyagtartalmáról biztosítanak információt, beleértve a makró tápanyagokat (pl. szénhidrátok, fehérjék és zsírok), mikró tápanyagokat (pl. vitaminok és ásványi anyagok) és egyéb összetevőket (pl. élelmi rostok és víz) *(Marconi és mtsai., 2018)*. A tárolt adatokat a táplálkozási szakemberek, dietetikusok és kutatók használják étkezések megtervezésére, az étrendek tápértékének meghatározására, illetve annak vizsgálatára, hogy az táplálkozás hogyan függ össze az egészséggel *(Ene-Obong és mtsai., 2019)*. Az élelmiszerek komplexitásának feltárásában, valamint az interneten tárolt tudás hasznosításában egy megfelelően felépített adatbázis, numerikus, illetve statisztikai eszközökkel kombinálva, hatalmas lehetőségeket rejt magában *(Barabási és mtsai., 2020; Morgenstern és mtsai., 2021)*.

2.2.3. Matematikai modellalkotás, elsődleges/másodlagos modell

Az anyatej összetevők időbeli változásának megértéséhez elengedhetetlen a matematikai modellalkotás. Ezen kulcsfontosságú eszköz az élettudományok különböző területein a matematikai fogalmak és nyelvezet segítségével segít megérteni az összetett rendszereket. A matematikai modellek előrejelzésekre, rendszeroptimalizálásra és a rendszerek különböző körülmények közötti viselkedésének megértésére is használhatók.

A modellalkotás filozófiájában rendszerint különbséget tesznek a matematikai, a számítástechnikai és a materiális (azaz tényleges vagy fizikai) modellek között *(O'Malley & Parke, 2018; Dym, 2004)*. A matematikai modellek alapját az egyenletek alkotják, a számítástechnikai modelleket algoritmusok szolgálják, a fizikai szerkezetek különböző fajtáit pedig a materiális modellek felépítéséhez használják *(Weisberg, 2013)*. A modelleket egy rendszer viselkedésének szimulálására, különböző forgatókönyvek kimenetelének előrejelzésére és a rendszer viselkedését befolyásoló kulcstényezők

azonosítására használják. A modellek lehetnek determinisztikusak vagy sztochasztikusak, és az egyszerű lineáris egyenletektől a komplex számítógépes szimulációkig terjedhetnek.

A matematikai modellalkotás alapvető a tudományban a rendszerek viselkedésének oly módon történő feltárása céljából, ahogy az a valóságban gyakran lehetetlen vagy túl kockázatos lenne. A matematikai módszerek segítségével felállított modell az összetett kérdéseket képes formalizálni és megválaszolhatóvá tenni, precizitást és problémamegoldási stratégiát adni, ezáltal a modellezett rendszer szisztematikus megértését kínálva. Ez jobb tervezéshez, a rendszerek irányításához és a modern számítástechnikai lehetőségek hatékony felhasználásához vezet *(Galit, 2010)*.

Élettudományokban a matematikai modellezés gyakorlatát gyakran a statisztikai elemzéssel azonosítják, holott ez két különböző, de egymást kiegészítő megközelítés az adatok elemzésére és előrejelzések készítésére. Mindkét megközelítés fontos eszköz az összetett rendszerek megértéséhez és a megalapozott döntéshozatalhoz, de míg a statisztika elsősorban az adatok elemzésére, összegzésére és leírására összpontosít, addig a modellezés egy rendszer vagy folyamat matematikai vagy fogalmi reprezentációjának létrehozására. A statisztikában a cél gyakran egy populáció leírása vagy következtetések levonása egy adatminta alapján, míg a modellezésben a cél a rendszer viselkedését szabályozó mögöttes mechanizmusok megértése *(Cox, 1990; Galit, 2010)*.

A statisztika és a modellezés közötti másik alapvető különbség a bizonytalanság megközelítésében rejlik. A statisztikában a bizonytalanságot gyakran valószínűségek, konfidenciaintervallumok és hipotézisvizsgálat segítségével számszerűsítik. A cél a különböző kimenetek valószínűségének megbecslése, vagy annak vizsgálata, hogy egy adott eredmény statisztikailag szignifikáns-e. A modellezésben a bizonytalanságot gyakran érzékenységi elemzés, forgatókönyv-elemzés vagy Monte Carlo-szimulációk segítségével építik be *(Salciccioli és mtsai., 2016; Banack és mtsai., 2021)*. A cél annak megértése, hogy a bemeneti paraméterek vagy feltételezések változásai hogyan befolyásolják a modell kimeneteit, és annak azonosítása, hogy mely tényezőknek van a legnagyobb hatása a rendszer viselkedésére (érzékenység-analízis).

E különbségek ellenére a statisztikát és a modellezést gyakran együtt használják a komplex rendszerek jobb megértéséhez, számos modellezési megközelítés statisztikai módszereket is tartalmaz. A regresszióanalízis például olyan statisztikai módszer, amely prediktív modellek kidolgozására használható, és a gépi tanulási algoritmusok is gyakran tartalmaznak statisztikai technikákat az adatelemzéshez és a modellépítéshez.

Az idővel változó rendszerek – mint például kutatásunkban az anyatej, vagy a modellként szolgáló mikrobiológiai rendszerek - megértéséhez értékes eszközök a dinamikus modellek *(Strogatz, 2018)*. Ezeket a modelleket olyan rendszerek viselkedésének szimulálására használják, amikor az rendszerek időbeli változását kell nyomon követni (dinamikus rendszerek). Széles körben alkalmazzák őket a mikrobiológia, az ökológia, a közgazdaságtan, környezettudományok, a szociológia, a politika, a mérnöki tudományok és a fizika területein is. A dinamikus modellek gyakran differenciálegyenletek, amelyek a rendszer változásának a sebességét a rendszer állapotának függvényében írják le. Ezek a modellek komplexek és nemlineárisok is lehetnek, így a természetes és dinamikus rendszerek megértésére alkalmasabbak, mint a lineáris statisztikai megközelítések *(Saltelli, 2019)*.

A dinamikus modellezés és az alapstatisztika közötti fő különbség az alkalmazásokban és a célkitűzésekben rejlik. A dinamikus modellezést jellemzően időfüggő rendszerek szimulációjára és előrejelzésére használják, míg a matematikai statisztika alapvető fontosságú az adatelemzés és következtetés során. Mindkettő valós problémák megértésének és megoldásának szerves részét képezi, de az adott problémától függően eltérő módon alkalmazhatóak. E két megközelítés (a dinamikus modellezés és a statisztika) integrációja számos tudományos törekvésben megfigyelhető, ahol statisztikai módszereket használnak a dinamikus modellek paramétereinek becslésére vagy e modellek empirikus adatok alapján történő validálására. Mindazonáltal a dinamikus modellek alkalmasabbak prediktív előrejelzésre mint a standard statisztikai elemzések *(Saltelli, 2019)*.

A dinamikus modellek általában differenciálegyenletek, amelyek a kiválasztott változók közötti kölcsönhatásokat és azok időbeli változásának a sebességét írják le az állapotuk függvényében. A "milyen változók és milyen kölcsönhatások" kérdésre adott válasz azonban nem mindig egyértelmű, és egyfajta egyszerűsítést, "a lényegtelen elhagyását" igényli *(Baranyi, 2005)*, ahol a "lényegtelen " kiválasztása a gyakorlati

szakemberek és a kapcsolódó szakterületek kutatói közötti konszenzus eredménye. A lényegtelen információk elhagyása után eredményül kapott modellnek kompromisszumot kell teremtenie a részletek (felbontás), az azonosíthatóság (a paraméterek a megfigyelt adatokból becsülhetők) és a komplexitás (azaz, hogy mennyire implementálható a döntéshozatalt segítő előrejelzési szoftverben) között.

A modell kidolgozása során a komplexitás csökkentése az egyik legfontosabb feladat. Egy differenciálegyenlet-rendszer adatokra való illesztése kihívást jelenthet, ezért a modell paramétereinek azonosításához az explicit algebrai függvényeket részesítik előnyben. Emellett bizonyos paraméterek (például biokémiai sebességek) magasabb prioritást élvezhetnek, mint mások, például, ha szoros kapcsolatuk van a mechanisztikus modellezéssel és/vagy megbízható ismeretekkel rendelkezünk a számítási tartományukról. Más paraméterek, mint például az előzmények hatását kifejező, a modell kezdeti értékeit képező paraméterek, ismeretlen tényezőktől függően sztochasztikusak lehetnek. Ezek a szempontok vezetnek bennünket ahhoz a modellezési eljáráshoz, amelyet Shenhav & Azad *(Shenhav & Azad, 2022)* tanulmánya, valamint a prediktív mikrobiológia elmúlt három évtizedben végbement fejlődése *(McMeekin és mtsai., 2002)* inspirált, az **elsődleges/másodlagos modellezés** módszeréhez.

Az "elsődleges/másodlagos modell" terminológia a prediktív élelmiszer mikrobiológia *(McMeekin és mtsai., 2002)* területéről származik, ahol ezeket a kifejezéseket a 80-as években vezették be. Ezen időszakban a prediktív élelmiszer mikrobiológia tudományának fejlődése, olyan szakaszokon ment keresztül, amely nagyon hasonló az anyatejkutatásban jelenleg tapasztalhatóhoz: hatalmas mennyiségű adatot generáltak anélkül, hogy elegendő erőfeszítést tettek volna az előállított adatok megfelelő, prediktív modellezéssel történő értelmezésére. A mikrobiológiában azt már korábban leírták, hogy az élelmiszerek feldolgozása, szállítása, forgalmazása és tárolása során a környezetben bekövetkező összetett fizikai, kémiai és biológiai változásoknak kitett mikroorganizmus-populációk dinamikája (növekedés és túlélés) olyan összetett folyamat, amelyet számos környezeti tényező befolyásol, mint például a hőmérsékletet, a pH, a vízaktivitás és a tápanyagok elérhetősége *(Pirt, 1975)*. Az ezen folyamatok matematikai modellekkel való leírása iránti igény azután erősödött, hogy felismerték élelmiszerek utólagos minőség-ellenőrzése drága és időigényes *(Mahdinia és mtsai., 2020)*.

A prediktív élelmiszer mikrobiológia fejlődésével számtalan modellt és modell osztályozási sémát dolgoztak ki *(Buchanan, 1992)*. Ezek közül az egyik leggyakrabban használt a Buchanan *(Buchanan, 1993; McDonald & Sun, 1999)* által javasolt osztályozás, amely a legtöbb modelltípust elsődleges, másodlagos modell és a modellek implementációja kategóriákba csoportosítja:

(i) **Elsődleges modellek**: a mikrobiális növekedési és inaktiválási fázisok kinetikai folyamatait írják le mindössze néhány paraméter segítségével, és a populáció sűrűségének időbeli növekedését (vagy csökkenését) rögzítik. Mikrobiológiában a leggyakrabban használt elsődleges modellek a Baranyi, és a Gompertz- egyenlet, amelyek a mikrobiális növekedés különböző fázisait írják le *(Baranyi és mtsai., 1993; Zwietering és mtsai., 1990; Buchanan és mtsai., 1997; Baranyi-Rockaya- és mtsai., 2024)*. Általánosságban az elsődleges modellek a matematikai modellezésben olyan alapmodellek, amelyek egy rendszer vagy folyamat alapvető kvantitatív összefüggéseit írják le. Ezeket gyakran közvetlenül elméleti alapelvekből vagy empirikus megfigyelésekből vezetik le *(Galit, 2010)*. Ezek a modellek jellemzően közvetlen méréseket és megfigyeléseket tartalmaznak, alapját képezve a bonyolultabb elemzéseknek. Az elsődleges modelleket gyakran használják az alapvető folyamatok leírására olyan területeken, mint a fizika, a kémia és a biológia *(May, 2001; Murray, 2003)*. Példaként említhetjük statisztikában a lineáris regressziós modelleket, a fizikában az alapegyenleteket, a biológiában pedig az alapvető populációnövekedési modelleket. Bár az elsődleges modellek alapvető fontosságúak az alapvető folyamatok megértéséhez, gyakran egyszerűsítéseket és feltételezéseket igényelnek, amelyek korlátozhatják pontosságukat az összetett valós alkalmazásokban *(Hartmann, 2005)*.

(ii) **Másodlagos modellek**: az elsődleges modell paramétereit modellezik a rájuk ható környezeti tényezők (például a hőmérséklet, a nedvesség vagy a pH) függvényében. A gyakori másodlagos modellek gyakran empirikus egyenletek, melyek regressziója a válaszfelület-módszertana közé sorolható *(Ratkowsky és mtsai., 1983)*.

(iii) **Modellek implementációi** egy vagy több elsődleges és másodlagos modellt kombinálnak számítógépes szoftver segítségével, és olyan modellrendszert hoznak létre (ilyen például a ComBase is *(Baranyi & Tamplin, 2004)*), amely felhasználóbarát felületet biztosít.

Ha egy tudományágra vonatkozóan meghatározó adatmennyiség áll rendelkezésre, a prediktív modellezés hasznos eszközzé válik ahhoz, hogy a tudományterületet ne csak leíróan, hanem előre jelzően is lehessen használni. Az ilyen törekvések egyes tudományágakban, például a biotechnológiában és az élelmiszermikrobiológiában is már sikeresek voltak *(Bailey, 1998; McMeekin és mtsai., 2002)*, és az anyatejkutatásban is érezhető az igény az adatok előrejelzésekre történő kiaknázására.

3. ANYAG ÉS MÓDSZER

A kutatás során alkalmazott módszereket 3 fő szakaszra lehet felosztani. Először az anyatej összetevők rögzítésére alkalmas adatbázis ontológiáját definiáltuk, és töltöttük fel adatokkal a publikációk alapján, majd az így kapott alap adatbázist bővítettük irányított kereséssel, végül a kapott adatokat elemeztük és modelleztük.

3.1.A MILKYBASE ONTOLÓGIA LÉTREHOZÁSA

3.1.1. Alapelvek

Az élelmiszer-összetételre vonatkozó adatok különböző formátumokban kerülnek bemutatásra. Rendkívül fontos ezeknek a formátumoknak a szabványosítása, az adatbázisok kompatibilissé tétele és teljes potenciáljának kihasználása. Az egységes struktúra megkönnyíti az adatkészleteken belüli és azok közötti összehasonlítást, és lehetővé teszi a gyors keresést az adatbázisban előre meghatározott kulcsparaméterek alapján.

Mivel célunk egy olyan ontológia létrehozása volt, amely a táplálkozási szakemberek és kutatók számára, valamint az iparban és a szabályozásban is jól használható eszköz lehet, ezért az ontológia definiálásakor számos kompromisszumot kellett kötni, hogy megtaláljuk az egyensúlyt Big Data a korábban ismertetett négy fő alappillére - ***méret, sebesség***, ***változatosság*** és ***hitelesség*** - között.

3.1.1.1. Méret

Az alapelvek szem előtt tartásával egy olyan adatbázist hoztunk létre, amely az anyatej molekuláris összetevőinek publikált méréseit tartalmazza. Az alap MilkyBase a maga kb. 10000 adatpontjával messze elmarad a Big Data projektektől elvárható adathalmaz nagyságtól. Reméljük azonban, hogy az általunk létrehozott ontológia olyan, amelyet a kutatók és a klinikusok is használhatnak saját adataik bevitelére, s így - a kollektív tudás tárházaként - további fontos élelmiszer-típusok "periódusos táblázatának " *(PTFI, 2021)* közös formátuma lehet.

Ahhoz, hogy ez a cél megvalósulhasson, az adatok bevitelére szolgáló sablonnak felhasználóbarátnak kell lennie, lehetőleg egy olyan széleskörűen ismert és használt platformon, amelyet az adatot szolgáltató felhasználók könnyen kezelnek. Ezen

irányelvek alapján a választásunk az adatok adatbázisba rögzítéséhez a Microsoft Excel táblázatkezelő programra esett. Ez a legelterjedtebb és legismertebb felhasználói program, amely képes táblázatok összekapcsolására, és emellett adatmegjelenítési és elemzési eszközöket is kínál. Mindemellett funkciói kibővíthetőek a Visual Basic for Applications (VBA) objektum orientált programozási nyelv segítségével. Az általunk létrehozott VBA segédprogramok lehetővé teszik a rekordok rögzítése során a bemeneti ellenőrzést, valamint segítik az adatelemzést is (például a saját és mások hasonló publikált mérés eredményeinek összehasonlítását), ösztönzve ezzel az adatkurátorokat a releváns adataik adatbázisunkba való rögzítésére. Ez az úgynevezett wiki-filozófia realizációja, azaz a tudás hozzáadása a közöshöz, amely potenciálisan jóval nagyobb adatbázishoz vezethet.

3.1.1.2. Sebesség

Az adatbázis aktuális mérete mellett az adatok közötti navigáció és az adatfeldolgozás az elvárásoknak megfelelő sebességgel működik. Azonban a jelenleg használt Excel platform az adatmennyiség növekedésével elveszítheti az egyszerű használhatóságából származó előnyét, ezért az adatbázis növekedésével célszerű lesz azt egy SQL szerverre importálni, és ezt követően a jelenlegi Excel táblázatokat az adatfeltöltők (az alapszintű felhasználók) tranzit területeként használhatják a kezdeti adatbevitelhez és adatbázis bővítéshez.

3.1.1.3. Változatosság és hitelesség

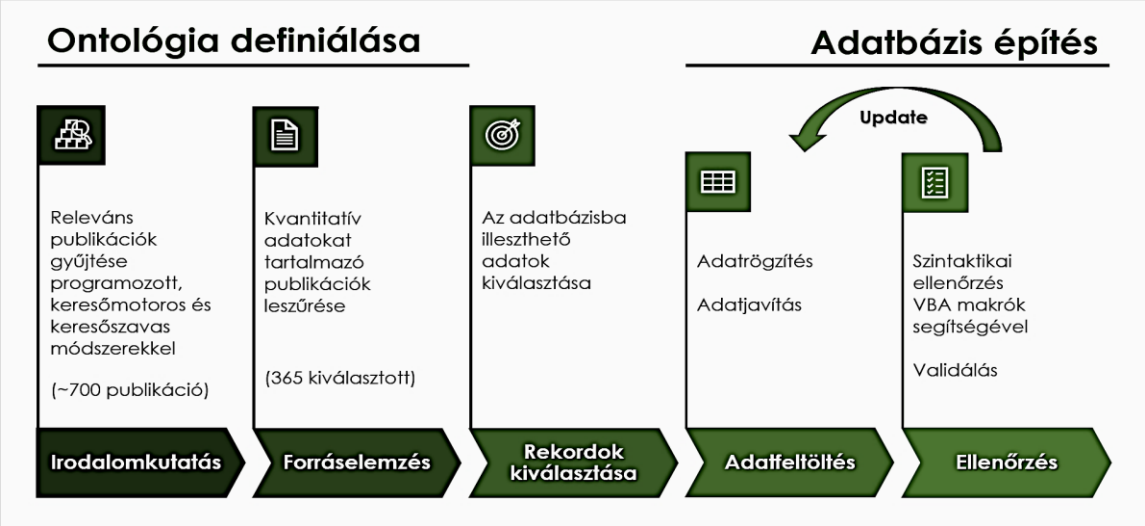
Célunk az volt, hogy a közzétett adatokat egy szigorúan szervezett adatbázisban digitalizáljuk, amely készen áll az anyatej összetételének az elemzésre, a különböző feltételek függvényében. A hitelesség fenntartására az adatok rögzítése során igyekeztünk elkerülni a közzétett adatok megváltoztatását. Kivételt képeztek azok a triviális esetek (mint például a mértékegységek átváltása), melyekben a publikált adatok rögzítése - az egységesség érdekében - a publikált adatok átalakításával történt. Azokban az esetekben, amikor az értékek g, mg-ban a 100 g élelmiszertömegre vonatkoztatva voltak megadva (pl. mg/g), akkor azokat g/literre (vagy g/l-re) váltottuk át, ahol 1 liter tejet 1 kg-nak feltételeztünk *(NIST, 2022)*.

Sokszor találtunk kétértelműséget vagy ellentmondást a szerzők által használt terminológiákban is. Erre példa amikor egy adott zsírsavmolekula koncentrációját (amelyet többnyire az *összes zsírsavhoz* viszonyított arányként közölnek) egyes szerzők

az *anyatej teljes tömegéhez* viszonyított arányként adták meg, sőt előfordult, hogy az összes *mért* zsírsav arányaként közölték. Ezekben az esetekben - a hitelesség szem előtt tartásával - legjobb tudásunk szerint definiáltuk és számszerűsítettük a publikált adatokat.

3.1.2. Az adatbázis-építés folyamata

Adatbázis építési folyamatunk (4.ábra.) során először az adatbázis „vázát" jelentő ontológiát definiáltuk, majd a felállított ontológia alapján létrehoztuk a tervezett adatbázist.



4.ábra. Az adatbáziselőpítés folyamata

Az ontológia definiálásához elengedhetetlen volt az ontológiában rögzítendő adatok felmérése és a legfontosabb rögzítendő adatcsoportok meghatározása. Ehhez a publikációk keresőmotoros és keresőszavas gyűjtése (**Irodalomkutatás**) után kapott tudományos cikkeket áttekintettük, értékeltük és kiválasztottuk az adatbázisba rögzíthetőket (**Forráselemzés**). A fő kiválasztási kritérium az volt, hogy az adatbázisba rögzítendő publikációnak mennyiségi adatokat kell tartalmaznia az anyatej összetevőiről; lehetőleg úgynevezett dinamikus adatokat, amelyek az összetevők időbeli változását mutatják. Ezeket ideális esetben a publikációkban olyan táblázatokba rendezve találhattuk meg, amelyeknek adatait könnyebb volt beilleszteni az Excel táblázatunk lapjaira.

Ezt a kiválasztott publikációk részletes elemzése követte, melynek célja az volt, hogy azonosítsuk a rögzítésre alkalmas adatkészleteket (**Rekordok kiválasztása**).

Mivel az ontológia építés egyik fő célja a lehető legszélesebb körű információ összegyűjtése volt, nemcsak az anyatej összetételére vonatkozó adatokat választottuk ki rögzítésre, hanem az arra vonatkozó adatokat is, hogy milyen körülmények között generálódtak az anyatej összetevőkre mért adatok.

Az ontológia definiálását követően, az adatbázis adatokkal való feltöltése során (**Adatfeltöltés**) folyamatosan validáltuk a rögzítendő és már rögzített adatokat (**Ellenőrzés**). Ehhez egyrészt a megfelelő szintaxis ellenőrzésére „ellenőrző makrókat" fejlesztettünk ki, másrészt szemantikai validációt alkalmaztunk a publikációkban található anomáliák azonosítására. A minőségellenőrzés (szemantika ellenőrzése) sokkal nehezebb feladat, mint a szintaktikai ellenőrzés, hiszen elvégzése emberi intelligenciát igényel.

Az **ontológia innovációja, hogy az anyatej komponensek időbeli lefutását** - amit a dolgozat további részében ***trajektóriának*** vagy *időbeli pályának* nevezünk- **egyetlen bejegyzésként (matematikai értelemben vett változóként) kezeli** - mint az adathalmaz elemi alanya - és helyezi a középpontba. Ezt a trajektóriát egy mutató (pointer) azonosítja, amely az [idő, koncentráció] párok mért vagy becsült értékeit tartalmazza. Az adatbázis ontológiája követi Shenhav és Azad *(Shenhav & Azad, 2022)* ajánlásait, akik az anyatejkutatás előmozdítása érdekében longitudinális adatgyűjtést - lehetőleg egyéni anyáktól - szorgalmaztak.

3.1.2.1. Irodalomkutatás

Az irodalomkutatás első lépéseként az anyatej összetételével kapcsolatos releváns tudományos publikációkat kerestünk nagy adatbázisokban, programozott keresőmotoros, illetve manuális kulcsszavas kereséssel.

A programozott keresőmotoros keresést, a kutatás első szakaszában részt vevő kollaborátorunk - Barabási-Albert László kutatócsoportja *(BarabasiLab, 2024)* – végezte el, az általuk létrehozott FoodMine *(Hooton és mtsai., 2020)* elnevezésű természetes nyelvi feldolgozó algoritmussal. A FoodMine keresőmotor a PubMed *(PubMed, 2022)* adatbázisból keresi meg egy célzott élelmiszer - esetünkben az anyatej-kémiai összetételével kapcsolatos publikációkat, a cikkek címét és absztraktját szisztematikusan elemezve.

A FoodMine keresőmotoros keresés eredményeként egy - 1308 publikáció adatait tartalmazó - listát kaptunk mely a következő adatokat tartalmazta:

- **PMID** A publikáció PubMed egyedi azonosító száma („PubMed IDentifier”)
- **abstract** A publikáció rövid összefoglalója (absztrakt)
- **journal** A publikációt megjelentető folyóirat
- **mesh_terms** A publikáció keresése során használt MeSH (Medical Subject Headings) *(MeSH, 2020)* kifejezés, mely az orvosi és egészségtudományi irodalom témáinak indexelésére és keresésére használt szabványosított kulcsszó.
- **paper** A publikáció címe
- **webpage** A publikáció internetes elérhetősége
- **year** A publikáció megjelenésének éve

A manuális keresés során további tudományos publikációs adatbázisokban végeztünk kereséseket, melyek során a következő a PubMedben *(PubMed, 2022)* található Medical Subject Headings (MeSH) *(MeSH, 2020)* kifejezéseket és logikai kifejezéseket *(Gries & Schneider, 1993)* használtuk:

("human milk" OR "mother milk" OR "mothers' milk") AND ("nutrients" OR "components" OR "composition" OR "biochemical" OR "quantification" OR "bioactive").

A keresés főként az angol nyelvű publikációkra összpontosított, de nem csak azokra korlátozódott.

3.1.2.2. Forráselemzés

A keresési eredményeket az előre meghatározott szókapcsolatok, a MeSH-kifejezések és a PubMed-bejegyzésben szereplő dolgozat kivonatának szöveges megfeleltetését alkalmazva szűkítettük a keresési eredményeket. Miután megkaptuk az eredmények részhalmazát, manuálisan megkerestük a publikációkat, és 551 anyatejre vonatkozó cikket töltöttünk le, ha teljes szöveges linket tudtunk elérni.

A forráselemzés során a fő kiválasztási kritérium az anyatej tápanyag- és/vagy nem tápanyag-összetevőire vonatkozó mennyiségi adatok voltak. Elsőbbséget élveztek azok az adatok, amelyek

(i) táblázatos formában, szisztematikusan rendszerezettek,

(ii) időbeli változásokat mutatnak (azaz úgynevezett dinamikus adatok);

(iii) bizonytalansági számszerűsítéssel vannak ellátva.

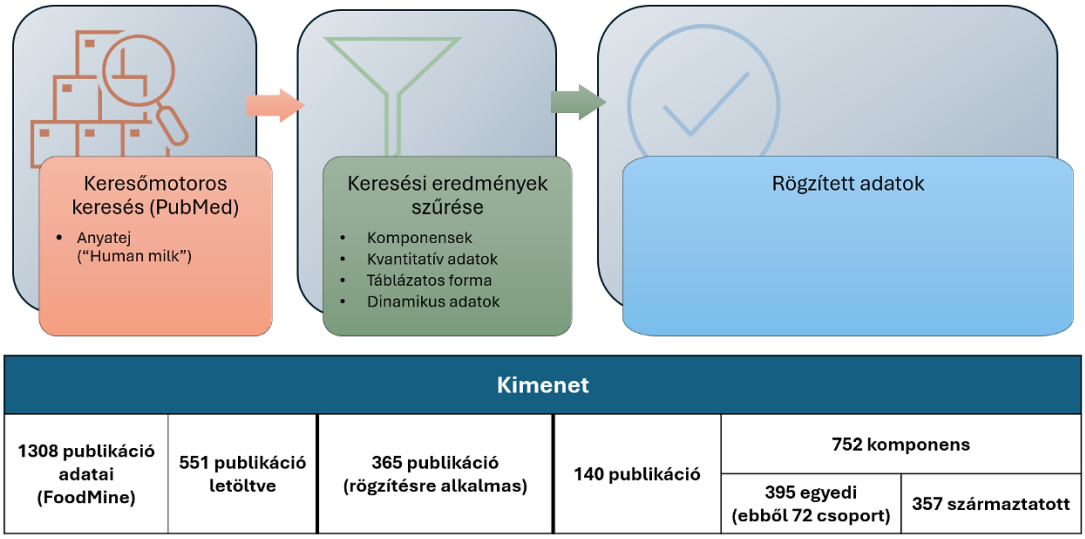
A nem releváns tanulmányok kizárása után összesen 365 potenciális tanulmányt azonosítottunk, amelyeket alkalmasnak találtunk az adatbázisba való rögzítésre. A **MilkyBase alap** adatbázisa 2022. július 1-jén **140 tanulmány adatait tartalmazta**.

Végül manuálisan kiértékeltünk minden egyes szűrésen fentmaradt publikációt, hogy releváns kémiai tartalmukat azonosítsuk, és kinyerjük az rögzíthető információkat belőlük.

3.1.2.3. Rekordok kiválasztása; Az összetevők azonosítása; Adatfeltöltés

Az MilkyBase adatbázis publikált 1. verziójában (úgynevezett alap adatbázis /core database) több mint 750 (egyedi vagy származtatott) komponenst azonosítottunk, amelyek vagy egy fa szerkezetű értékkészlet (*lsd. 4.2.4. Általános fa adatstruktúra*) csomópontjai vagy levelei, vagy a köztük lévő kapcsolatok lehetnek. Az így kapott adathalmazban néhány egyedi molekula explicit és implicit módon *(lsd. 4.2.5. Az adatok direkt és indirekt (származtatott) formában is rögzíthetők*) is reprezentálva van. Ez azt jelenti, hogy például egy adott zsírsav egyrészt g/liter mértékegységgel, másrészt az összes zsírsav mennyiséghez - amelyet grammban mérnek – viszonyított arányszámként is megjelenhet az adatbázisban. Az ilyen „duplikátumokat” leszámítva körülbelül 400 "valódi" komponensről léteznek explicit mérések. Ezek közül nagyjából

70 csoportként is funkcionál, azaz vagy további csoportokra, vagy a fa végső „leveleiként” definiált molekulákra bonthatóak (5. ábra.).



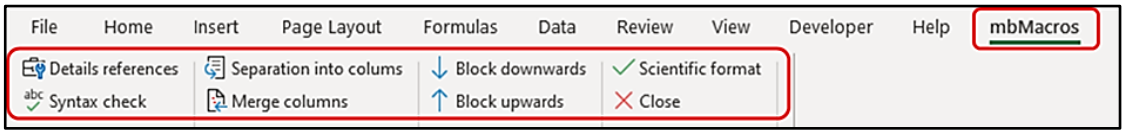
5. ábra. Az adatkiválasztási folyamat eredménye

3.1.2.4. Ellenőrzés, Hibajavítás

A Big Data két korábban említett „változatosság – valódiság” kérdése szorosan összefügg az adatbázis mezőinek szintaktikájával és szemantikájával. Míg a szintaktika automatizált módon ellenőrizhető, a szemantika gyakran mutat anomáliákat, amelyek befolyásolják, hogy milyen adatok adhatók meg (változatosság) és hogyan ellenőrizhetők (hitelesség).

3.1.2.4.1. Rögzítési hiba

Az adatbázis szintaktikai ellenőrzését MS Excel VBA makrók (6. ábra.) segítik. A makrókat tartalmazó MBmacros.xlsm fájl bármely felhasználó számára elérhető a Figshare *(Pacza és mtsai., 2022)* adattárban. Az adatok rögzítésekor „Syntax check” validáló makrót rendszeresen futtatva elérhető, hogy a rögzített adatok a mezők előírt formátumának megfelelően legyenek rögzítve, elősegítve ezzel az adatelemzést.



6. ábra. Az MBmacros eszköztára

3.1.2.4.2. Publikációs hiba feltárása

A szintaktikai hibák a kifejlesztett "Syntax check" segítségével könnyen észlelhetők, de a szemantikai ellenőrzés emberi figyelmet és speciális ismereteket igényel.

Különböző összehasonlító ábrákat használtunk a publikációkban található anomáliák felfedezésére. Ilyen anomáliák például a helytelen mértékegységek, az ábrák és táblázatok közötti ellentmondások, illetve a félreértelmezett adatszórások és bizonytalansági számszerűsítések.

Ha a publikációban olyan triviális hibákat találtunk (például az egyik mértékegységről egy másikra való átváltáskor keletkező hibát), melyek könnyen javíthatók voltak, akkor ezt az adatrögzítés folyamán megtettük; ellenkező esetben kihagytuk a rekordot, vagy "gyanúsnak" jelöltük.

Az így kapott alap-adatbázis, a legnagyobb gondossággal végzett ellenőrzések ellenére is, elkerülhetetlenül tartalmazhat hibákat. Ezek az eltérések azonban az adatbázis használata során észlelhetőek és korrigálhatóak. Mi magunk is felfedeztünk ilyen ellentmondásokat a modellalkotási folyamatunk során, melyeket egy újabb verzió publikálásával korrigáltunk is.

A valósághűséget a statisztikai/számtani fogalmakkal kapcsolatos félreértések is befolyásolják. Például a mért értékek szórását néha összetévesztik az átlaguk standard hibájával. Erre a tévedésre korábban már több publikáció is felhívta a figyelmet *(Vaux, 2012; Chavalarias és mtsai., 2016)* , de ez a fogalomtévesztés sajnos még mindig gyakran előfordul. Hasonlóképpen előfordul, hogy összekeverik a kvantiliseket (amelyek a nyers adatok szórására vonatkoznak) a konfidenciaintervallumokkal (amelyek a becslés pontosságára vonatkoznak). Amikor ilyen hibákat észleltünk, vagy kijavítottuk őket (ha nyilvánvaló), vagy az adatbázisban jelöltük a gyanús adatot (kevésbé nyilvánvaló helyzetekben).

3.2.AZ ALAP MILKYBASE ADATBÁZIS KIBŐVÍTÉSE IRÁNYÍTOTT KERESÉSSEL

A kutatásom második szakaszában az alap MilkyBase adatbázist a pontosabb adatkiértékelés érdekében további adatokkal egészítettük ki. Az adatbázis bővítésekor, az irodalomkutatás során az anyatej kutatások regionális helyzetére koncentráltunk, és két kiemelt régió adataival bővítettük, Ázsiával és Dél-Amerikával.

Az irodalomfeltárás során a SciELO – Scientific Electronic Library Online (SciELO) *(SciELO, 2022)*, a Scopus *(Elsevier, 2022)*, PubMed *(PubMed, 2022)*, Web of Science (WOS) *(WOS, 2022)* online publikációs adatbázisokban végeztünk kereséseket.

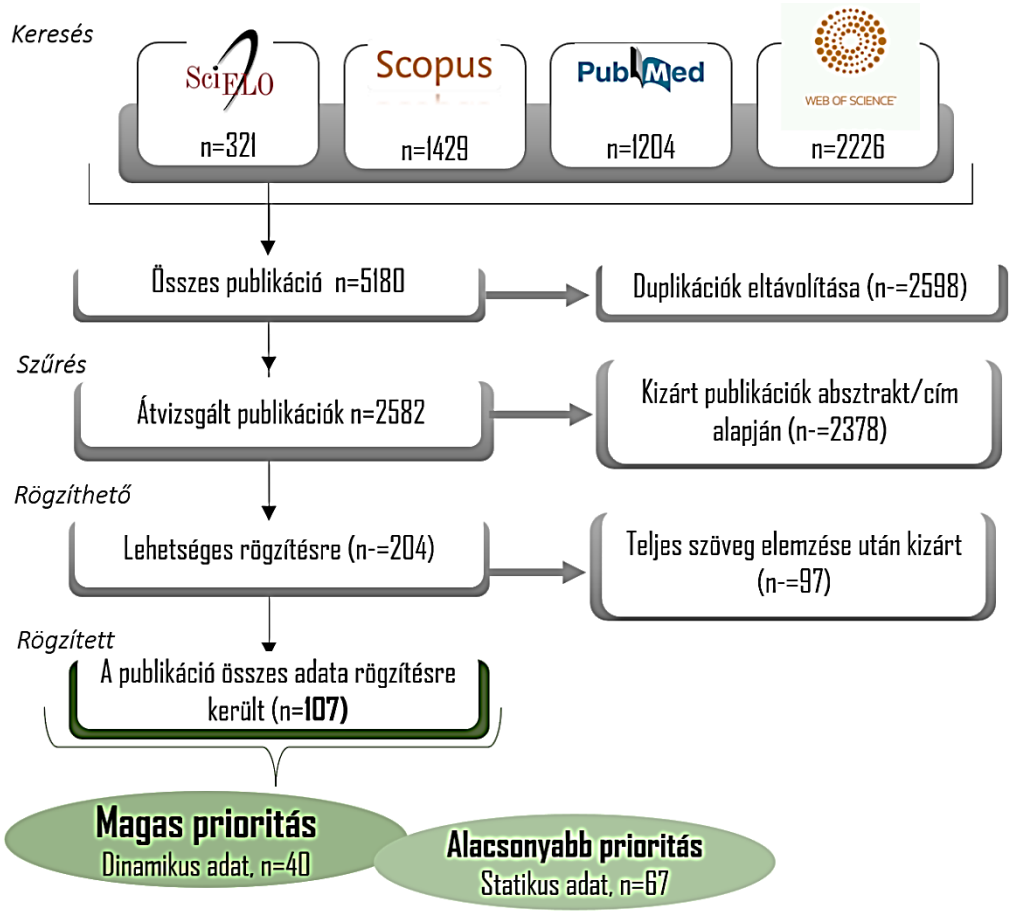
A korábbiakban ismertetett (*lsd. 3.1.2 Az adatbázis-építés folyamata*) adatfeltöltési metódussal (irodalomkutatás, forráselemzés, rekordok kiválasztása) az alap MilkyBase adatbázis a következőképpen egészült ki.

3.2.1. Irodalomkutatás, Forráselemzés:

3.2.1.1. Dél Amerika:

A használt keresés és az eredményül kapott publikációk szűrési folyamata (7. **ábra**):

Search: ((Brazil) OR (Brazilian) OR (Brasileira) OR (Brasileiro) OR (Brasil) OR (brasileña) OR (brasileño)) AND ((human milk) OR (breast milk) OR (breastmilk) OR (leite materno) OR (leite humano) OR (leche humana) OR (leche materna))

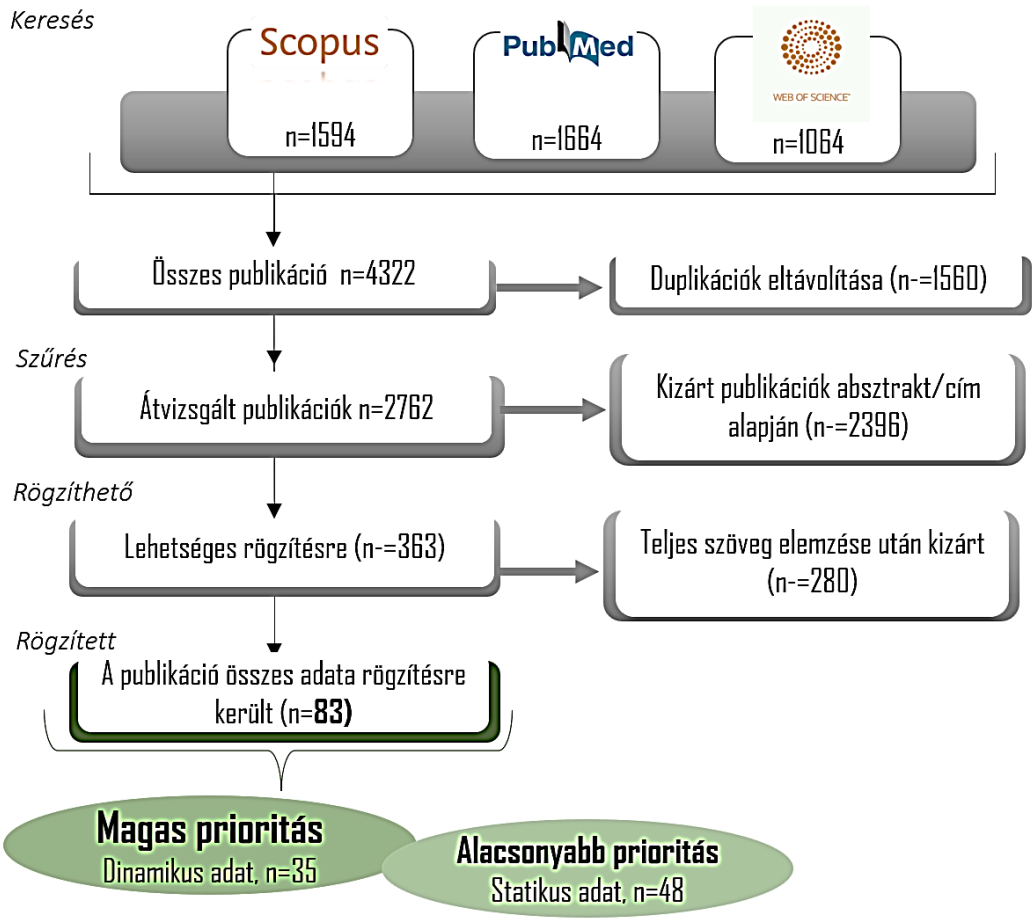


7. ábra. Adatbázis bővítés, irodalomkutatás - Dél-Amerika

3.2.1.2. Ázsia:

A használt keresés és az eredményül kapott publikációk szűrési folyamata (8. **ábra**):

Search:((breastmilk) OR (breast_milk) OR (humanmilk) OR (human_milk)) AND ((india[Title/Abstract]) OR (china[Title/Abstract]) OR (chinese[Title/Abstract]) OR (asia[Title/Abstract]))



8. ábra. Adatbázis bővítés, irodalomkutatás - Ázsia

3.2.2. Rekordok kiválasztása

Az adatbázisbővítésre irányuló irodalomkutatás során talált 190 publikációkból végül 23-at találtunk alkalmasnak adatbázisunk bővítésére. Mivel nagy hangsúlyt fektettünk a dinamikus adatok felkutatására, az adatbázisbővítésünk során a dinamikus adataink száma közel másfél szeresére nőtt (9. ábra).

MilkyBase version	Publikáció	Rekord (Master lap)			Összetevő	DynVal
		Összesen	Statikus	Dinamikus		
Core (1.0.)	**140**	**840**	205	635	**752**	**7666**
Adatbázis bővítés	**23**	**54**	11	43	**71**	**2965**
2.0.	**163**	**894**	216	678	**823**	**10631**

9. ábra. A MilkyBase adatbázisban található rekordok az adatbővítés után

3.3.AZ ANYATEJ ÖSSZETEVŐK IDŐBELI VÁLTOZÁSÁNAK ELEMZÉSE ÉS MODELLEZÉSE

3.3.1. Modellkészítés

Az ontológia létrehozásán túl, az adatbázisban összegyűjtött adatok alapján egy olyan modell elkészítése volt a cél, mely leírja az anyatej komponensek időbeli változásának pályáját az emberi élet kritikus első négy hónapjában, amikor a csecsemők tápanyag igényeinek az ideális esetben elsősorban az anyatejből származik.

Ezek a szempontok vezetnek bennünket ahhoz a modellezési eljáráshoz, amelyet Shenhav & Azad *(Shenhav & Azad, 2022)* tanulmánya, valamint a prediktív mikrobiológia elmúlt három évtizedben végbement fejlődése *(McMeekin és mtsai., 2002)* inspirált:

- Gyűjtsünk adatokat az anyatej összetevőinek longitudinális (azaz időbeli pályák) alakulásáról, lehetőleg egyénileg az anyáktól, egy meghatározott intervallumban (mondjuk a csecsemő első 120 napja alatt), amely alatt a körülmények legalább megközelítőleg állandóak, és a csecsemő tápláléka várhatóan kizárólag anyatej.
- Definiáljuk a matematikai függvények egy olyan osztályát, amely elég általános ahhoz, hogy ezeket az időbeli pályákat reprezentálja. Ezt nevezzük *elsődleges modell*nek; paraméterei a modellnek a megfigyelt adatokra való illesztésével becsülhetők meg.
- Modellezzük a feljegyzett tényezők (pl. a földrajzi elhelyezkedés vagy az anya/gyermek körülményei) hatását a fenti elsődleges modell paramétereire (nem közvetlenül az időbeli pálya mért egyes pontjaira!). Ezt nevezzük *másodlagos modell*nek. Ez lehet csak egy egyszerű minősítő is például, ha az elsődleges paraméterekre gyakorolt pozitív vagy negatív hatásokat jelzi.
- Határozzuk meg a bizonytalanságok várható forrásait. Rangsoroljuk őket, hogy a gyakorlati alkalmazásokkal kapcsolatos döntéshozatalt segítse, például a csecsemőtápszer összetételének optimalizálásával, étrend-kiegészítéssel stb. kapcsolatban.

Tehát a modellezéshez, először az anyatej komponensek időfüggő koncentrációjának mérését kell elvégezni, lehetőleg minél többször. Mivel esetünkben a

publikált adatokra építettük a modellt, így azok kiválasztásánál kritériumaink a következők voltak: az anyatej komponensek mért vagy becsült koncentrációi a vizsgált szoptatási időszakon (ebben a tanulmányban 4 hónap) belül legyenek, kellően hosszú intervallumon keresztül. Egy elsődleges modell kialakításához, ahhoz, hogy bármilyen tendenciát láthassunk, legalább 4-5, de lehetőleg több mint 10, [idő , koncentráció] adatpárra van szükségünk. Az ilyen adatsorok ritkák a szakirodalomban, különösen az egyéni anyák esetében, de természetesen használhatunk származtatott adatokat is (leggyakrabban a vizsgált komponensek homogén kohorszból előállított átlagait és standard eltéréseit).

3.3.2. Matematikai modellezés, szaturációs modell

Elsődleges modellünk a következőképp írható le:

A fókuszintervallum a szülést követő első négy hónap. Anyatejkutatásra specializálódott kutató kollégáinkkal való konzultációk során arra jutottunk, hogy az anyatej komponensek *y(t)* koncentrációjának időbeli változását két fázissal írjuk le;

1. először egy gyors kezdeti lineáris fázis a [0, λ] időintervallumban (kolosztrum),
2. majd egy exponenciálisan konvergens fázis, amelyet egy szaturációs modell ír le:

Ezzel a következő **kétfázisú modellhez** jutottunk:

(1.)
$$y(t)=\begin{cases} y_0 + a \cdot t & (0 \le t < \lambda) \\ y_\lambda \cdot e^{-r\cdot(t-\lambda)} + y_{End}(1 - e^{-r\cdot(t-\lambda)}) & (\lambda \le t) \end{cases}$$

$$ahol\ y_\lambda = y_0 + a\cdot\lambda\ ,\qquad 0\le r,\quad 0 \le \lambda.$$

Jelölések:

- ***y(t)*** : egy anyatej komponens koncentrációja a születéstől mért idő függvényében
- ***t*** : a szülés után eltelt idő (*t=0* a szülés időpontja)

- y_0 : egy anyatej komponens koncentrációja a szülés időpontjában (kezdeti koncentráció)
- a : az anyatej komponens koncentrációjának változási gyorsasága a kezdeti szakaszban
- $y_\lambda = y(\lambda)$: az anyatej komponens koncentrációja a kezdeti fázis végén
- r : szaturációs ráta
- λ : a kezdeti (kolosztrum) fázis időtartama
- y_{End} : az anyatej komponens koncentrációja a vizsgált időszak végén (végső koncentráció)

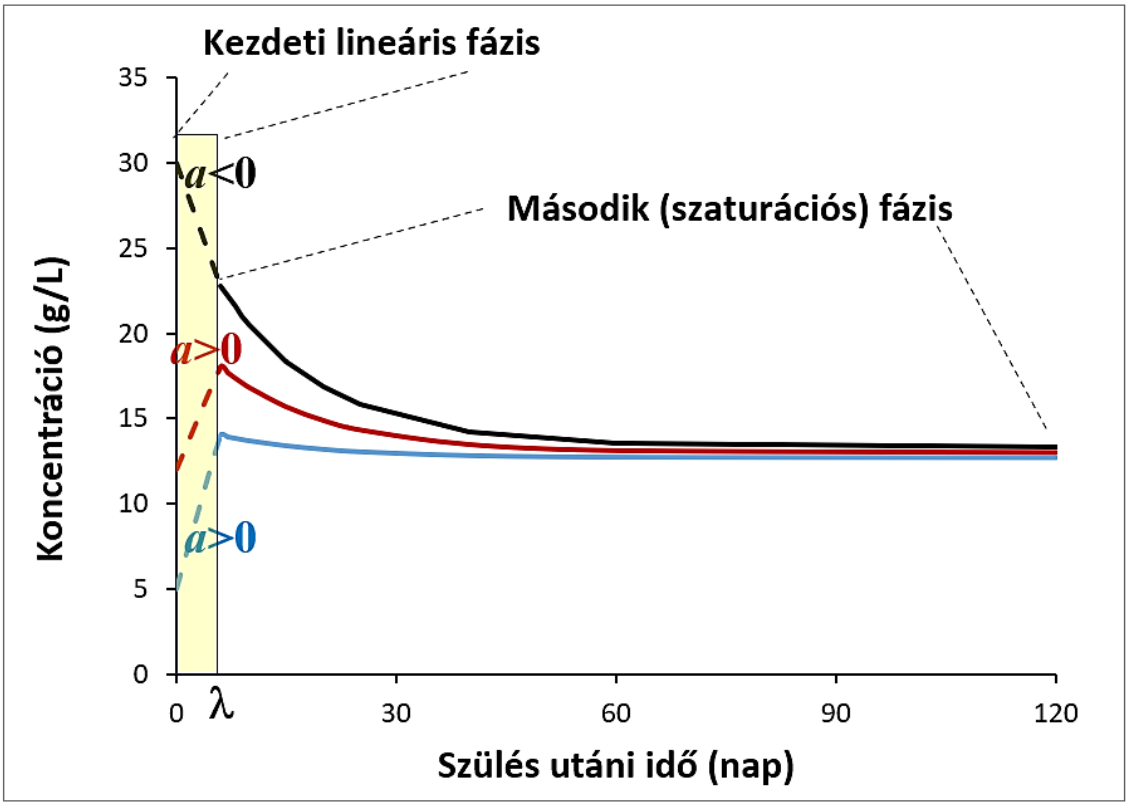
A modell a következőképpen értelmezhető:

1. Az első, úgynevezett kolosztrum fázisban a vizsgált anyatej komponens koncentrációja a kezdeti y_0 szintről indulva a sebességgel, gyorsan változik.
2. Ezután a rendszer egy autonóm fázisba lép, ahol a koncentráció exponenciális sebességgel konvergál egy stacionárius szinthez.

A kezdeti fázis paraméterei y_0 és a, csakúgy, mint a második fázis paraméterei r, és az y_{End} végső koncentrációs szint, számos elsősorban az anyát jellemző tényezőtől függ.

A második, stacionárius állapothoz konvergáló fázist **szaturációs modell**nek nevezzük. Ezt az elnevezést eddig főként olyan helyzetekre alkalmazták, amikor a válaszváltozó értéke az idő előrehaladtával növekszik, de az egyszerűség kedvéért a tükörkép folyamatokra, azaz az idővel csökkenőkre is ezt használjuk most.

Az általános elsődleges modellre a 10. ábra-n mutatunk be három példát, ahol a komponensek koncentrációja a kezdeti kolosztrum időszakot követően csökken. Minden görbe esetében szaturációs sebességként r=0,07/napot használtunk, és λ =6 napban határoztuk meg a kezdeti szakasz (kolosztrum) hosszát. A $\lambda = 0$ eset alternatívája (amikor az a paraméter önmagában jelentéktelenné válik) a $\lambda = 6$ eset, ahol $a < 0$, $a = 0$, vagy $a > 0$ lehet. Hasonlóképpen, a szignifikancia teszt azt mutathatja, hogy az r nullának vehető, ami azt jelenti, hogy $y_0 + a\ \lambda = y_{End}$.



10. ábra. Általános kétfázisú szaturációs modell.

Szaggatott vonal: kezdeti fázis (kolosztrum).

Folyamatos vonal: második (szaturációs) fázis.

A kezdeti lineáris fázis meredeksége a kékkel és pirossal jelölt két alsó görbe esetében, ***a***>0, míg a felső, fekete görbe esetében ***a***<0.

Forrás: *(Baranyi-Pacza- és mtsai., 2024)*

A legalsó, kék görbe alig tér el az ***r***=0 esettől, amely egy olyan **kétfázisú függvényt** eredményezne, ahol a kezdeti, kolosztrum ideje alatti lineáris változást egy állandó ($y_{End}= y_\lambda$) koncentráció követ:

$$y(t)= \begin{cases} y_0 + a \cdot t & (0 \le t < \lambda) \\ y_\lambda & (\lambda \le t) \end{cases} \qquad \textbf{(2.)}$$

$$\text{ahol } y_\lambda = y_0 + a \cdot \lambda\,, \qquad r \ge 0, \quad \lambda \ge 0.$$

A legfelső az λ =0, a=0 esethez hasonlít, amely az **egyfázisú egyszerű szaturációs** modellt képviseli, három paraméterrel.

(3.) $$y(t) = y_0 \cdot e^{-r \cdot t} + y_{End}(1 - e^{-r \cdot t}) \qquad (0 \leq t, r \geq 0)$$

Az **egyfázisú egyszerű szaturációs** modell három paramétere:

1. a kezdeti koncentráció (y_0)
2. és a végső koncentráció, amelyhez a trajektória konvergál, (y_{End})
3. valamint ennek az exponenciális konvergenciának a sebessége (r).

A kezdeti időszakot követően növekvő koncentrációkat mutató változatok tükrözéssel kaphatók.

Az anyatej komponensek koncentrációjának modellezése során az itt leírt szaturációs modellt egy általunk fejlesztett, Visual Basic for Applications nyelven írt MS-Excel kiegészítő programmal illesztettük, melyet közvetlenül a MilkyBase adatbázisból kinyert adatkészletekre futtattunk az Excelben. A változékonyság forrását standard ANOVA-eljárásokkal elemeztük.

Azt, hogy az adathalmaz leírásához szükséges-e a teljes kétfázisú modell négy paraméterrel (y_0, a, r, y_{End}), vagy bármelyik paraméter fix értéknek vehető a modell dimenziójának csökkentése érdekében, F-próba döntötte el. Fontos megemlíteni, hogy λ értékkészletét binárisnak tekintettük, azaz vagy 6 (alapértelmezett), vagy 0 értéket kapott. Az utóbbi esetben az első fázis beágyazódik a szaturációs modellbe, és az eredmény egy egyfázisú, egyszerű szaturációs modell lesz, három paraméterrel. Szintén F-próba döntötte el, hogy elegendő-e az egyfázisú egyszerű szaturációs modell az adott adatsorra való illeszkedéshez, vagy a λ =6 eset a meredekséggel a szignifikáns.

Hasonlóképpen, az F-próbát használtuk annak eldöntésére, hogy egy vagy két paraméter azonosnak tekinthető-e egy adatkészlet-pár vonatkozásában.

A másodlagos modellel számszerűsíthető a különböző tényezők (az anyai előzmények és egyéb jellemzők) hatása az elsődleges modell illesztett paramétereire (pl. y_{End}). A számításokat egy Visual Basic nyelven írt Microsoft Excel Add-In-ben felépített nemlineáris regressziós algoritmussal végeztük, a standard Levenberg-Marquardt-

módszer *(Levenberg, 1944) (Marquardt, 1963)* implementálásával. Az Excel integrált Data Analysis Add-In programját használtuk a lineáris regresszió és ANOVA számolások elvégzésére, 5%-os szignifikancia-szinttel.

4. EREDMÉNYEK

A kutatás eredményeként létrehoztuk a MilkyBase *(Pacza és mtsai., 2022)* adatbázist, mely a tej összetételére vonatkozó rekordokat tárolja összekapcsolt Excel táblázatokban, majd ezeket az adatokat felhasználva matematikai modellezést végeztünk.

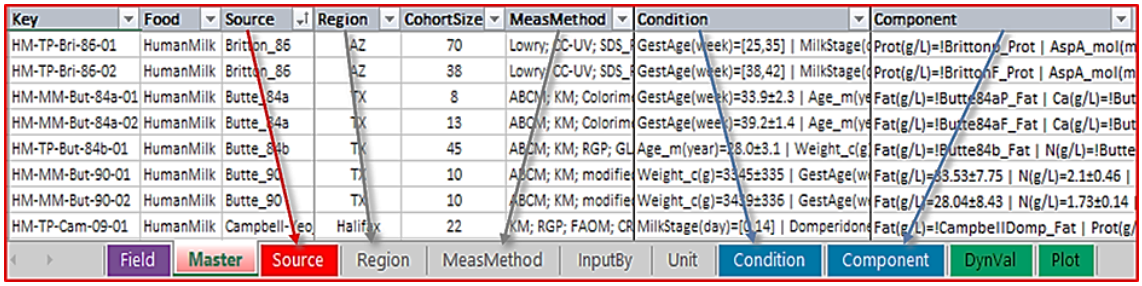
A létrehozott adatbázis egy olyan Excel munkafüzet, amely az adatok rögzítését a nem számítástudományos érdeklődésű táplálkozástudományi szakemberek számára is egyszerűvé teszi. A mezők hierarchikus szervezése biztosítja az adatok elemzésére szolgáló statisztikai és grafikai módszerek programozhatóságát. Természetesen jóval haladottabb adatbázis kezelő szoftverek is léteznek (pl. SQL) mint az összekapcsolt Excel munkalapok, azonban alkalmazás-centrikus céljainknak megfelelően olyan eszközt kívántunk fejleszteni, ami közel van a várható felhasználók számítástudományi ismereteihez, ezzel lehetővé téve, hogy az adatbázis ontológiáját a saját igényeikhez igazíthassák. Ezt szolgálja az is, hogy az adatbázis és leírása, valamint a hozzá készített Excel makrók, mind szabadon és díjmentesen hozzáférhetők és szükség szerint módosíthatók.

Valamint egy, – az anyatej komponensek időbeli változását leíró - prediktív modell-struktúrát alkottunk, melyet ezután a MilkyBase adatbázisban már rögzített – a tudományos irodalomban rendelkezésre álló, publikált - adatokra illesztettünk. A struktúra alkalmasnak bizonyult mind egyedi molekulák (például specifikus zsírsav-, oligoszacharid- és ásványi anyag molekulák), mind molekulacsoportok (például összes fehérje, teljes zsírtartalom) időbeli trajektóriáinak a leírására is.

4.1.A MILKYBASE ADATBÁZIS

4.1.1. A MilkyBase adatbázis felépítése

Az adatbázis építési elveinket követve a létrehozott MilkyBase adatbázis 11 összekapcsolt táblázat rendszere, egyetlen Microsoft Excel munkafüzetbe rendezve (11. ábra).



11. ábra. A MilkyBase adatbázis összekapcsolt táblázatai

Az adatbázis munkalapjai 2 fő csoportba sorolhatók: Fő munkalap és Definíciós lapok. A fő munkalap rekordjait egy egyedi kulcs azonosítja, és a rekord egyes mezőinek lehetséges értékeit -a fejlécnek megfelelő- azonos nevű definíciós lapok tárolják.

1. **Fő munkalap**
   a. Master

2. **Definíciós lapok.**
   a. Field
   b. Source
   c. Region
   d. MeasMethod
   e. InputBy
   f. Unit
   g. Condition
   h. Component
   i. Dynval
   j. Plot

A fő munkalap szolgál a publikált mért adatok rögzítésére a keletkezés és publikálás körülményeivel együtt, míg a definíciós lapok funkciója a fő munkalap bejegyzéseivel kapcsolatos részletek definiálása, a mezők szintaxisa és leírása ezeken lapokon követhető. Ezeket a definíciókat használja az formai ellenőrzésekre használt

"Syntax check" makró is. A MilkyBase adatbázis tábláinak kapcsolódási sémáját a Melléklet 1. Kiegészítő információs táblája tartalmazza.

4.1.2. Fő munkalap

Az adatbázis fő munkaalapja, a ***Master*** (Elsődleges).[1] Munkalap (12. ábra), mely egyedi kulcsokkal azonosított rekordokból áll. Az ***Master*** munkalap mezőit három fő csoportba sorolhatjuk: *Adminisztratív mezők*, *Magyarázó változók mezői* és *Válaszváltozók mezői.* Az *Adminisztratív mezők* a publikált mérési adatok adatbázisba történő rögzítésével kapcsolatos, specifikus adminisztratív részleteket tartalmaznak, a *Magyarázó változók* mezői az adatok keletkezésének körülményeit tartalmazzák míg az anyatej komponenseinek mért adatait a *Válaszváltozók* mezői. Ez a terminológia a matematikában használt függő és független változók terminológiához hasonlít, ahol a függő változó (esetünkben a Válaszváltozó) értéke a független változó (Magyarázó változó) hatására változik.

1	Key	Food	Source	Region	Cohort	MeasMethod	Condition	Component	Comment
325	HM-TP-Gax-14-01	HumanMil	Gax_14	Baja	108	AAS	MilkStage(day)=[7,10] | StorageT	Hg(g/L)=2.52E-6@[3.00E-8,2.48	Component: Mean@[Min,Max]
326	HM-TP-Gom-17-01	HumanMil	Gomez-Gallego_17	Spain	10	HPLC	MilkStage(day)=30	Putr_mol(mol/L)=3.69E-7 | Sprd_mol(mol/L)=4.012E-6 | Sper_m	
327	HM-TP-Gom-17-02	HumanMil	Gomez-Gallego_17	Spain	10	HPLC	OtherAtBirth_c=CesareanDelivery	Putr_mol(mol/L)=3.176E-6 | Sprd_mol(mol/L)=5.031E-6 | Sper_	
328	HM-TP-Gom-17-03	HumanMil	Gomez-Gallego_17	Finland	10	HPLC	MilkStage(day)=30	Putr_mol(mol/L)=0 | Sprd_mol(mol/L)=6.086E-6 | Sper_mol(mo	
329	HM-TP-Gom-17-04	HumanMil	Gomez-Gallego_17	Finland	10	HPLC	OtherAtBirth_c=CesareanDelivery	Putr_mol(mol/L)=0 | Sprd_mol(mol/L)=5.20E-6 | Sper_mol(mol/	
330	HM-TP-Gom-17-05	HumanMil	Gomez-Gallego_17	SouthAfrica	8	HPLC	MilkStage(day)=30	Putr_mol(mol/L)=4.48E-7 | Sprd_mol(mol/L)=3.628E-6 | Sper_m	
331	HM-TP-Gom-17-06	HumanMil	Gomez-Gallego_17	SouthAfrica	10	HPLC	OtherAtBirth_c=CesareanDelivery	Putr_mol(mol/L)=0 | Sprd_mol(mol/L)=3.329E-6 | Sper_mol(mo	
332	HM-TP-Gom-17-07	HumanMil	Gomez-Gallego_17	China	10	HPLC	MilkStage(day)=30	Putr_mol(mol/L)=4.54E-7 | Sprd_mol(mol/L)=3.357E-6 | Sper_m	
333	HM-TP-Gom-17-08	HumanMil	Gomez-Gallego_17	China	10	HPLC	OtherAtBirth_c=CesareanDelivery	Putr_mol(mol/L)=1.93E-7 | Sprd_mol(mol/L)=3.234E-6 | Sper_m	
334	HM-TP-Hig-82-01	HumanMil	Higashi_82	Japan	65	FAAS	Age_m(year)=27.3@[21,37] | we	Cu(g/L)=!Hig_Cu | Zn(g/L)=!Hig_Zn	
335	HM-TP-Jag-20-01	HumanMil	Jagodic_20	Koper	36	CGC	MilkStage(day)=[35,77] | Age_m(	C18:1n-9/FAc(-)=0.322±0.0344 | C18:3n-3/FAc(-)=0.0071±0.002	
336	HM-TP-Jag-20-02	HumanMil	Jagodic_20	Pomurje	38	CGC	MilkStage(day)=[35,77] | Age_m(	C18:1n-9/FAc(-)=0.293±0.0583 | C18:3n-3/FAc(-)=0.0101±0.004	
337	HM-TP-Jan-81-01	HumanMil	Jansson_81	Malmo	34	HPLC	MilkStage(day)=[4,150] | Storage	AlphaTocoph_mol(mol/L)=!Jans_AlphaTocoph | BetaTocoph_mo	
338	HM-TP-Joh-19-001	HumanMil	John_19	TX	1	FTIR	Age_m(year)=27 | MilkStage(day)	Prot(g/L)=!John_001_Prot | LAC(g/L)=!John_001_LAC | Fat(g/L)=	
339	HM-TP-Joh-19-002	HumanMil	John_19	TX	1	FTIR	Age_m(year)=33 | MilkStage(day)	Prot(g/L)=!John_002_Prot | LAC(g/L)=!John_002_LAC | Fat(g/L)=	
340	HM-TP-Joh-19-003	HumanMil	John_19	TX	1	FTIR	Age_m(year)=36 | MilkStage(day)	Prot(g/L)=!John_003_Prot | LAC(g/L)=!John_003_LAC | Fat(g/L)=	

Field | Master | Source | Region | MeasMethod | Condition | Component | InputBy | Unit | DynVal | Plot | +

12. ábra. A Master (Elsődleges) munkalap.

A ***Master (****Elsődleges)* munkalap egy-egy rekordja a következő mezőket tartalmazza:

Key | ***Food*** | ***Source*** | ***Region*** | ***CohortSize(cap)*** | ***MeasMethod*** | ***Condition*** | ***Component*** | ***Comment***

[1] Leírásban a dőlt, félkövérrel szedett, nagybetűs kezdőbetűkkel szedett szavak a munkafüzet ***Lapjaira***, a dőlt betűs `Courier` típusúak pedig a munkafüzet `mezőire` utalnak.

4.1.2.1. Adminisztratív mezők:

- *Key (Kulcs)*: A ***Master*** (Elsődleges) lap rekordjainak egyedi azonosítója.

Az alap MilkyBase adatbázis rekordjainak rögzítése során a kulcsok definiálására és egyediségének megtartására az alábbi metódust definiáltuk, mely szabályokat alkalmazva a felhasználók hozzák létre a kulcsokat manuálisan, egyediségüket a „Syntax Check" makró segítségével ellenőrizve:

A kulcs 5 részből épül fel kötőjelekkel (-) elválasztva. AA-BB-Ccc-xx-yy, ahol

- **AA** – A rekordban rögzített élelmiszer kódja. Jelenleg az anyatejre vonatkozó adatokat tartalmazza az adatbázis, ezért a HM (HumanMilk) anyatej rövidítést használtuk, szem előtt tartva azt a lehetőséget, hogy későbbiekben a szarvasmarha-, növényi alapú tej stb. összetétele is rögzíthető legyen ugyanebben az adatbázisban)
- **BB** - A rekordot rögzítő személy nevének kezdőbetűi (pl. JS John Smith számára). Ez a módszer több ember közös munkája során az adatbázisban megkönnyítette a kétszintű ellenőrzést, lehetővé téve, hogy ugyanazt a rekordot a rögzítőn kívül más ellenőrizhesse.
- **Ccc** – A rekordban rögzített adatokat tartalmazó publikáció első szerzőjének vezetéknevének első három betűje. (pl. Dan Daniels első szerzőnév esetében)
- **xx** - A forráspublikáció megjelenése évének utolsó 2 számjegye. (pl. -04- 2004-et jelöli)
- **yy** – A forráspublikáción belüli azonosítószám (pl. 01, 02...) az egyedi kulcs biztosítására, mivel egy publikációból több különböző feltételmezővel rendelkező rekord is rögzíthető, például különböző régiókban végzett mérések, különböző szülési módokra (normál, császármetszés). (pl. HM-JS-Dan-04-01, HM-JS-Dan-04-02...)

- *Food (Élelmiszer):* A vizsgált élelmiszer neve vagy kategóriája.
- *Source (Forrás):* A rekord által tárolt információ forrása. Ez a rövidítés a ***Source*** (***Forrás)*** lapon kerül definiálásra.

4.1.2.2. Magyarázó változók mezői:

- *Region (Régió):* A kohorsz földrajzi régiója. A (faszerkezetű) értékkészlet a ***Region*** lapon van definiálva.
- *CohortSize(cap) (Kohorszméret(fő)):* A rekordban rögzített vizsgálat kohorsz mérete, melyet általában a mintaméretnek is tekintünk, amikor az összetevőkre vonatkozó átlagokat veszünk.
- *MeasMethod (Mérési Módszer):* A rekordban rögzített kísérlet során az anyatej összetevők mérésére alkalmazott analitikai módszerek, illetve eszközök részletei.
- *Condition (Feltétel):* (magyarázó vektorváltozó) Függőleges vonalakkal (|) elválasztott különböző körülmények adatait tartalmazó vektormező, mely az egész vektort egyetlen cellában egyesíti. Az anyatej keletkezésének körülményeit leíró mezők (pl. a gyermek súlya vagy az anya bizonyos életmódbeli jellemzői) lehetséges szabályos értékei a ***Condition*** lapon definiáltak.

4.1.2.3. Válaszváltozók mezői:

- *Component (Összetevők):* (válasz vektorváltozó) A Condition mezőhöz hasonlóan egy olyan vektormező, melynek vonalakkal (|) elválasztott bejegyzései a rekordban rögzített élelmiszer biokémiai összetevőire (pl. B1-vitamin vagy tiamin) vagy több összetevőt tartalmazó csoportra (pl. zsírok csoportja) közölt számszerűsített (alapértelmezés szerint g/liter-ben megadott) adat. Az értékei lehetnek a korábban bemutatott "kiterjesztett numerikus" értékek, valamint dinamikus értékek is, a lehetséges értékkészleteket és az összetevők csoportosítását (azaz a faszerkezet vázát) az ***Összetevő*** lapon definiáljuk.

A ***Master*** (Elsődleges) munkalap mezői közül minden egyes rekord esetében az információ forrásának, a mérés földrajzi régiójának, a kohorsz méretének, az anyatejben található komponenseket mérő analitikai módszernek, valamint legalább egy feltételnek és legalább egy válaszváltozó értéknek a kitöltése kötelező.

4.1.3. Definíciós lapok

A definíciós lapok funkciója, hogy további részleteket jelenítsen meg a ***Master*** (Elsődleges) munkalap bejegyzéseivel kapcsolatosan. Minden egyes definíciós lap neve megjelenik a lap első oszlopának első cellájában is, elősegítve az ellenőrző makrók működését. Az első oszlopban alfanumerikus karakterláncok formájában tárolt változók vannak, amelyek az adott lap által meghatározott típusú rekordokat azonosítják, és melyeket rekord további mezőiben rögzített adatok definiálnak.

A lapok második oszlopában további értelmező elnevezések (*Alternative)* is hozzáadhatók, melyek a különböző publikációkban eltérő írásmóddal, illetve névvel szereplő alternatívákat rögzítik. Ily módon elkerülhető, hogy ugyanaz a rekord többször szerepeljen az adatbázisban különböző név alatt, illetve az adatelemzés is ezáltal válik elvégezhetővé. A változók pontos definiálására A Field, Condition és Component lapokon a *ValueSet, Unit*, *TYPE* mezők, míg további szabadszöveges leírására a *Description* mező nyújt lehetőséget.

A csoportosításhoz szükséges csomópont nevét a lap harmadik mezője (*Group)* tartalmazza. Itt tárolható annak a csoportnak a neve, amelyhez a bejegyzés közvetlenül tartozik. Ezen csoportok segítségével végezhető el a fastruktúra adatbázisba való átemelése, azaz egy-egy csoport a fa csomópontjait, elágazásait definiálja, fa kiindulási pontja („gyökere”) maga a vizsgált élelmiszer (jelenleg tehát az anyatej), míg a molekuláris összetevők a fastruktúra végpontjai („levelek”).

Az alap MilkyBase adatbázis definíciós lapjai a következőképp kerültek publikálásra.

4.1.3.1. Field lap

A ***Field*** (mezők) lap (13. ábra) az adatbázis átfogó keretéhez tartozó definíciókat tartalmazza, elengedhetetlen a szintaktikai ellenőrzéshez.

Field	Alternative	Group	Description	ValueSet	Unit	TYPE	Comment
Key		Field	Unique record ID in the Master sheet			Alphanumeric	Unique one word (can be an integer, to
Food		Field	Food name	HumanMilk; BovineMilk; InfantFormula		Alphanumeric	*Likely to be extended, in which case Valu*
Source		Field	Source of information	_Source		Alphanumeric	Syntax checked like the Master sheet
Region		Field	Geographic region of cohort	_Region		Alphanumeric	Tree-structured definition sheet
CohortSize		Field	Cohort size	[1, 99999]	cap	Numeric	
MeasMethod	MeasurementMethod	Field	Method of collecting data	_MeasMethod		Alphanumeric	Tree-structured definition sheet
Condition		Field	Various conditions and history around birth	_Condition		Alphanumeric	Tree-structured definition sheet. When
Component		Field	Food component per-volume concentration or rela	_Component		Alphanumeric	Tree-structured definition sheet. When
Comment		Field	Comment			FreeText	
Title		Field	Title of paper / web-doc etc			FreeText	
SourceDOI		Field	DOI adress of paper web-doc			FreeText	
PublDate		Field	Year of publication	[1950,2025]	year	Numeric	
InputDate		Field	Year of input	[2020,2025]	year	Numeric	
InputBy		Field	Responsible for inputting the record	_InputBy		Alphanumeric	Simple definition sheet containing the re
Comment/Summary		Field	Comment to source record			FreeText	

13. ábra. A Field (Mezők) munkafüzetlap

4.1.3.2. Source (Forrás) és InputBy (Rögzítette) lap

A ***Source*** (Forrás) lap (14. ábra) az adatbázisban rögzített adatok forrásaival kapcsolatos információkat (forrás címe, DOI száma, a publikálás és az adatbázisba rögzítés éve, valamint a rekordért felelős azonosítója melyet az ***InputBy*** *(Rögzítette)* lapon (15. ábra) definiálunk tartalmazza. Az első mezőben található forrás kulcs, mely a publikáció első szerzőjének nevéből és a publikálás évéből generált, egyértelműen összeköti a ***Master*** *(Elsődleges)* lap megfelelő rekordjaival. Az alap MilkyBase adatbázis 140 forrás adatait tartalmazza.

	Source	Title	SourceDOI	PublDate(year)	InputDate(year)	InputBy	Comment/Summary
2	Ahmed_04	Antioxidant Micronutrient Profile (Vitan	doi.org/10.1093/tropej/50.6.357	2004	2020	UD	
3	Akinbi_10	Alterations in the host defense properti	doi.org/10.1097/MPG.0b013e3181e07f0a	2010	2020	UD	The paper is on studying bacterial growth in frozen
4	Ala-Houhala_88	25-Hydroxyvitamin D and Vitamin D in H	doi.org/10.1093/ajcn/48.4.1057	1988	2020	UD	
5	Anderson_83	Length of Gestation and Nutritional Con	doi.org/10.1093/ajcn/37.5.810	1983	2020	UD	
6	Antonakou_10	Breast milk tocopherol content during t	doi.org/10.1007/s00394-010-0129-4	2010	2020	UD	
7	Aparicio_20	Human Milk Cortisol and Immune Facto	doi.org/10.1371/journal.pone.0233554	2020	2020	UD	
8	Arnold_87a	Protein, Lactose and Fat Concentration	doi.org/10.1111/j.1440-1754.1987.tb0027	1987	2020	UD	
9	Atkinson_80	Macro-mineral Content of Milk Obtaine	doi.org/10.1016/0378-3782(80)90003-1	1980	2020	UD	
10	Atkinson_81	Human Milk Feeding in Premature Infan	doi.org/10.1016/s0022-3476(81)80275-2	1981	2020	UD	
11	Atkinson_87b	Vitamin D activity in maternal plasma a	doi.org/10.1016/S0271-5317(87)80171-9	1987	2020	UD	
12	Azeredo_08	Retinol, carotenoids, and tocopherols in	doi.org/10.1016/j.nut.2007.10.011	2008	2020	UD	
13	Bao_07	Simultaneous Quantification of Sialylolig	doi.org/10.1016/j.ab.2007.07.004	2007	2020	UD	
14	Barrera_18	The Impact of Maternal Diet during Pre	doi.org/10.3390/nu10070839	2018	2020	UD	
15	Bauer_11	Longitudinal Analysis of Macronutrients	doi.org/10.1016/j.clnu.2010.08.003	2011	2021	UD	
16	Beer_20	The Effect of Physical Activity on Humai	doi.org/10.1089/bfm.2019.0292	2020	2020	UD	Effect of moderate- to high-intensity PA on human
17	Britton_86	Milk Protein Quality in Mothers Deliveri	doi.org/10.1097/00005176-198601000-00	1986	2021	UD	
18	Butte_84a	Longitudinal Changes in Milk Compositic	doi.org/10.1016/0378-3782(84)90096-3	1984	2021	UD	

14. ábra. A Source (Forrás) munkafüzetlap

1	InputBy	Alterna	Group	Description	Commer
2	UD	University	InputBy	University of Debrecen, Hungary	
3					
4					
5					

15. ábra. Az InputBy (Rögzítette) munkalap

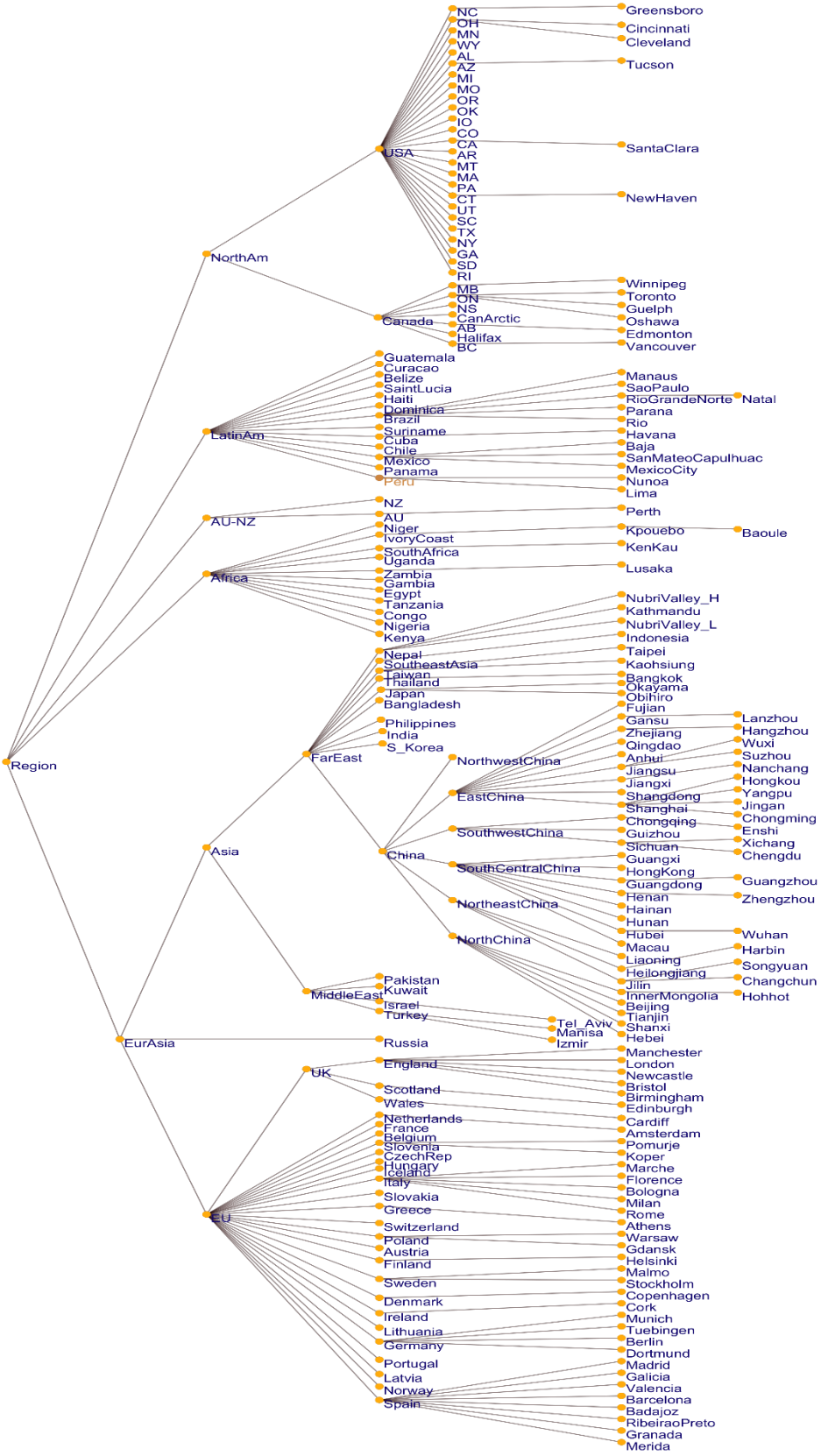
4.1.3.3. Region (Földrajzi hely) lap

A Region (Földrajzi hely) lap (16. ábra) a publikált adatok kohorszainak földrajzi helyzetére vonatkozó adatokhoz szolgáltat hátteret, a fastruktúrájú csoportosítással. Az adatok ilyen struktúrában való rögzítése elősegíti azon kutatások elvégzését melyek egy-egy meghatározott régióra, vagy országra vonatkoznak. Ilyen típusú vizsgálat a dolgozatomban későbbiekben bemutatott is (lsd 4.3.2.1 Földrajzi különbségek.).

Az alap adatbázisunk 241 csomópontot tartalmaz, melyből 69 országot és 123 várost fed. A földrajzi helyek hierarchikus csoportosítása a 17. ábra-n látható.

1	Region	Alternative	Group	Description	Comment
2	AB	Alberta	Canada	Province in Canada	
3	Africa		Region	Continent	
4	AL	Alabama	USA	State in the US	
5	Amsterdam		Netherlands	Capital of Netherlands	
6	Anhui		EastChina	Province in China	
7	AR	Arkansas	USA	State in the US	
8	Asia		EurAsia	Part of EurAsia	
9	Athens		Greece	Capital of Greece	
10	AU	Australia	AU-NZ	Australia, part of AU-NZ	
11	AU-NZ		Region	Australia and NewZeland	
12	Austria		EU	Country in Europe	
13	AZ	Arizona	USA	State in the US	
14	Badajoz	Badajos	Spain	City in Spain, formerly written Badajos	
15	Baja	Baja_California_Sur	Mexico	State in Mexico	
16	Bangkok		Thailand	Capital of Thailand	
17	Bangladesh		FarEast	Country in the FarEast	
18	Baoule		Kpouebo	Ethnic group in Kpouebo	
19	Barcelona		Spain	City in Spain, Catalonia	
20	BC	British_Columbia	Canada	Province in Canada	
21	Beijing		NorthChina	Capital of the People's Republic of China	

16. ábra. A Region (Földrajzi hely) lap



17. ábra. Region (Földrajzi hely) fastruktúra

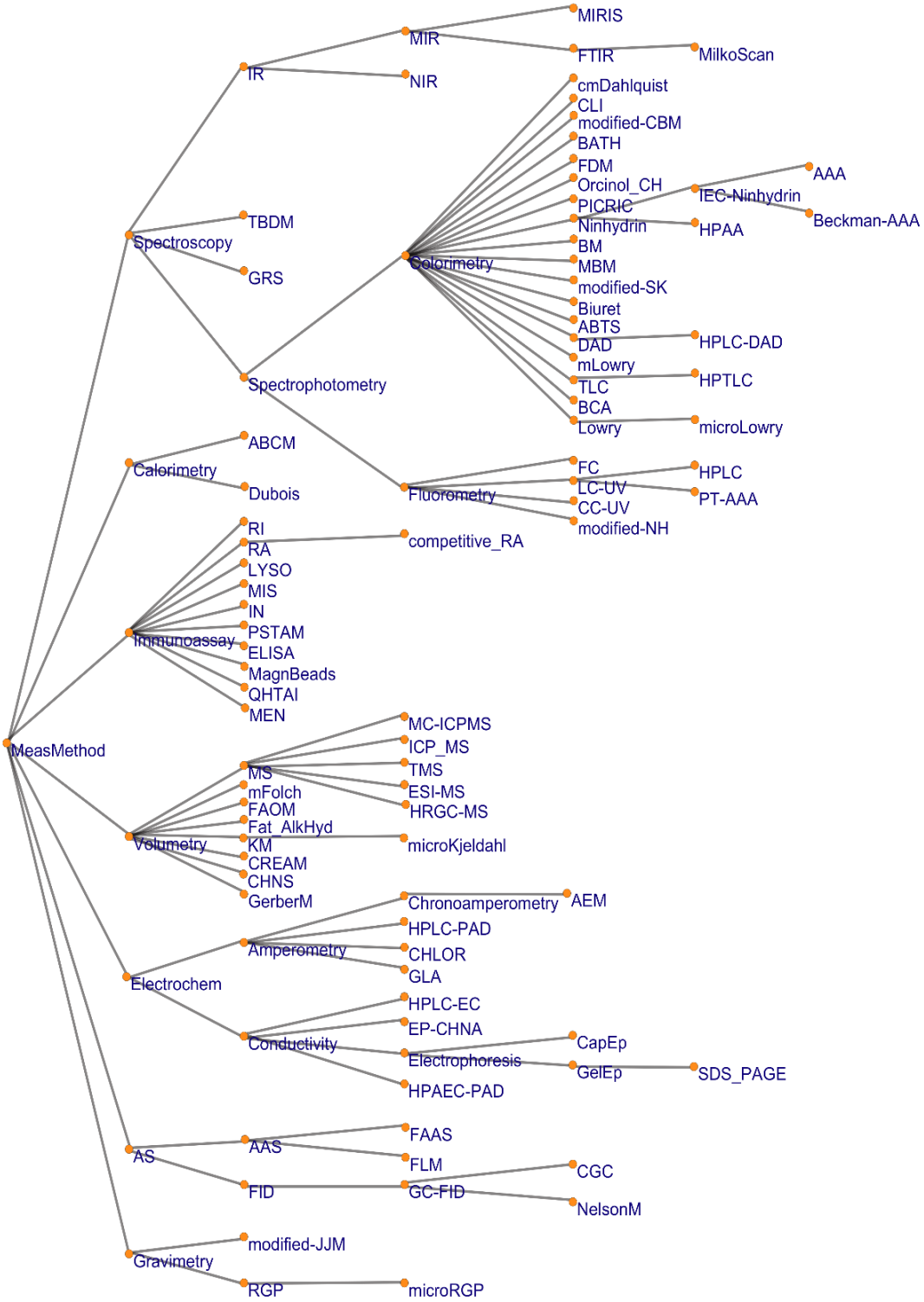
4.1.3.4. MeasMethod (Mérési módszerek) lap

Az összetevők mennyiségi meghatározását a kutatási metódusok is befolyásolják. Függhet az anyatej mintagyűjtés módjától (manuális vagy gépi), a minták tárolásától és tartósításától (hűtés, fagyasztás/olvasztás, pasztőrözés hossza, módja, hőmérséklete). *(Kim és mtsai., 2019)* Ezeken túl az összetevők mennyiségi meghatározását befolyásolhatják a minták előkészítése *(Fusch és mtsai., 2015)*, a feldolgozásukra használt analitikai eljárások is. Nagyon fontos a megfelelő analitikai módszer kiválasztása a vizsgált összetevő kvantitatív mérésére, különösen az összehasonlító vizsgálatokban, hiszen az eltérések nemcsak a környezeti feltételekből fakadhatnak, hanem módszertani torzításokat és hibákat is tükrözhetnek. *(Samuel és mtsai., 2020)*

Adatbázisunkban a MeasMethod lap tartalmazza a mennyiségi elemzésekre használt műszeres analitikai módszereket (18. ábra), melyek hierarchikus csoportosítása a 19. ábra-n látható.

1	MeasMethod	Alternative	Group	Description	Comment
2	AAA		IEC-Ninhydrin	JLC-500V amino acid analyzer	https://doi.org/10.1016/j.jtemb.2005.05.001
3	AAS		AS	Atomic absorption spectrometry (AAS) is a very sensitive method of e	belongs to Spectometry. https://doi.org/10.101
4	ABCM		Calorimetry	Adiabatic Bomb Calorimeter	https://www.researchgate.net/publication/296
5	ABTS		Colorimetry	2-20-azinobis 3-ethylbenzothiazoline-6-sulfonic acid radical cation assa	https://doi.org/10.1080/10715760500404805
6	AEM		Chronoamperometry	Amperometric enzymatic method to analyze lactose content	https://doi.org/10.1016/j.jtemb.2005.05.001
7	Amperometry		Electrochem	Amperometry is an electroanalytical technique that involves the application of a constant reducing or oxidizing potent	
8	AS		MeasMethod	Atomic spectroscopy (including atomic absorption spectrometry, atomic emission spectrometry, and atomic fluoresce	
9	BATH		Colorimetry	Bathophenanthroline Method	DOI: 10.1016/s0899-9007(02)00813-4
10	BCA		Colorimetry	Bicinchoninic acid (BCA) assay or Smith assay is a copper-based colorir	https://www.sciencedirect.com/science/article
11	Beckman-AAA		IEC-Ninhydrin	Beckman 121 Amino Acid Analyser	https://doi.org/10.1093/ajcn/40.5.1042
12	Biuret		Colorimetry	The biuret method is a colorimetric technique specific for proteins an	https://biocheminsider.com/biuret-test/#:~:tex
13	BM		Colorimetry	Bradford assay, a colorimetric protein assay	https://doi.org/10.1016/0003-2697(76)90527-3
14	Calorimetry		MeasMethod	Calorimetry is the process of measuring the amount of heat released or absorbed during a chemical reaction. By knov	
15	CapEp		Electrophoresis	Capillarity of narrow bore tube is employed to separate the samples based on their size: charge ratio.	
16	CC-UV		Fluorometry	Column chromatography works on a much larger scale by packing the	DOI: 10.1055/s-2008-1033924
17	CGC		GC-FID	Capillary Gas Chromatography	belongs to GC: https://application.wiley-vch.de/
18	CHLOR		Amperometry	Chloridometer	https://doi.org/10.1203/00006450-198202000-
19	CHNS		Volumetry	Multiple analysis options including Carbon (C), Hydrogen (H), Nitrogen (N), Sulfur (S) and Oxygen (O) (CHN or CHNS an	
20	Chronoamperometry		Amperometry	Chronoamperometry	
21	CLI		Colorimetry	Chemi-Luminescenc Immunoassay	https://doi.org/10.1016/B978-012214730-2/50

18. ábra. A MeasMethod (Mérési módszerek) lap



19. ábra. Az anyatej összetevők mérésére alkalmazott módszerek fastruktúrája

4.1.3.5. Unit (Mértékegység) lap

A ***Master*** (Elsődleges) adatlapon rögzített mért adatok esetében minden esetben megadásra került a mértékegység is melyet a ***Unit*** (Mértékegység) lapon (20. ábra) definiálunk. A mért adatok elemzésének megkönnyítése érdekében az eltérő módokon publikált azonos típusú adatokat a mértékegységek átváltásával rögzítettük. Az átváltásokhoz a ***Unit*** (Mértékegység) lapon definiáltuk a fő csoportokat alapegységükkel együtt (az arány (-), az idő (nap), a hossz (mm), a tömeg (g), a térfogat (L), a dózis (g/nap), a sűrűség (g/l) arány), valamint megadtuk az átváltási arányszámokat (`FaktorToBase`) minden használt mértékegységre. Azokban az esetekben, amikor az értékek g, mg stb.-ban voltak 100 g ehető élelmiszerre vonatkoztatva (pl. mg/g), akkor azokat g/literre (vagy g/l-re) váltottuk át, ahol 1 liter tejet 1 kg-nak feltételeztünk.*(NIST, 2022)*

1	Unit	Alternative	Grou	Description	Comment	FactorToBase	Basic_Unit
2	ratio		Unit				
3	-		ratio	dimensionless	mostly the proportion	1	-
4	time		Unit				
5	min		time	minute	=1/60*24 day	0.000694444	day
6	hour		time	hour	=1/24 day	0.041666667	day
7	day		time	day	basic unit of time	1	day
8	week		time	week	= 7 days	7	day
9	month		time	month	= 30 days	30	day
10	year		time	year	= 365 days	365	day
11	length		Unit				
12	m		length	Metre	= 1.e3 mm	1000	mm
13	cm		length	centimetre	= 10 mm	10	mm
14	mm		length	milliMetre	basic unit of length	1	mm
15	mass		Unit				
16	kg		mass	kilogram	= 1000 gram	1.E+03	g
17	g		mass	gram	basic unit of mass	1	g
18	mg		mass	milliGram	=10^(-3) gram	1.E-03	g
19	mcg		mass	microGram	=10^(-6) gram	1.E-06	g
20	ng		mass	nanoGram	=1.e-9 gram	1.E-09	g
21	pg		mass	picoGram	=1.e-12 gram	1.E-12	g
22	mol		mass	The mass of one mole of a chemical compound, in gram, is numerically equal to	basic unit of mole-mas	1	mol
23	mMol		mass	milliMole	=1.e-3 mol; mcM (mi	1.E-03	mol

20. ábra. A Unit (Mértékegység) lap

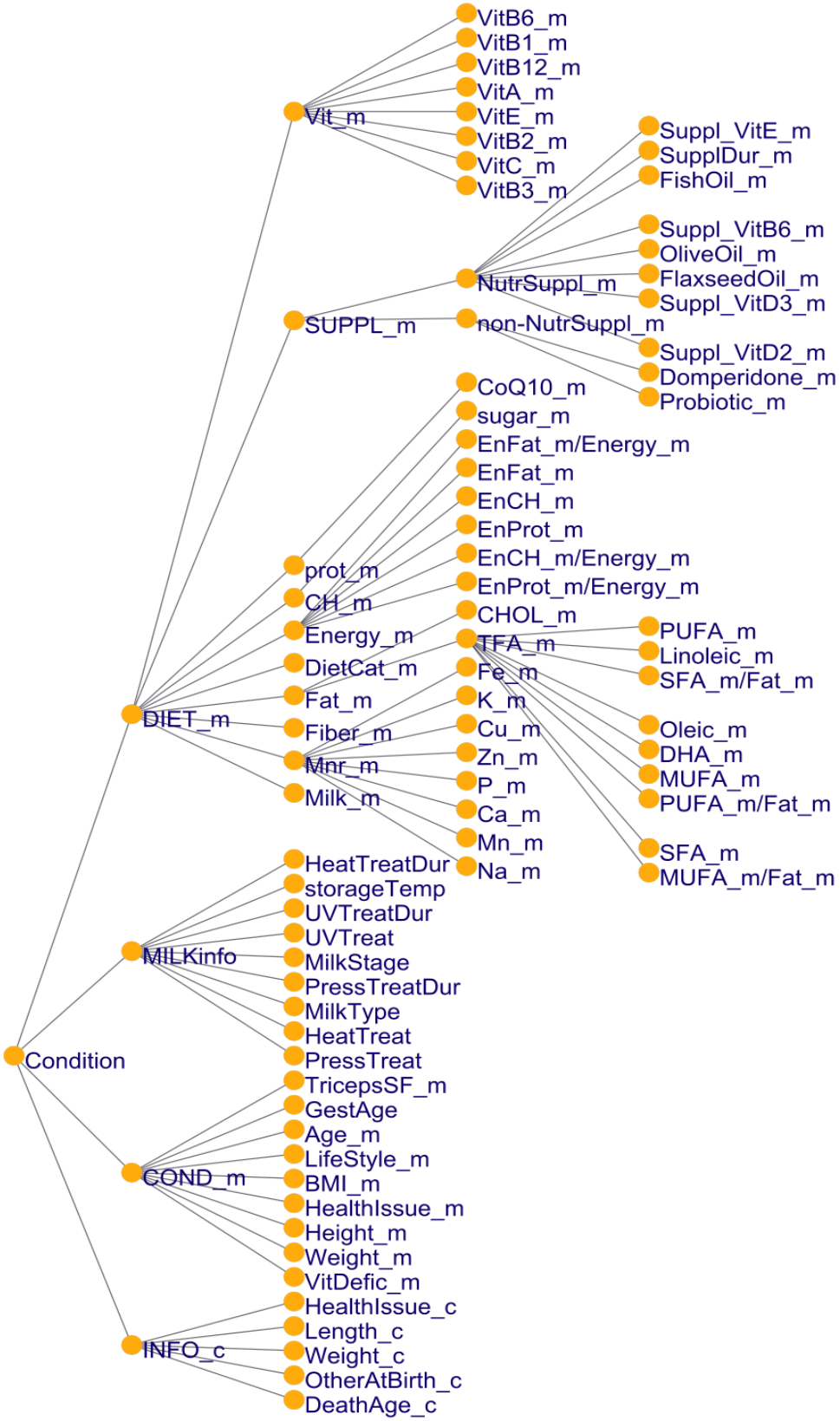
4.1.3.6. Condition (Feltétel) lap

A ***Condition*** (Feltétel) magyarázó változó definíciós értékeit tartalmazó tábla (21. **ábra**), mely értékek az anyatej összetételét befolyásoló tényezők. Ezek vonatkozhatnak az anyára (pl. terhességi kor, testsúly, életmód, diéta, gyógyszeres kiegészítés), a csecsemőre születéskor (pl. súly, egészségügyi probléma, hossz) és a tejmintára (pl. pasztőrözés, laktációs stádium, tárolási hőmérséklet). Az adatok sokféleségét befolyásolja azoknak a körülményeknek a fontossága, amelyekről a publikációk beszámolnak. Például az anyatej -összetételt ritkán vizsgálják az újszülött nemének függvényében, ezért az adatbázisban nem definiáltunk külön mező erre a magyarázó változóra, de a nemet rögzíteni lehet a OtherAtBirth_c változó segítségével, amely a releváns csecsemőjellemzőket tartalmazza.

	Condition	Alternative	Group	Description	ValueSet	Unit	TYPE	Comment
2	**COND_m**	**Condition_mother**	**Condition**	**Various info on the mother's condition**				
3	**GestAge**	Gestational_Age	**COND_m**	Gestational age refers to the 'age'	[0, 50]	week	Numeric	
4	**Age_m**	AgeAtBirth_mother	**COND_m**	Mother's (or cohort's mean-) age a	[10, 60]	year	Numeric	
5	**BMI_m**	BMI_mother	**COND_m**	Mother's BMI. Default: Pre-pregna	[5, 50]	kg/m2	Numeric	Formula: BMI=mass/(height^2)(kg/m^2). Can be dyn
6	**Height_m**	Height_mother	**COND_m**	Mother's height	[50, 250]	cm	Numeric	Can be dynamic.
7	**Weight_m**	PrePregnancyWeight_	**COND_m**	Weight of Mother. Default: pre-pr	[0, 200]	kg	Numeric	Can be dynamic.
8	**TricepsSF_m**	Triceps_Skinfold_moth	**COND_m**	Mother's triceps skinfold thickness	[0, 1e4]	mm	Numeric	Can be dynamic.
9	**HealthIssue_m**	HealthIssue_mother	**COND_m**	Issue with the Mother's health	allergy; asthma; IBD; CVD; HIV		AlphaNumeric	
10	**LifeStyle_m**	LifeStyle_mother	**COND_m**	Mother's lifestyle features	smoker; physAct		AlphaNumeric	Smoker: This could be numeric like number of cigare
11	**VitDefic_m**	Vitamin_deficiency	**COND_m**	Deficiency of vitamins for mother	A; B1; B2; B3; B5; B6; B12; C; D; E; K		AlphaNumeric	
12	**INFO_c**	**INFO_child**	**Condition**	**Info on child at birth**				
13	**Length_c**	Length_child	**INFO_c**	Length of child	[10, 1e2]	cm	Numeric	Can be dynamic (<14 yrs old)
14	**Weight_c**	Weight_child	**INFO_c**	Child's weight	[0, 1e4]	g	Numeric	Can be dynamic (<14 yrs old)
15	**DeathAge_c**	DeathAge_child	**INFO_c**	Death Age of the Child	[0, 1e3]	day	Numeric	DeathAge_c(day)=0 : stillborn
16	**HealthIssue_c**	HealthIssue_child	**INFO_c**	Issue(s) with health of child (<14	ergy_c; asthma_c; IBD_c; obese_c; CVD_c; Small_		AlphaNumeric	If some of the parameters are later quantified, those
17	**OtherAtBirth_c**	AtBirth_child	**INFO_c**	Other conditions of child at birth	bornInSummer; BornInWinter; CesareanDelivery		AlphaNumeric	Summer runs from June 1 to August 31 in Northern
18	**MILKinfo**	**MILKinfo**	**Condition**	**Info on milk samples.**				
19	**MilkStage**	MilkStage	**MILKinfo**	PostPartum stage of milk in days. (	[0, 1500]	day	Numeric	
20	**MilkType**	MilkType	**MILKinfo**	The type of Milk. Hind milk of a fe	*eed; completeBreast; notFeeding; whileFeeding; poc*		AlphaNumeric	*#whileFeeding/notFeeding* : Refering to whether milk
21	**storageTemp**	storageTemperature	**MILKinfo**	Temperature at which the sample	[-200, 50]	C	Numeric	doi: 10.1097/MPG.0000000000000641; Frozen: stor

21. ábra. A Condition (Feltétel)lap

A bejegyzések közötti kapcsolatok a korábbiaknak megfelelően fa szerkezetűek (**22. ábra**). A ***Condition*** (*Feltétel*) mező bejegyzései lehetnek numerikusak, de lehetnek egymásba ágyazottan definiált kategóriák is. A ***Condition*** mezőhöz tartozó információk fastruktúrába rendszerezettek: 4 fő csoportba sorolva 80 változó van rögzítve.



22. ábra. A Condition (Feltétel) fastruktúra

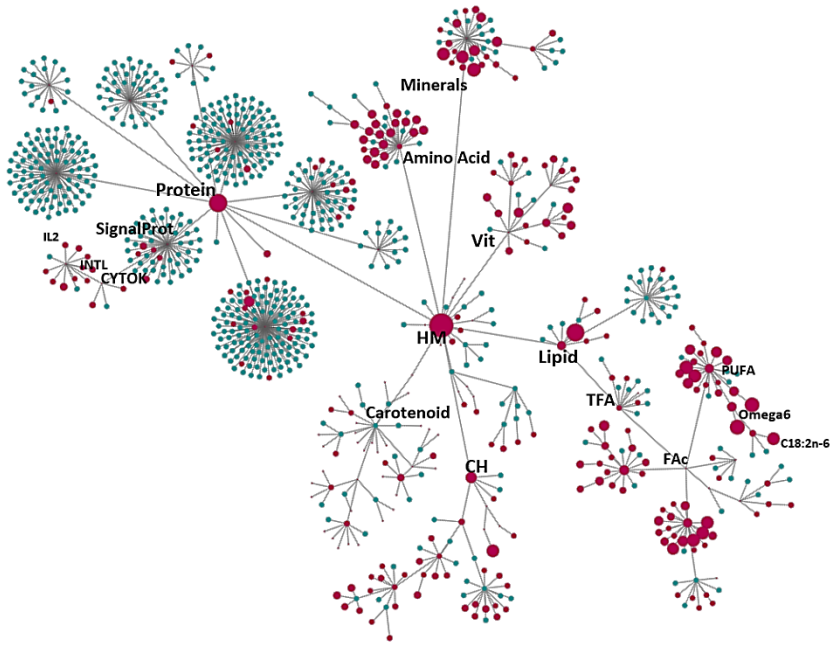
4.1.3.7. Component (Összetevő) lap

A válaszváltozók, az élelmiszer összetevők lehetséges értékeit tartalmazó definíciós táblázat a ***Component*** (Összetevő) lap (23. ábra). A csoportosításon és leíráson kívül tartalmazza a molekuláris szintű összetevők mol súlyát és az egyedi „Nemzetközi Kémiai Azonosítóját” (International Chemical Identifier, InChiKey *(Heller és mtsai., 2015)*).

	Component	Alternati	Group	Description	ValueSet	Unit	Type	MolWeight(g/mo	ChemID
2	25OHD2		VitD	25-Hydroxyvitamin-D2. A metabolite of vitamin	[0, 1.e-3]	g/L	Numeric	412.6	InchiKey=KJKIIUAXZGLUND-ICCVIKJNSA-N
3	25OHD3	Calcifediol	VitD	25-Hydroxyvitamin-D3. A metabolite of vitamin	[0, 1.e-3]	g/L	Numeric	400.6	InchiKey=JWUBBDSIWDLEOM-DTOXIADCSA-N
4	2FL		FS-OS	2-Fucosylatedlactose	[0, 100]	g/L	Numeric	488.4	InChiKey=SNFSYLYCDAVZGP-ZTFPOLCTSA-N
5	2FL_mol				[0, 0.1]	mol/L	Numeric		
6	3FL		neutOS	3-Fucosyllactose	[0, 100]	g/L	Numeric	488.4	InchiKey=WJPIUUDKRHCAEL-YVEAQFMBSA-N
7	3FL_mol				[0, 0.1]	mol/L	Numeric		
8	3SL		SL-OS	3-SialylLactose	[0, 1]	g/L	Numeric	633.6	InchiKey=OIZGSVFYNBZVIK-FHHHURIISA-N
9	3SL_mol				[0, 0.01]	mol/L	Numeric		
10	6SL		SL-OS	6-SialylLactose	[0, 1]	g/L	Numeric	633.5	InchiKey=MBSOBTIATUNICK-GIGDJUIZSA-N
11	6SL_mol				[0, 0.01]	mol/L	Numeric		
12	AAc	AminoAcid	Peptide	Alpha-amino carboxylic acid	[0, 1000]	g/L	Numeric		
13	AACH		Prot	Alpha1-antichymotrypsin	[0, 1]	g/L	Numeric		

23. ábra. A Component (Összetevő) lap

A MilkyBase alap adatbázis fastruktúrája (24. ábra) 394 csomóponttal rendelkezik, és további 357 indirekt (származtatott) változó definiált a ***Component*** lapon.



24. ábra. Az anyatej összetevőinek csoportosítását ábrázoló fa struktúra az alap MilkyBase adatbázisban rögzített adatok alapján

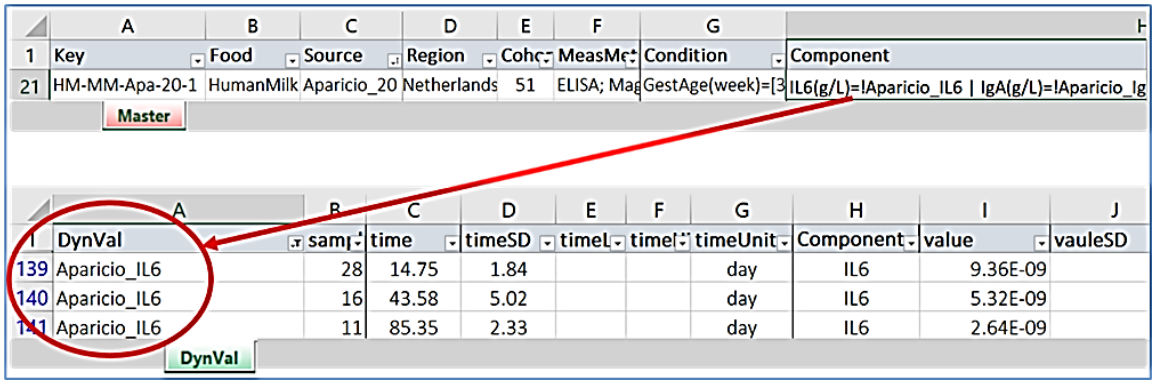
4.1.3.8. DynVal (Dinamikus értékek) lap

Mint korábban már említettem, a ***Master*** (Alap) lap a `Component (Összetevő)` és `Condition (Feltétel)` mezői numerikus értékek időfüggő sorozatai, úgynevezett dinamikus értékek is lehetnek. Ezeket az értékeket tárolja a **DynVal (**Dinamikus értékek) dedikált lap (25. ábra).

	DynVal	sampleSiz	time	timeSD	timeLo	timeHi	timeUnit	Component	value	vauleSD	valueLo	valueHi	valueUni	Comment
2	Ala_ForeSummCalcitriol		53		50	56	day	Calcitriol_mol	1.47E-09		1.35E-10	4.08E-09	mol/L	
3	Ala_ForeSummCalcitriol		137		134	140	day	Calcitriol_mol	3.25E-10		1.23E-10	1.54E-09	mol/L	
4	Ala_ForeSummCalcitriol_25VitD2_Suppl		53		50	56	day	Calcitriol_mol	9.98E-10		2.48E-10	3.29E-09	mol/L	
5	Ala_ForeSummCalcitriol_25VitD2_Suppl		137		134	140	day	Calcitriol_mol	8.85E-10		1.35E-10	1.93E-09	mol/L	
6	Ala_ForeSummVitD3		53		50	56	day	VitD3_mol	5.28E-10		2.31E-10	9.02E-10	mol/L	
7	Ala_ForeSummVitD3		137		134	140	day	VitD3_mol	4.11E-10		5.50E-11	6.27E-10	mol/L	
8	Ala_ForeSummVitD3_25VitD2_Suppl		53		50	56	day	VitD3_mol	5.67E-10		2.76E-10	8.50E-10	mol/L	
9	Ala_ForeSummVitD3_25VitD2_Suppl		137		134	140	day	VitD3_mol	6.76E-10		1.33E-10	8.22E-10	mol/L	
10	Ala_ForeWinCalcitriol		53		50	56	day	Calcitriol_mol	1.13E-10		6.30E-11	2.90E-10	mol/L	
11	Ala_ForeWinCalcitriol		137		134	140	day	Calcitriol_mol	2.83E-10		1.08E-10	4.80E-10	mol/L	
12	Ala_ForeWinCalcitriol_25VitD2_Suppl		53		50	56	day	Calcitriol_mol	3.60E-10		7.50E-11	4.50E-10	mol/L	
13	Ala_ForeWinCalcitriol_25VitD2_Suppl		137		134	140	day	Calcitriol_mol	6.53E-10		2.40E-10	1.67E-09	mol/L	
14	Ala_ForeWinVitD3		53		50	56	day	VitD3_mol	3.15E-10		1.33E-10	5.90E-10	mol/L	
15	Ala_ForeWinVitD3		137		134	140	day	VitD3_mol	7.02E-10		9.10E-11	9.78E-10	mol/L	
16	Ala_ForeWinVitD3_25VitD2_Suppl		53		50	56	day	VitD3_mol	5.02E-10		1.30E-10	7.20E-10	mol/L	
17	Ala_ForeWinVitD3_25VitD2_Suppl		137		134	140	day	VitD3_mol	5.75E-10		2.81E-10	9.36E-10	mol/L	
18	Ala_hindSummCalcitriol		53		50	56	day	Calcitriol_mol	1.35E-09		3.55E-10	4.05E-09	mol/L	
19	Ala_hindSummCalcitriol		137		134	140	day	Calcitriol_mol	5.83E-10		3.83E-10	2.05E-09	mol/L	
20	Ala_hindSummCalcitriol_25VitD2_Suppl		53		50	56	day	Calcitriol_mol	1.56E-09		5.85E-10	3.90E-09	mol/L	
21	Ala_hindSummCalcitriol_25VitD2_Suppl		137		134	140	day	Calcitriol_mol	1.11E-09		5.58E-10	2.11E-09	mol/L	

25. ábra. A DynVal (Dinamikus értékek) lap

Ahhoz, hogy a ***Master*** *(Elsődleges)* lap dinamikus értékeket tartalmazó mezőit a ***DynVal*** lap megfelelő [idő, érték] párjaival összekapcsolhassuk egy "!" karakterrel kezdődő mutató szükséges. Az alábbi példán látható (26. ábra),hogy a ***Master*** lapon található Il6(g/L) változó időfüggő értékeit a „! Aparicio_IL6” mutató összekapcsolja a ***DynVal***-lapon tárolt [idő, érték] párok időfüggő sorozatával.



26. ábra. A Master lap időfüggő mezőinek összekapcsolása a DynVal (Dinamikus értékek) tábla megfelelő rekordjaival

4.2. MILKYBASE ADATBÁZIS ÚJDONSÁGAI

Mivel adatbázisunk nem csak azzal a céllal készült, hogy az anyatej összetevőket egy táblázatba összegyűjtsük, hanem hogy megvizsgálhassuk ezek az összetevők hogyan függenek a különböző faktoroktól, így egy olyan adatstruktúra definiálására törekedtünk mely alkalmas ennek tükrözésére.

Az adatbázisunk fő újdonsága az ontológia, amely az anyatej összetétel mérési körülményeinek hatására, az anya/gyermek különböző jellemzőire, valamint a környezeti és előzményi körülményekre, adott válaszként tekinti az összetételre vonatkozó adatokat, valamint ezen adatok dinamikájára és bizonytalansági jellemzőire összpontosít, amelyek a magyarázó és válaszváltozókban kerülnek beírásra.

Ehhez a következő újításokat vezettünk be:

1. A MilkyBase ontológia definiálása a magyarázó változók (azaz a feltételek) és a válasz változók (azaz az összetevők) szem előtt tartásával.
2. A változók statikus és dinamikus (időfüggő) formában is rögzíthetőek
3. A „kiterjesztett numerikus” változók, melyek lehetőséget adnak a bizonytalanságok mérésére is
4. Általános fa adatstruktúra
5. Az adatok direkt és indirekt (származtatott) formában is rögzíthetők

4.2.1. Magyarázó- és válaszváltozók

Adatbázis-építési elvünk alapja, hogy egy-egy rekordot úgy tekintünk, mint a különböző magyarázó körülmények - amelyek mellett a megfigyelések történtek - leképezését az anyatej összetételére, mint válaszváltozóra. Azokat a változókat, melyek az adatok keletkezésének körülményeit írják le ***Magyarázó-***, míg az anyatej komponenseinek mért adatait a ***Válasz-változók*** reprezentációjának tekintjük, ezzel előkészítve a többváltozós függvényekkel való regressziót. A magyarázó változók az anya, gyermek, illetve a születés részleteit (anya kora, diétája, gyerek súlya, terhesség hányadik hetében született stb.) írják le, és az anyatej úgynevezett „történelmét” (a földrajzi helyzet, mérési és tárolási módszerek stb.), míg a válaszváltozók az anyatej összetevők (nutriensek, bioaktív komponensek stb.) mennyiségi adatai. A magyarázó és

válaszváltozókat vektoroknak tekintettük, ahol az elsőben minden egyes bejegyzés egy (többnyire számszerűsített) érték egy adott feltételre vonatkozóan, amely a válaszváltozókat eredményezte, akár közvetlen, akár közvetett formában.

E válaszváltozók rögzített értékei lehetnek úgynevezett "kiterjesztett numerikus" formátumúak (*lásd 4.2.3 A „kiterjesztett numerikus" változók*), illetve akár statikus (azaz egy időpontra vonatkozó) akár dinamikus (azaz időbeli változást követő) formátumú is, azaz a dinamikus (időfüggő) állapotok is rögzíthetők. (*lásd 4.2.2 A statikus és dinamikus (időfüggő) változók*), ahol az összetevők időbeli változását egy táblázatban tároljuk, és a táblázatra irányító mutató a változó beviteli értéke.

A magyarázó változók numerikus értékei (feltétel mezők) szintén rögzíthetők a válaszváltozókhoz hasonló módon, de nem feltétlenül csak numerikus értékeket tartalmazhatnak. Lehetnek Boolean értékek vagy kategória listák is. Ugyanúgy, ahogy egy szám egy intervallumhoz tartozik, egy kategóriaérték is tartozhat egy csoporthoz vagy több csoporthoz. Ilyen kategóriaértéket felvevő magyarázó változó például a földrajzi régió: Kína kategóriacsoportja például lehet "Ázsia" vagy "Távol-Kelet".

4.2.2. A statikus és dinamikus (időfüggő) változók

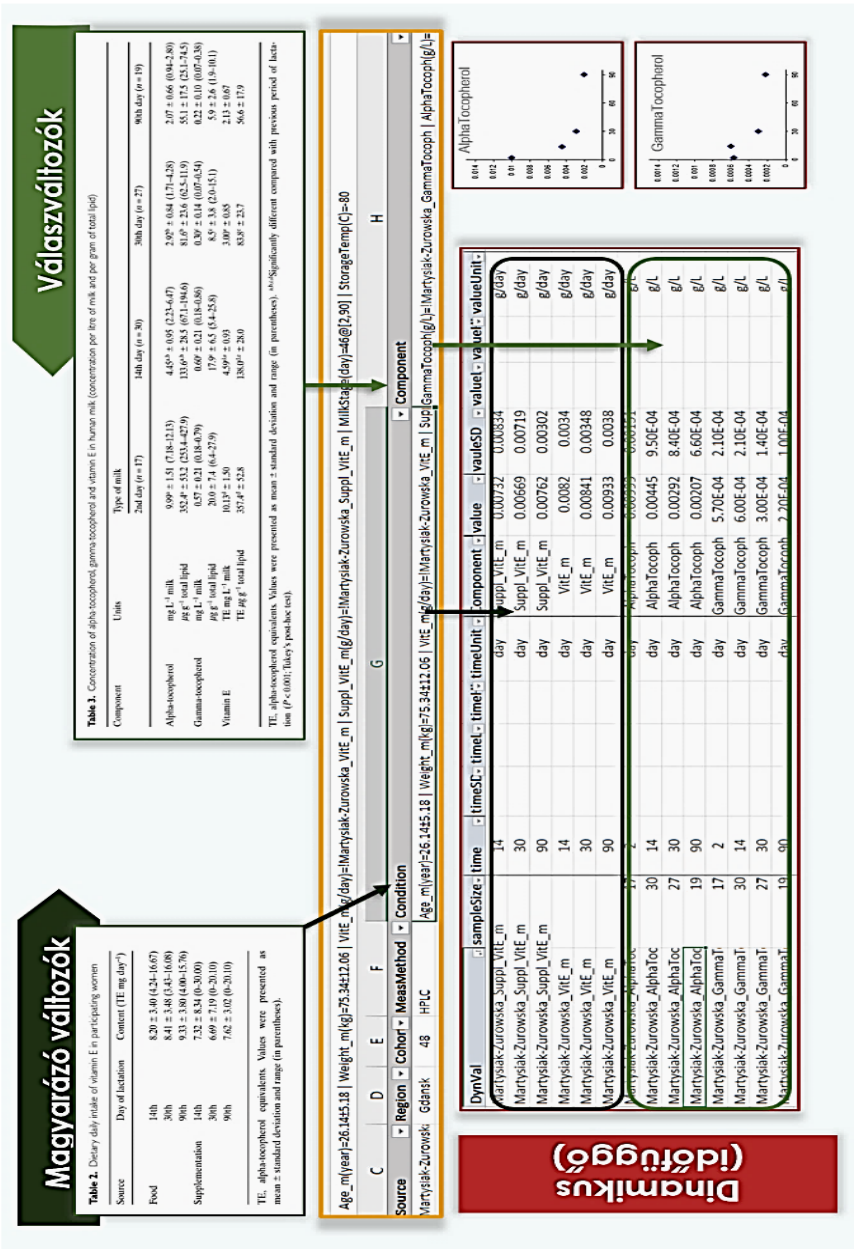
Ahhoz, hogy az anyatej komponensek trajektóriáit megfelelően bemutathassuk, elengedhetetlen eszköz, hogy a változók időbeli változását egy dedikált táblával reprezentáljuk, amelyek a pályák („időpont, mért érték") értékpárjait tartalmazza.

Adatbázisunkban az időfüggő magyarázó és válaszváltozókat [idő, érték] adattáblák reprezentálják, míg az fő lap vonatkozó bejegyzése csak egy mutató erre a táblára. Ezen mutatók biztosítják, hogy az időfüggőség az adott mezők természetes attribútuma.

Az időfüggő adatokra illeszkedő pályák (trajektóriák) származtatott paraméterei, mint például a sebesség vagy az állandósult állapot szintje, a pályák lehetséges skaláris reprezentánsai. Ezt a struktúrát az "elsődleges - másodlagos modell" megközelítés *(lsd. 2.2.3. Matematikai modellalkotás, elsődleges/másodlagos modell*) ihlette, amely már a prediktív mikrobiológia alapjává vált, mind a matematikai modellezés, mind az adattárolás tekintetében *(Baranyi & Tamplin, 2004)*. Vagyis egy változó időbeli profilját néhány kulcsparaméter („elsődleges modell") írja le, míg e paraméterek változása a választ befolyásoló körülmények (feltételek) függvényében a másodlagos modellekkel

írható le. Ez az adatstruktúra tette lehetővé az anyatej összetevők prediktív modellezését is. *(lsd. 4.3. Az anyatej molekuláris összetételében lévő mintázatok felismerése (Prediktív anyatej komponens Modellezés*)

Az adatbázisunkban egyetlen rekordba rögzített többváltozós dinamikus válaszra mutat egy példát a 27. ábra, mely Martysak- Zurowska és munkatársai által egyetlen publikációban *(Martysiak-Żurowska és mtsai., 2013)* közölt, egymást követő időpontokban mért anyai diétára vonatkozó adatokat és az anyatejben mért vitaminok koncentrációját tartalmazza.



27. ábra. Dinamikus magyarázó- és válaszváltozók

4.2.2.1. A dinamikus változókra, mint alapértelmezett bejegyzésekre való összpontosítás további előnyei

Az egyetlen rekordba rögzített többváltozós dinamikus válaszok megkönnyítik az időben változó adatok vizualizációját és ezzel azok összehasonlítását, elemzését. Erre láthatunk egy példát az ábrán, ahol az alap MilkyBase adatbázisban rögzített adatok alapján, a kolosztrumban található összzsírsav koncentrációjának időfüggő változását ábrázoltuk 5 kínai város esetében.

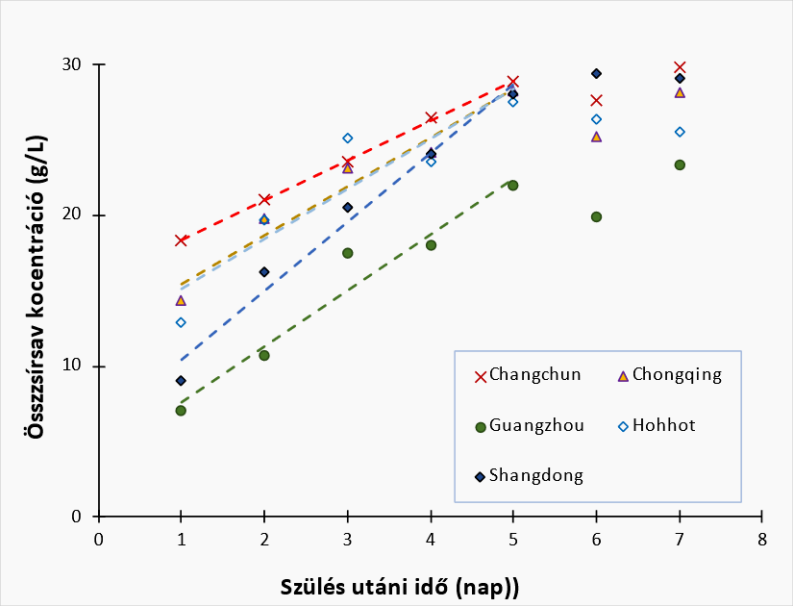
Egy ilyen típusú adatvizualizáció segítséget nyújthat

(i) az adatokban lévő minták és kiugró értékek felismeréséhez;

(ii) az adathiányok azonosításához;

(iii) a hibák esetleges azonosításához.

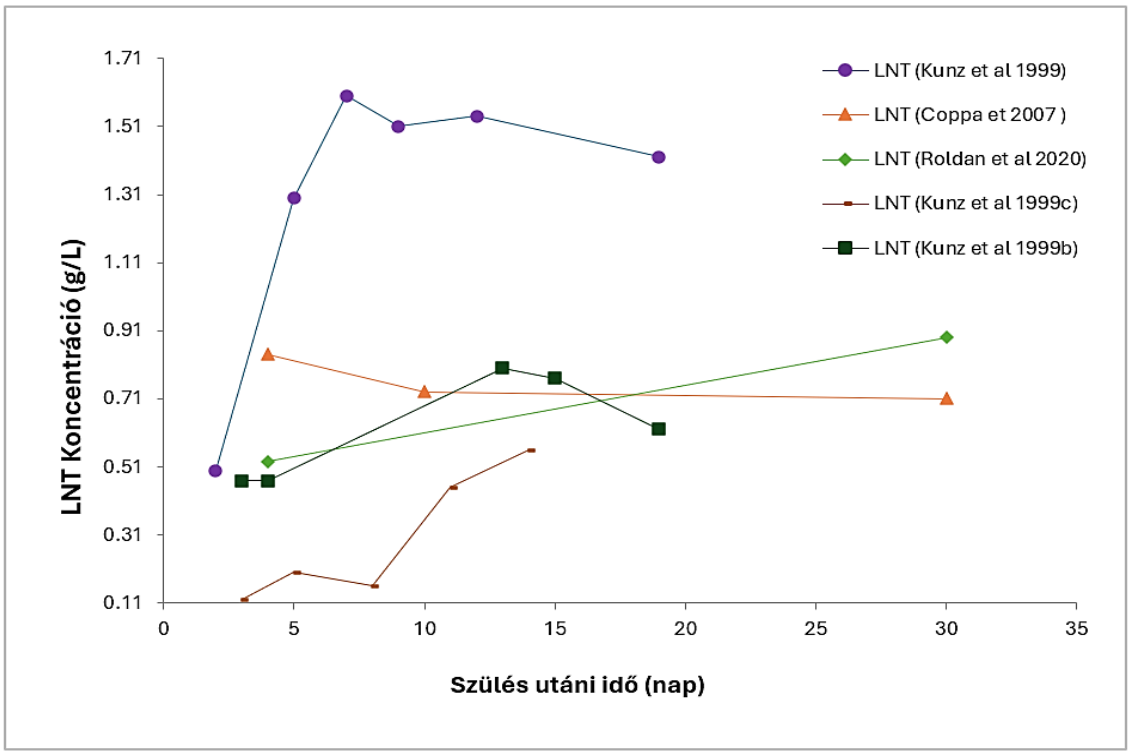
Ezen ábra alapján merült fel az az az ötlet például (*lásd 4.3.1.2. Kétfázisú szaturációs modell*), hogy a kolosztrum időszak végét úgy lehet meghatározni, mint azt az időpontot, amikor a zsírsavkoncentráció lineáris növekedése véget ér.



28. ábra. Az anyatej összzsírsav koncentrációjának változása a kolosztrumban 5 kínai város esetében Liu és munkatársainak *(Liu és mtsai., 2019)* adatai alapján. A mért adatok a zsírsav koncentráció gyors lineáris növekedését mutatják az 5. napig, hasonló meredekséggel.

Forrás: *(Pacza és mtsai., 2022)*

Másik lehetséges mód az adatok összehasonlítására, amikor egy adott összetevőre vonatkozó különböző publikációk adatait ábrázoljuk együtt. Erre a ábrán látható egy példa, melyen különböző szerzők által publikált és a MilkyBase adatbázisban rögzített adatok alapján hasonlítjuk össze az anyatejben lévő Lacto-N-tetraóz (LNT) koncentrációjának trajektóriáit. Itt a többitől eltérő, kiugró értékre figyelhetünk fel. A Kunz és munkatársai *(Kunz és mtsai., 1999)* által közölt adatok jelentősen eltérnek a többi publikáció adataitól, ami annak további vizsgálatára irányíthatja a figyelmet, hogy mi okozhatta ezeket a különbségeket.



29. ábra. Lacto-N-tetraoz koncentrációjának változása

Forrás: *(Pacza és mtsai., 2022)*

(A forráscikkek adatait a Melléklet tartalmazza.)

4.2.3. A „kiterjesztett numerikus” változók

A mérési bizonytalanságok adatbázisba való rögzítésének elősegítésére bevezettük az adatbázis *"numerikus mező"* fogalmának kibővített definícióját. A magyarázó és válaszváltozók rögzített numerikus értékeit úgynevezett "kiterjesztett numerikus" formátumban adjuk meg. Ez alatt azt értjük, hogy a beírt számot vagy a ± szórással, vagy a körülötte lévő intervallummal (mint minimum-maximum, vagy kvantilis) lehet ellátni. Mindkettő az adatok bizonytalanságát jellemzi. Mindezen túl különbséget teszünk a nyers megfigyelések és a becslések között is, oly módon, hogy a változók értékeinek rögzítésére egy speciális, kettős formátum is megengedett. A pontosvesszővel elválasztott két részből, az első a nyers (mért) adatokra vonatkozik szórásukkal (vagy interkvantilis tartományukkal vagy [Min, Max] intervallummal); a második rész egy becslésre és a becslés megbízhatóságára vonatkozik (standard hiba vagy 95% -os konfidencia intervallum). Míg az első rész leíró jellegű (jellemzően a megfigyelések átlaga és a körülötte lévő szórás számszerűsítése), addig a második prediktív (vagy becslés) az előrejelzés/becslés standard hibájával vagy konfidencia intervallumával.

Alapértelmezett formája egy közönséges valós szám, a leíró statisztikák által kalkulálható érték, rendelkezésre álló releváns adatok bizonyos középértéke, melyet ezen adatok szóródásának mértékével is el lehet látni. Ez általában vagy a szórás, vagy a minimum-maximum tartomány. Egy további, pontosvesszővel elválasztott második részben rögzíthető a valós átlag előrejelzése (vagy becslése). Ez utóbbi becsült érték is megadható bizonytalansági számszerűsítéssel, ami általában vagy a becslés standard hibája, vagy annak 95%-os konfidenciaintervalluma (**1. táblázat**).

1. táblázat: Kiterjesztett numerikus változók

Leírás	Adatbeviteli formátum
Mért / számolt érték	x_1
Mért / számolt érték **szórással**	$x_1 \pm y_1$
Mért / számolt érték, az azt tartalmazó **intervallummal**	$x_1 @ [y_1, z_1]$
Mért / számolt érték és a **becsült érték**	$x_1 ; x_2$
Mért / számolt érték **szórással**, és **becsült értéke**	$x_1 \pm y_1 ; x_2$
Mért / számolt érték, az azt tartalmazó **intervallummal**, és **becsült értéke**	$x_1 @ [y_1, z_1] ; x_2$
Mért / számolt érték , és a **becsült értéke annak standard hibájával.**	$x_1 ; x_2 \pm y_2$
Mért / számolt érték **szórással**, és a **becsült értéke annak standard hibájával.**	$x_1 \pm y_1 ; x_2 \pm y_2$
Mért / számolt érték , az azt tartalmazó **intervallummal**, és a **becsült értéke annak standard hibájával.**	$x_1 @ [y_1, z_1] ; x_2 \pm y_2$
Mért / számolt érték , és a **becsült értéke annak 95%-os konfidencia intervallumával.**	$x_1 ; x_2 @ [y_2, z_2]$
Mért / számolt érték **szórással**, és a **becsült értéke annak 95%-os konfidencia intervallumával.**	$x_1 \pm y_1 ; x_2 @ [y_2, z_2]$
Mért / számolt érték , az azt tartalmazó **intervallummal**, és a **becsült értéke annak 95%-os konfidencia intervallumával.**	$x_1 @ [y_1, z_1] ; x_2 @ [y_2, z_2]$

Ez a formátum legalább egy numerikus érték rögzítését igényli, az összes többi fent ismertetett érték nem kötelező.

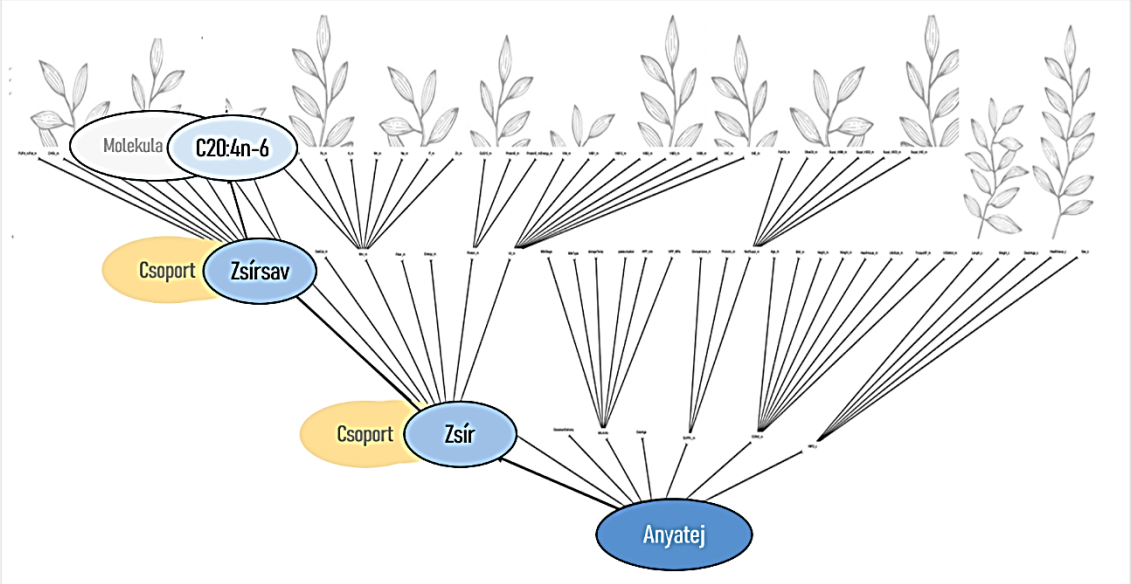
Általában az első rész a mért adatok számtani átlaga a megfelelő szórásmutatókkal együtt. Ez az átlag jellemzően a valós átlag előrejelzésére/becslésére szolgál, de nem feltétlenül minden esetben. Ez a becsült valós átlag bizonytalansági számszerűsítéssel is ellátható. Ilyen például az átlag becslésének standard hibája vagy annak 95%-os konfidenciaintervalluma. Megjegyzendő, hogy általában az átlagolás valóban a valódi átlag becslése, de nem feltétlenül a legjobb, és előfordulhat, hogy a két középérték nem azonos. Emellett az adatok szórása soha nem lehet kisebb, mint az átlag becslésének standard hibája. Ez utóbbi ok miatt a publikációk adatainak rögzítésekor, nagyon sok esetben találtunk anomáliákat a publikációkban, függetlenül attól, hogy a szerzők az adatok szórásáról beszéltek-e vagy a valós átlag becslésének standard hibájáról. Egy másik tipikus példa a félreértelmezésre a „± hiba” kifejezéssel kapcsolatos, mivel a szórást (amely a statisztikai minta átlaga körüli szórást számszerűsíti) gyakran összetévesztik az adatok valós átlagának becslésére vonatkozó

standard hibával. *(Barde & Barde, 2012) (Nagele, 2003)* Mivel adatstruktúránk szétválasztja a nyers adatokat és a becslést/előrejelzést, az adatok rögzítése során annak eldöntésében, hogy hová helyezzük a közzétett hiba adatokat sokszor, csak a többi adattal való összehasonlítás volt az egyetlen támpont.

Ha a rögzített adatok intervallumokat képviselnek, akkor a későbbi adatelemzések során sztochasztikus intervallum elemzés használhatók, amely sokkal erőteljesebb, mint a determinisztikus értékekkel végzett számítások, mivel nem csak a mennyiségi következtetéseket lehet levonni, hanem a következtetések megbízhatósága is számszerűsíthető.

4.2.4. Általános fa adatstruktúra

Az egyes biokémiai összetevők esetében - kompatibilitás érdekében - az összetevő tömegének tejtérfogathoz viszonyított koncentrációját határoztuk meg válaszértékként. Összetevő („*Component*”) alatt egy molekulát vagy molekulacsoportot értünk. Összetevő például a C20:4n-6 („Arachinodic acid”) molekula, és a "zsírsav" is, azonban míg az első egy molekula összetevő, addig a második egy molekulacsoport, melynek eleme az említett C20:4n-6 molekula is. Az ilyen típusú csoportosítás egy hierarchikus fa struktúrát követ (30. ábra). Ezt a struktúrát követve nem csak egy adott molekula koncentrációja, hanem bármely molekulacsoport (azaz az anyatej gyökér felletti szintjeiről bármely összetevő) koncentrációja is rögzíthető (32. ábra).



30. ábra. Általános fastruktúra

A MilkyBase adatbázisban nemcsak az összetevők alkotnak hierarchikus faszerkezetet, hanem az egész adatbázisra jellemző ez a struktúra. Az egyes fastruktúrák MilkyBase adatbázis elemeinek részletes ismertetésekor (*lsd. 4.1.1 A MilkyBase adatbázis felépítése*) kerülnek bemutatásra.

4.2.5. Az adatok direkt és indirekt (származtatott) formában is rögzíthetők

A kutatás során felfedeztük, hogy az összetevők mért koncentrációi helyett számos szerző csak átváltott vagy származtatott értékeket publikál egy-egy komponens esetén.

Erre példa a C20:4n-6 zsírsav („Arachinodic acid") is, mely mérési eredményeit különféle módokon publikálják (31. ábra.)

- vagy mért koncentrációval (g/L)
- vagy az összes zsírsav arányában („C20:4n-6/FAc"),
- vagy az összes zsírsav metil észter arányában („C20:4n-6/FAME")
- vagy az összes trigliceridhez viszonyítva („C20:4n-6/TG") .

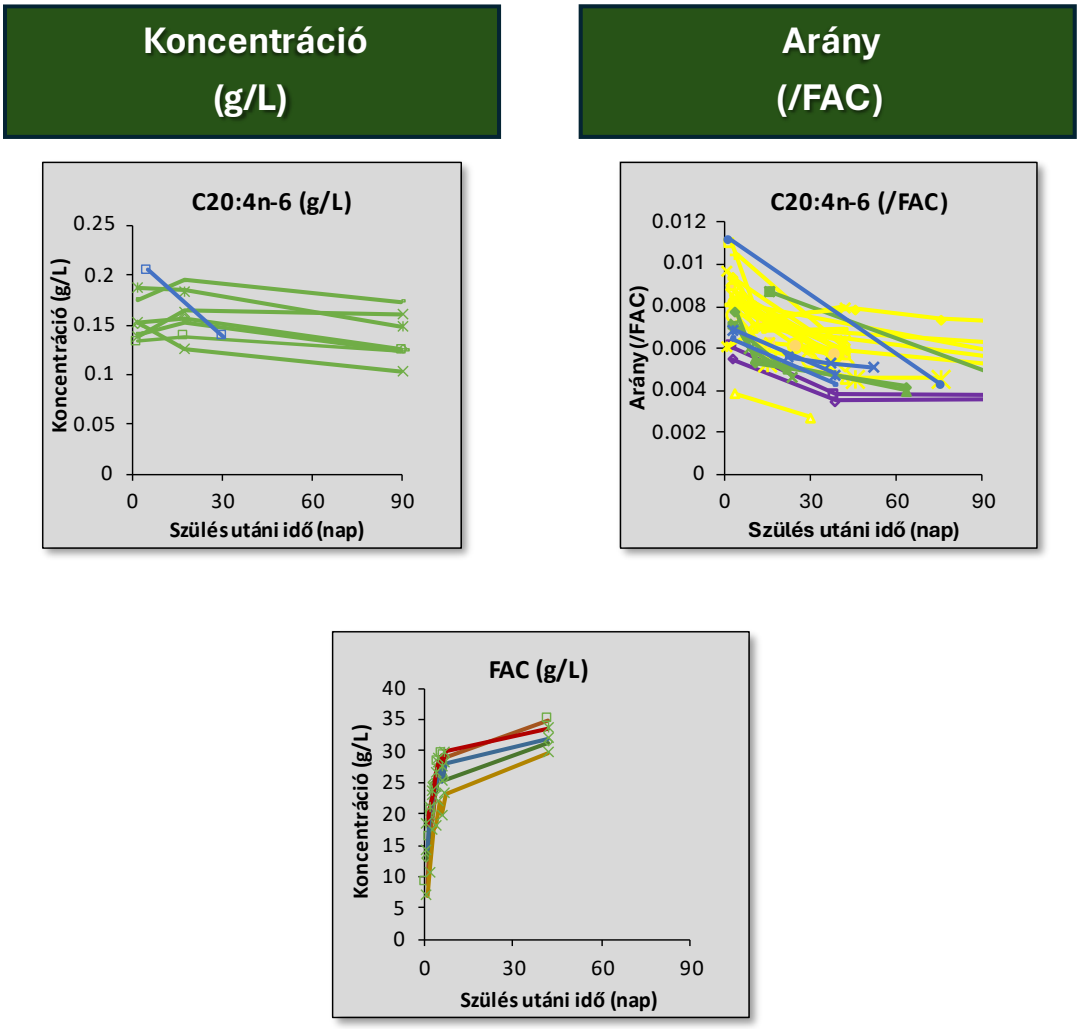
1	Component	Alternati	Group	Description	ValueSet	Unit	Type	MolWeight(g/mo	ChemID
235	C20:4n-6	AA; ARA	Omega6	Arachidonic acid OR Arachidonate	[0, 100]	g/L	Numeric	304.5	InchiKey=YZXBAPSDXZZRGB-DOFZRALJSA-N
236	C20:4n-6/FAc				[0, 1]	-	Numeric		
237	C20:4n-6/FAME				[0, 1]	-	Numeric		
238	C20:4n-6/TG				[0, 0.1]	-	Numeric		

31. ábra. A C20:4n-6 zsírsav rögzíthető formái a Milky Base adatbázisban

Az ilyen esetek egyértelmű kezelése érdekében a C20:4n-6-re vonatkozó (mért adat) számértéket **direkt (közvetlen),** míg a C20:4n-6/FAc arányra vonatkozó értéket **indirekt (származtatott)** formának nevezzük.

Hasonlóképpen, a "C18:1n-9 + C18:3n-3" változó elnevezése azt jelzi, hogy a két zsírsavat együttesen mérték. Adatbázisunk lehetőséget nyújt mindezen formátumok rögzítésére, tehát egy változó neve tartalmazhatja a ":" karaktert, hogy a lehető legközelebb álljon a biokémiai jelöléseikhez, valamint a "/" és "+" speciális karaktereket is, mint az **indirekt változók jelzőkódjait**.

A hierarchikus osztályozás és az adatok indirekt formában való rögzíthetősége lehetővé teszi, hogy nem csak egy adott molekula koncentrációja, hanem az molekulát tartalmazó csoportoké is elemezhető. (*32. ábra*)



32. ábra. A C20:4n-6 (Arachidonsav) zsírsav molekula koncentrációja és az összes zsírsavhoz viszonyított aránya, valamint az molekulát tartalmazó összes zsírsav (FAC) koncentrációja az anyatejben a MilkyBase-ben rögzített adatok alapján

Azt tapasztaltuk, hogy az indirekt válaszok nagy részét az arányok teszik ki, leginkább egy adott zsírsavmolekula koncentrációja az összes zsírsavhoz viszonyítva. Ezekből az adott zsírsavmolekula koncentrációja csak akkor becsülhető meg, ha az összes zsírsav koncentrációja ismert. Hasonlóan, amikor egy összetevőt molekulatömegben mérnek; csak akkor lehet koncentrációra átszámítani, ha a Mólsúly ismert (ezeket a Master lap külön mezőjében megadjuk).

4.3. AZ ANYATEJ MOLEKULÁRIS ÖSSZETÉTELÉBEN LÉVŐ MINTÁZATOK FELISMERÉSE (PREDIKTÍV ANYATEJ KOMPONENS MODELLEZÉS)

A "prediktív anyatej-modell készítés", azaz az anyatej molekuláris összetételében lévő mintázatok felismerésének érdekében két fő feladatra koncentráltunk:

1. Először egy olyan elsődleges modell létrehozása, amely képes leírni az anyatej komponenseinek időbeli pályáját,
2. majd annak bemutatása, hogy az elsődleges modell paraméterei hogyan függenek olyan tényezőktől, mint például a földrajzi elhelyezkedés.

Ahhoz, hogy a lehető legjobb modellt kaphassuk, az anyatej komponensekre vonatkozó időbeli adatoknak legalább 30 napra kell kiterjedniük, és legalább 4-5, de lehetőleg 10-nél több adatpontot kell tartalmazniuk a vizsgált időintervallumon belül. Az ilyen adatsorok ritkák a szakirodalomban, de felhasználhattuk a származtatott adatokat, leggyakrabban a vizsgált komponensek átlagait és standard eltéréseit egy-egy homogén kohorszon belül.

4.3.1. Elsődleges modell elkészítése

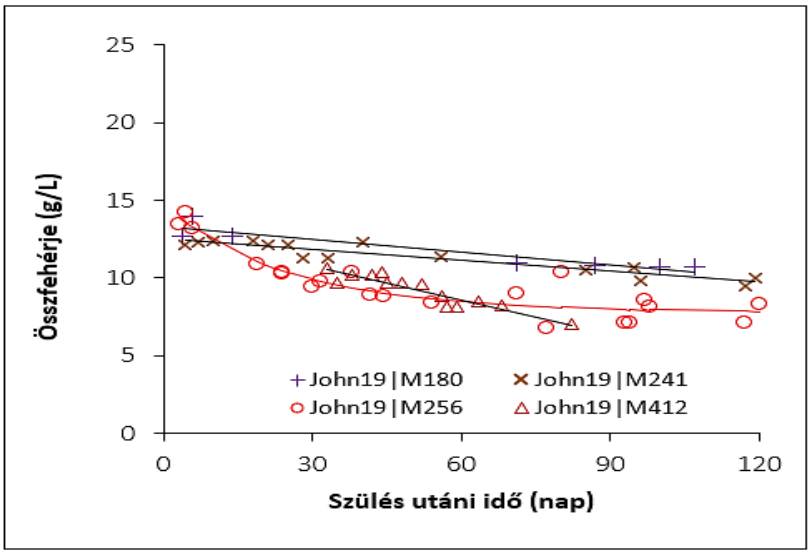
Ahhoz, hogy az elsődleges modell felépítéséhez, paramétereinek meghatározásához ötleteket kaphassunk, egy-egy anya esetében mért részletes egyedi adatokra volt szükségünk az anyatej összetevők időbeli változásával kapcsolatosan. Ideális esetben több száz ilyen idő függvényében leírt pályát tudtunk volna összegyűjteni a publikált szakirodalomból, de sajnos azt tapasztaltuk, hogy az ilyen mélységű részletes adathalmaz nagyon ritka.

4.3.1.1. Egyfázisú szaturációs modell a fehérjekoncentrációk időbeli változásának jellemzésére

John és munkatársainak publikációjában találtunk egy ilyen, induló elemzésre alkalmas adathalmazt *(John és mtsai., 2019)*. Az anyatej fehérjéiről publikáltak adatokat (443 anya, különböző időpontokban mért adatai); melyek a MilkyBase *(Pacza és mtsai., 2022)* adatbázisban is rögzítésre kerültek.

4.3.1.1.1. Több mint 10 mérést biztosító egyedi anyákra vonatkozó mérések

Először az egyedi anyákra vonatkozó mérések adatait vizsgáltuk meg, azokra az esetekre összpontosítva, ahol a lehető legtöbb mérés történt egy-egy anya esetében. A **33. ábra** négy olyan anyára vonatkozóan mutat egyéni fehérjepályákat (fehérjekoncentrátumok időbeli változását mutató pályákat), akik a születéstől számított 120 napon belül, legalább 30 napos időintervallumban, több mint 10 mintát szolgáltattak.



33. ábra. Az anyatej összfehérje koncentrációjának időbeli változását mutató trajektóriák a *(John és mtsai., 2019)* publikáció alapján.
A vékony piros vonal az M256-os kódjelű anya -aki a legtöbb mintát szolgáltatta a kísérleti megfigyelés 4 hónapos időszakában- adataira illesztett egyszerű szaturációs modell *(Baranyi-Pacza- és mtsai., 2024)*

Megfigyeléseink azt mutatták, hogy ezek az - ábrán fekete vonallal ábrázolt - egyéni pályák, egyszerű ereszkedő görbék. Azaz mindegyik lineárisan csökken a megfigyelési időintervallum alatt, továbbá az M256-os jelzésű anya esetében – piros vonallal jelölt pálya- már viszonylag korán egy úgynevezett állandó állapothoz konvergál a csökkenő fehérjekoncentráció szint.

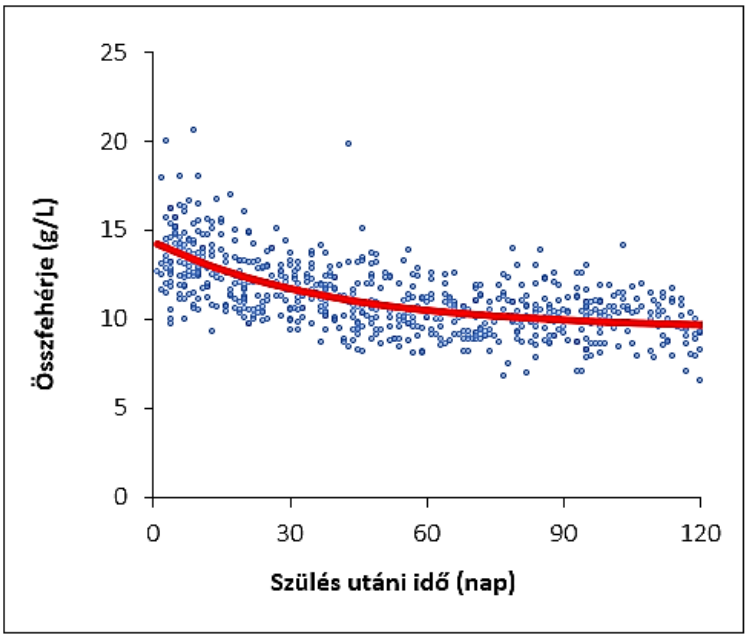
Feltehetjük, hogy még ha egy egyedi fehérje koncentráció pálya a megfigyelési időintervallum alatt lineáris ereszkedik is, későbbiekben egy majdnem statikus állapotba kerül (ellaposodik), azaz egy közel állandó (esetenként nulla) értékhez konvergál. Ezért mind a négy egyéni pálya leírható monoton konvergens csökkenő matematikai

függvénnyel, mint például a korábban említett szaturációs modellel. Fontos megjegyezni, hogy a közel lineáris lefutású görbék nagyon kicsi, közel nulla exponenciális konvergencia hányadossal rendelkeznek.

4.3.1.1.2. Összes anyára megadott egyéni mérések

Az egyedi anyákra vonatkozó adatok vizsgálatát követően, a John és munkatársai *(John és mtsai., 2019)* publikációjában található összes anyára megadott egyéni mérési adatot elemeztük.

Az 34. ábra az anyatejben mért teljes fehérjekoncentrációt mutatja a szüléstől eltelt idő függvényében, egyedi anyákra mérve, a szülést követő első 120 napban . Mivel a 177 anya vonatkozásában összesen 545 mérés történt, így az egy anya által adott átlagos mintaszám körülbelül 3 volt. Ez az adatszám nem elég az egyéni pályák azonosítása, azonban az átlagos koncentrációk az idő függvényében jól illeszthetőek a szaturációs modellel.



34. ábra. Az anyatej fehérje koncentrációjának időbeli változása a *(John és mtsai., 2019)* publikáció összes közölt mérési adata alapján. Az ábra kék pontjai a 177 anyától származó, összesen 545 darab minta egyedi értékeit mutatja a szülést követő 120 napban, míg a piros vonal az összes pontra illesztett egyszerű szaturációs görbét ábrázolja. Forrás: *(Baranyi-Pacza- és mtsai., 2024)*

Mivel az exponenciális konvergencia sebessége itt egy populáción alapul, ezért ezt *populációs konvergencia sebességnek* nevezzük, szemben az egyes anyák *egyéni konvergencia sebesség*évek. Mivel a modell nem lineáris, ezért a populációs konvergencia sebesség nem az egyéni konvergencia sebességek számtani átlaga, hanem azok eloszlásának függvénye, és amint azt a 34. ábra mutatja, az egyszerű szaturációs függvény még mindig jól alkalmazható modell a tipikus (átlagos) viselkedésükre.

Az 34. ábra összes pontjára illesztett szaturációs modell paramétereit a **2. táblázat** mutatja:

2. táblázat A DmFit4 illesztő VBA program által kapott értékek**.**

curve	**JOHN_PROT**
nData	545
yDatMin	6.5
tObsMin	1
tObsMax	120.148
tIntval	119.1
yDatMax	20.6
model	Saturation
rate	-0.031
se(rate)	0.005
y_0	14.245
se(y_0)	0.232
y_{End}	9.691
se(y_{End})	0.189
se(fit)	1.474
R^2_stat	0.443
initVal	12.7

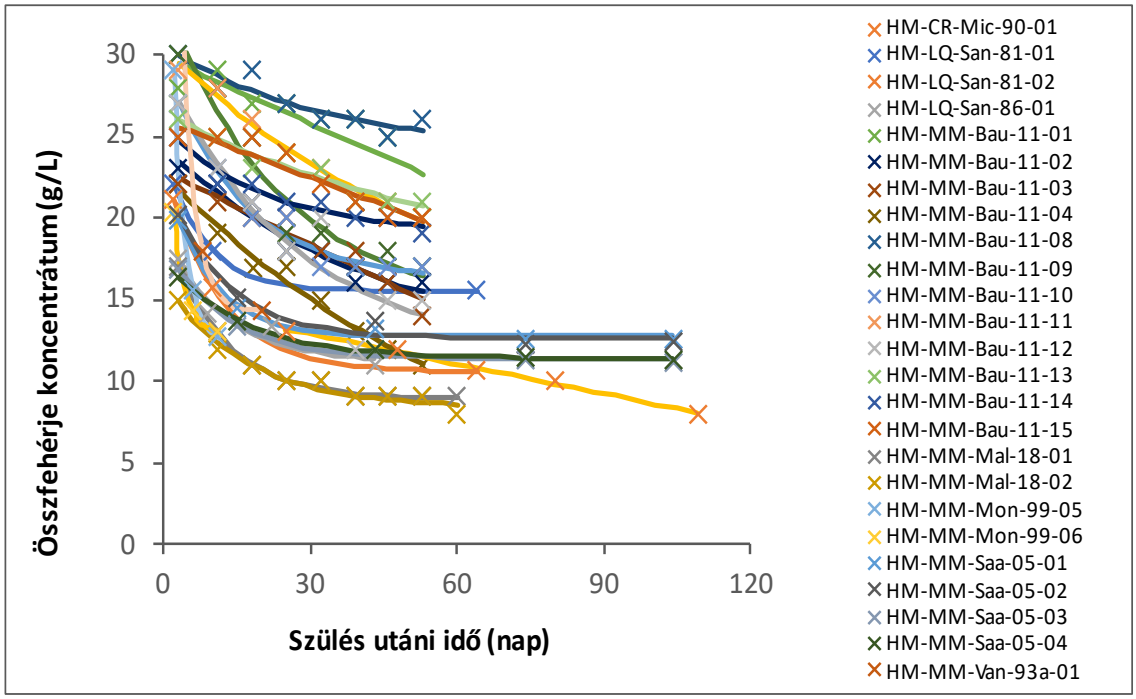
Tehát az illesztett modell a következőképpen írható le:

$$y(t) = 14.245 \cdot e^{-0.031 \cdot t} + 9.691 \cdot (1 - e^{-0.031 \cdot t}) \qquad (0 \leq t)$$

A kapott szaturációs modell a mikrobiológiai „felezési idők” -höz hasonlóan a következőképpen interpretálható: az exponenciális konvergencia sebessége 0,031/nap, ami körülbelül 22 napos "feleződési időnek" felel meg; azaz a végső stacionárius szintig hátralévő távolság nagyjából 3 hetente feleződik. A sebesség becslésének relatív hibája 15%, az illeszkedés standard hibája 1,5 g/L, míg a végső telítettségi szint 9,691 ± 0,19 g/L-re becsült. Tekintettel arra, hogy a vizsgálat kohorszát (egészséges, fiatal texasi nők) a lehető leghomogénebbnek tekinthetjük, a kapott eredmények (és az R^2=0,443 érték) azt mutatják - ami az **33. ábra** is látható -, hogy **nagyobb változatosság** mutatható ki a teljes fehérjekoncentrációban az **egyes anyák közötti biológiai különbségeknek köszönhetően** (úgynevezett keresztmetszeti eltérésnek), és jóval **kisebb mértékű az idő függvényében történő** (longitudinális) eltérés. Megjegyzendő, hogy a keresztmetszeti változatosság elkerülhetetlenül sztochasztikus, míg a longitudinális változást az egyes anyák esetében a determinisztikus egyszerű szaturációs modellünk jól leírja.

4.3.1.1.3. Európai kohorszok vizsgálata

Hipotézisünk további vizsgálatához a MilkyBase *(Pacza és mtsai., 2022)* egy olyan adathalmazát használtuk, melyben a kohorsz Európából származott. A 35. ábra az anyatej teljes fehérje koncentrációjának pályáit mutatja a születéstől eltelt idő függvényében, 7 publikáció közölt adatai alapján (összesen 167 adatpont). Az egy-egy pályát meghatározó adatpontok az adott kohorszra vonatkozó mérések adataiból származtatott átlagos fehérjekoncentrációk.



35. ábra. Az anyatej összfehérje koncentrációjának 24 trajektóriája Európai mintákból származó adatok alapján.

A MilkyBase adatbázisban *(Pacza és mtsai., 2022)* tárolt 7 közlemény publikált adatai. A pályák elnevezései a MilkyBase adatbázisban használt kulcs a pálya adatait tároló rekordhoz. .A méréseket különböző Európai országokban és különböző születési körülmények között végezték. Az adatpontok átlagos koncentrációk, jellemzően több mint 10 anya adataiból. A teljes variáció főként olyan körülményeknek köszönhető, mint a terhességi kor, a szülés módja, a földrajzi elhelyezkedés stb. A trajektóriák illesztésére a háromparaméteres egyszerű szaturációs modellt használtuk. Forrás: *(Baranyi-Pacza- és mtsai., 2024)*

Azokban az esetekben, ahol a publikációból le tudtuk vezetni a kohorsz egyedei által generált szórást, az általában 1 - 2 g/L közé esett. Ez megerősíti John és munkatársai *(John és mtsai., 2019)* (34. ábra) előzőekben ismertetett eredményét, ahol az illeszkedés standard hibája, azaz a keresztmetszeti eltérés nagyjából 1,5 g/L volt. Olyan inhomogén populációk esetében, amely például véletlenszerűen császármetszéses, koraszülött stb. szülést tartalmaz (mint például a ábrán szereplő publikációk adatai is), a várható hibának nyilvánvalóan sokkal nagyobbnak kell lennie; a keresztmetszeti szórás akár 5 g/L is lehet. Ezt megerősítették Samuel és munkatársai *(Samuel és mtsai., 2022)* is, akik az Európában élők véletlenszerű kohorszai által generált anyatej-fehérje koncentrációk esetében átlagosan 4,8 g/L körüli szórásáról számolnak be. Más szóval: két v*életlenszerűen kiválasztott anya* tejfehérje-koncentrációja közötti különbség *a szülés után azonos időpontban mérve* várhatóan 3-4-szer nagyobb lenne, mint *egy ugyanazon anyától* származó két *véletlenszerűen kiválasztott időpontban* vett minta – pl. egy kolosztrum és egy érett anyatej minta - fehérjekoncentrációja közötti különbség.

Az anyatej komponensek koncentrációi által alkotott időbeli pályák leírására szolgáló *elsődleges modellek* kialakításakor - ideális esetben - nagy részletességű longitudinális adatokra lenne szükség a vizsgált komponensekre vonatkozóan, lehetőleg anyánként.

Mint láttuk, az egyszerű szaturációs modell **elég rugalmas** ahhoz, hogy mind **egyéni**, mind **populációs szinten illeszkedjen** a fehérjepályákhoz.

A modellnek három paramétere van;

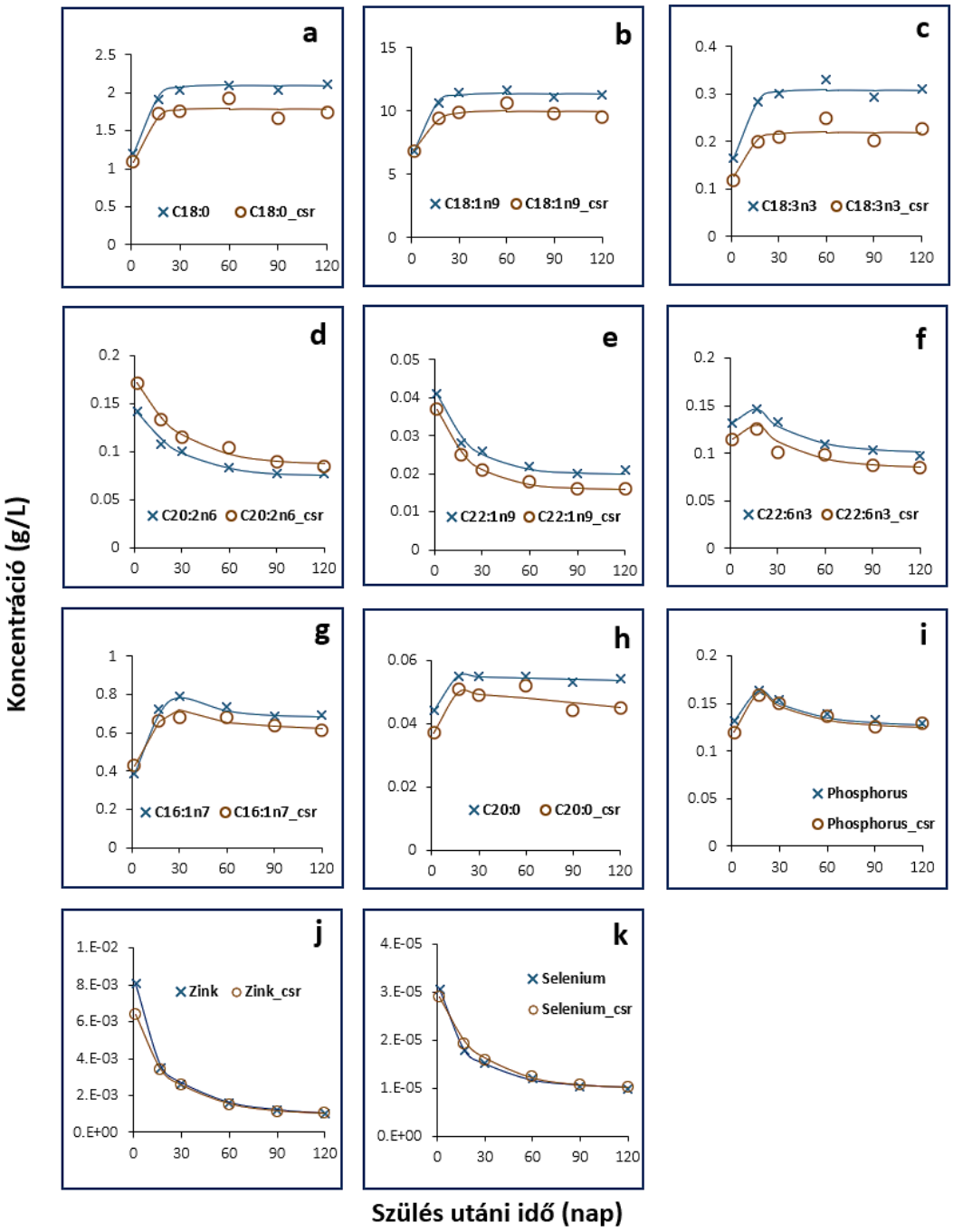
1. a kezdeti koncentráció
2. és a végső koncentráció, amelyhez a trajektória konvergál,
3. valamint ennek az exponenciális konvergenciának a sebessége.

Az 33. ábra és 34. ábra demonstrálja, hogy az egyszerű szaturációs modell illeszkedése populációs szinten **meglehetősen erős** (mindhárom paramétert 20%-nál kisebb relatív hibával becsültük). Egyes egyéni pályák lineárisok voltak ezekben a megfigyelt időintervallumokban, míg mások kifejezett görbületet mutattak.

4.3.1.2. Kétfázisú szaturációs modell

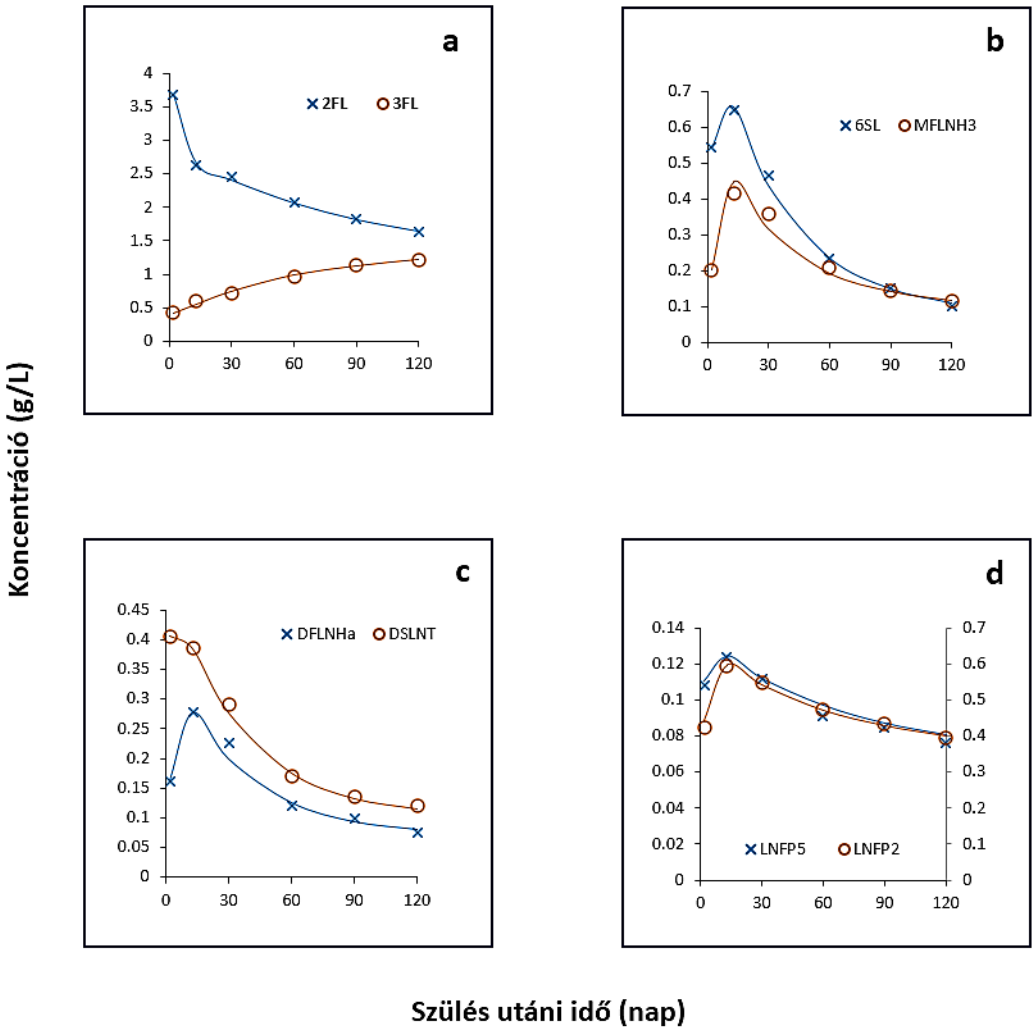
4.3.1.2.1. Egyedi molekulák mintázatai

Felmerült a kérdés, hogy a molekulák egy csoportjára - például a fehérjékre, zsírsavakra, oligoszacharidokra, ásványi anyagokra, vagy vitaminokra - azonosított mintázat ugyanúgy érvényes-e az adott csoport egyes molekuláira is. Ilyen specifikus molekulákra vonatkozóan is álltak rendelkezésre adatok, de egyedi anyák esetén sajnos nagyon kevés. Samuel és munkatársai *(Samuel és mtsai., 2022; Samuel és mtsai., 2019)* publikációiban találtunk ilyen egyéni anyáktól származó egyedi molekulákra vonatkozó adatokat, melyekre populációs pályákat illesztettünk (**36. ábra** és **37. ábra**).



36. ábra. A kétfázisú elsődleges modellünk jól használható a zsírsavmolekulák, valamint ásványi anyagok trajektóriáinak bemutatására.

Mivel a publikáció *(Samuel és mtsai., 2022)* a szülési mód (normál vagy császármetszés) hatását vizsgálták Samuel és munkatársai az anyatej összetételére vonatkozóan, a különböző szülési módokhoz tartozó összeillő pályákat egy-egy grafikonra összevonva ábrázoltuk. (A molekulák nevében a *_csr* jelölés a császármetszést jelöli.) Forrás: *(Baranyi-Pacza- és mtsai., 2024)*



37. ábra. Anyatej oligoszacharidjainak mért koncentrációinak trajektóriái kétfázisú elsődleges modellel illesztve,
Samuel és munkatársai *(Samuel és mtsai., 2019)* publikációjából származó adatok alapján. Forrás: *(Baranyi-Pacza- és mtsai., 2024)*

Jelölések: 2FL: 2'-Fukoszilaktóz; 3FL: 3'-Fukoszilaktóz; 6SL: 6′-szialilaktóz; DFLNHa: Difukoszilakto-N-hexaóz-a; DSLNT: Diszialilakto-N-tetraóz; LNFP5: Lacto-N-fucoPentaose-V; LNFP2: Lacto-N-fucoPentaose-II. A d ábrán az LNFP2 komponenst a másodlagos, jobboldali függőleges tengelyen, a mérjük.

Mivel ezek az adatok nagy kohorszokból származó átlagok, ezért a standard hibájuk kicsi volt. A pályák többsége az egyszerű szaturációs modellt követte, de néhány csak egy gyorsan változó kezdeti, lineáris szakasza után.

Az publikációkban nagyon kevés adat állt rendelkezésre annak bizonyítására, hogy a szaturációs fázist megelőző kezdeti lineáris időszak valóban általánosan létezik. A MilkyBase *(Pacza és mtsai., 2022)* adatbázisban egyetlen ilyen rekord található, Liu és munkatársaié *(Liu és mtsai., 2019)*. Az adatsor meggyőzően bizonyította (lásd **28. ábra**), hogy az anyatej összzsírsav koncentrációja lineáris növekedett a kezdeti kolosztrum időszakban, mielőtt a pálya második egyszerű szaturációs fázisba lépett. Az illesztett meredekségek átlaga 3,5 g/L/nap volt a szülést követő első 6 napon, nagyjából. 12%-os relatív standard hibával. Ezek alapján azt mondhatjuk, hogy az anyatej zsírtartalma nagyjából megduplázódik az első 6 napban.

Tehát feltételezhetjük, hogy egy általános modell esetében a szaturációs fázist először egy kezdeti lineáris „felszálló" időszak előzi meg. Az ilyen általános modellt nevezzük **kétfázisú szaturációs modell**nek, melynek 5 paramétere van, szemben az (kezdeti lineáris szakasz nélküli) **egyszerű szaturációs modellel**, amely csak 3 paraméterrel rendelkezik. (lásd a 3.3.2 Matematikai modellezés, szaturációs modell részt). Ha a két fázis azonos tendenciát mutat, nehéz szétválasztani őket, hiszen a rövid lineáris fázis jól beágyazható az egyszerű szaturációs modellbe. Azonban, ha a két fázis eltérő trenddel rendelkezik, nagy biztonsággal feltételezhetjük a kezdeti lineáris fázis létezését. Ennek fiziológiai magyarázata lehet, hogy az ilyen eltérő tendenciát mutató molekulák azok lehetnek, amelyek termelődését (vagy kiválasztódását) a csecsemő születése indítja be hirtelen, majd a rendszer egy „telítettségi állapothoz" konvergál.

Az ilyen kétfázisú modell felépítésének a nehézségét a gyakorlati validálása jelenti. Sajnos a kezdeti fázis létezése csak olyan adatsorok alapján mutatható ki, amelyek legalább két mintát tartalmaznak a kolosztrumból, de az ilyen adatsorok nagyon ritkák.

Ezért illesztettük a kétfázisú modellt a Samuel és munkatársai *(Samuel és mtsai., 2022; Samuel és mtsai., 2019)* azon adataira (**36. ábra** és **37. ábra**), ahol a kolosztrum utáni adatokra illesztett szaturációs modellből az egyetlen kolosztrumból származó adatpont kiugró volt. A modell illesztése során, először a kolosztrum időszak utáni adatokra egy egyfázisú, háromparaméteres szaturációs modellt illesztettünk nemlineáris regresszióval, majd a kolosztrumból származó egyetlen pontot kombináltuk az első kolosztrum utáni pont illesztett értékével, ami a kezdeti lineáris fázis sebességének alsó

(vagy felső) határát mutatja. Azaz a lineáris fázis meredekséget az első két pont határozza meg.

Ez a kétfázisú modell jellemzi a Samuel és munkatársai *(Samuel és mtsai., 2022; Samuel és mtsai., 2019)* által közzétett molekulák összes trajektóriáját. A **36. ábra a-c** grafikonjain olyan három zsírsavmolekula (C18:0, C18:1n9, C18:3n3) populációs pályája látható, amelyek a végső koncentrációs szintig gyorsan növekednek. A **d-e** grafikonok (C20:2n6, C22:1n9) mindvégig csökkenő tendenciát mutatnak ott, ahol a lehetséges kezdeti lineáris fázis beágyazódhat a későbbi szaturációs modellbe. Az **f-h** grafikonokon ábrázolt C22:6n3, C16:1n7 és C20:0 zsírsav molekulák koncentrációja az anyatejben először lineáris növekszik, majd egy exponenciálisan csökkenő pályát követve konvergál a végső szinthez. Az **i-k** grafikonok hasonlóan jó illeszkedést mutatnak az ásványi anyagok (foszfor, cink szelén) esetében, valamint az **37. ábra** grafikonjai az oligoszacharidokra (2FL: 2'-Fucosyllactose; 3FL: 3'-Fucosyllactose; 6SL: 6′-Sialyllactose; MFLNH3: Monofucosyllacto-N-hexaose-III; DFLNHa: Difucosyllacto-N-hexaose-a; DSLNT: Disialyllacto-N-tetraose; LNFP5: Lacto-N-fucoPentaose-V; LNFP2: Lacto-N-fucoPentaose-II.).

A modellek illesztési adatait az alábbi táblázatok mutatják.

3. táblázat A 36. ábra a-c grafikonjain ábrázolt három zsírsavmolekula (C18:0, C18:1n9, C18:3n3) illesztési adatai

ábra	**36_a**	**36_a**	**36_b**	**36_b**	**36_c**	**36_c**
trajektória	**C18:0**	**C18:0 _csr**	**C18:1n-9**	**C18:1n-9 _csr**	**C18:3n-3**	**C18:3n-3 _csr**
a	0	0	0.5	0.5	0	0
λ	1.5	1.5	6	6	1.5	1.5
r	0.12	0.12	0.12	0.12	0.12	0.12
y_0	1.195	1.093	6.811	6.866	0.166	0.128
y_{End}	2.1	1.8	11.4	10	0.31	0.22
se(fit)	0.02779	0.0624	0.13937	0.253	0.00827	0.01148

4. táblázat A 36. ábra d-e grafikonjain ábrázolt két zsírsavmolekula (C20:2n6, C22:1n9) illesztési adatai

ábra	36_d	36_d	36_e	36_e
trajektória	**C20:2n6**	**C20:2n6 _csr**	**C22:1n9**	**C22:1n9 _csr**
a	0	0	0.5	0.5
λ	1.5	1.5	6	6
r	0.035	0.035	0.05	0.05
y_0	0.142	0.172	0.041	0.037
y_{End}	0.074	0.086	0.02	0.016
se(fit)	0.00191	0.00254	0.00073	0.00035

5. táblázat A 36. ábra f-h grafikonjain ábrázolt három zsírsavmolekula (C22:6n3, C16:1n7 és C20:0) illesztési adatai

ábra	36_f	36_f	36_g	36_g	36_h	36_h
trajektória	**C22:6n3**	**C22:6n3 _csr**	**C16:1n7**	**C16:1n7 _csr**	**C20:0**	**C20:0 _csr**
a	0.008	0.008	0.02	0.02	0.0025	0.0025
λ	6	6	23	23	6	6
r	0.035	0.035	0.035	0.035	0.001	0.001
y_0	0.132	0.115	0.384	0.427	0.044	0.037
y_{End}	0.1	0.084	0.68	0.62	0.04	0.002
se(fit)	0.00179	0.00423	0.01199	0.01454	0.00039	0.00156

6. táblázat A 36. ábra f-h grafikonjain ábrázolt három ásványi anyag (foszfor, cink szelén) illesztési adatai

ábra	36_i	36_i	36_j	36_j	36_k	36_k
trajektória	**Phosphorus**	**Phosphorus _cs**	**Zink**	**Zink _csr**	**Selenium**	**Selenium _csr**
a	0.011	0.011	-0.0007	-0.0007	0	0
λ	6	6	6	6	6	6
r	0.035	0.035	0.035	0.035	0.035	0.035
y_0	0.1316	0.1196	0.0080478	0.0064171	0.000306	0.000029
y_{End}	0.127	0.1245	0.001	0.001	0.00001	0.00001
se(fit)	0.00201	0.00213	6.23E-05	3.52E-05	1.7E-07	2.6E-07

7. táblázat A **37. ábra** a-b grafikonjain ábrázolt az oligoszacharidok (2FL, 3FL, 6SL, MFLNH3) illesztési adatai.

ábra	37_a	36_a	37_b	37_b
trajektória	**2FL**	**3FL**	**6SL**	**MFLNH3**
a	-0.230	0.01	0.06	0.08
λ	6	6	6	6
r	0.01	0.015	0.028	0.028
y_0	3.691	0.422	0.543	0.201
y_{End}	1.1	1.4	0.08	0.1
se(fit)	0.01523	0.01670	0.00921	0.01735

8. táblázat A **37. ábra** c-d grafikonjain ábrázolt az oligoszacharidok (DFLNHa, DSLNT, LNFP5, LNFP2) illesztési adatai.

ábra	**37_c**	**37_c**	**37_d**	**37_d**
trajektória	**DFLNHa**	**DFLNHa**	**LNFP5**	**LNFP2**
a	0.04	0.01	0.005	0.045
λ	6	6	6	6
r	0.028	0.028	0015	0.015
y_0	0.162	0.405	0.108	0.422
y_{End}	0.07	0.1	0.07	0.35
se(fit)	0.00937	0.00547	0.00253	0.00349

A **36. ábra** szemlélteti azt, amikor trajektóriákat úgy hasonlítunk össze, hogy egyetlen körülmény (esetünkben a szülési módok, normál/császármetszés)) hatását vizsgáljuk meg. Ez a másodlagos modell felállításához vezet

4.3.2. Másodlagos modell

Mint korábban is említettük (lásd.3.3.2 Matematikai modellezés, szaturációs modell), az anyatej komponensek elsődleges modelljei a komponensek koncentrációjának az időbeli változását, azaz trajektóriáját írják le többé-kevésbé állandó körülmények között, míg a másodlagos modellek e körülményeknek az elsődleges modell paramétereire gyakorolt hatását jellemzik. A fő ok, amiért a két típusú modell nem összevonható, mechanisztikus: a körülmények azt befolyásolják, hogy egy komponens mennyisége hogyan változik az idővel (milyen gyorsan nő vagy csökken, azaz a változás sebességét), nem pedig közvetlenül magának a komponensnek a szintjét. A komplexitás redukálására az elsődleges modell paraméterkészlete használható, mellyel az egész időfüggő pálya helyettesíthető.

Mint láttuk, az egyszerű szaturációs modellnek három paramétere van: a kezdeti és a végső koncentrációs szint, valamint a szinthez való exponenciális konvergencia sebessége. Az opcionális kezdeti kolosztrumbeli lineáris fázisnak két paramétere van: a vizsgált komponens koncentrációja a születés napján, és a koncentráció változás lineáris növekedésének (vagy csökkenésének) a sebessége.

Kutatásunk során megvizsgáltuk azt is, hogy az elsődleges modell harmadik paraméterét, a végső koncentrációs szintet (másodlagos paraméter) hogyan befolyásolja a földrajzi helyzet.

4.3.2.1. Földrajzi különbségek.

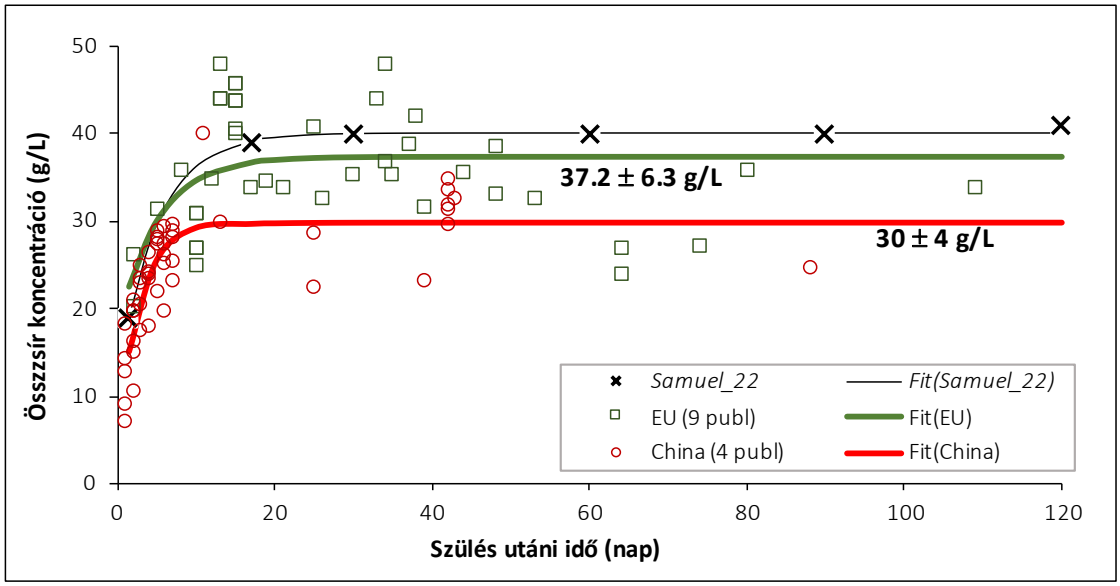
A MilkyBase adatbázisban a földrajzi régió az egyik magyarázó változó, míg a különböző anyatej komponensek időbeli pályái dinamikus válaszváltozók *(lásd 4.1.1 A MilkyBase adatbázis felépítése*), így könnyen vizsgálható volt a földrajzi helyzettől való függőség. A pontosabb adatkiértékelés érdekében további adatokkal egészítettük ki az alap MilkyBase-t (*Lásd 3.2 Az Alap MilkyBase adatbázis kibővítése irányított kereséssel*).

A teljes zsírkoncentráció nem változott jelentősen az átmeneti és az érett fázisban sem a John és munkatársai *(John és mtsai., 2019)* Texasi kohorszában (35±7 g/L), sem a Samuel és munkatársai *(Samuel és mtsai., 2022)* Európai kohorszában (40±15 g/L). Az utóbbi adatsor azonban kezdeti növekedést mutatott a kolosztrumban is. Az ok, amiért ezt a kezdeti időszakot nem tekintjük a szaturációs modell részének, a Liu és társainak

(Liu és mtsai., 2019) adatai, amelyek robusztus lineáris trendet mutatnak a kolosztrumban lévő összes zsírtartalomra vonatkozóan.

Liu és munkatársai *(Liu és mtsai., 2019)* kínai kohorszában az anyatej összes zsírtartalma a kezdeti gyors növekedés után a 42. napig 25 és 30 g/l között maradt. Wu és munkatársai *(Wu és mtsai., 2021)* szintén megerősítették, hogy az érett anyatej zsírkoncentrációja 25 g/l körül volt kínai kohorszukban

A Samuel és munkatársai *(Samuel és mtsai., 2022)* adataira illesztett két fázisú szaturációs görbéket (elsődleges modellek) vettük alapul az előrejelzéshez és az összehasonlításhoz. Ennek oka, hogy itt a mintákat nagyjából azonos időpontokban gyűjtötték az anyáktól, és a közzétett koncentrációk nagy minták átlagai, ezért standard hibáik kicsik. A MilkyBase *(Pacza és mtsai., 2022)* adatai alapján létrehozott **38. ábra**, az anyatej -zsírkoncentrációi közötti különbséget mutatja az európai és kínai kohorszok esetében. Mindkét esetben a végső zsírkoncentráció nagyjából kétszerese a kiindulási értéknek, de az európai kohorszra jellemző (átlagos) zsírkoncentráció (ábrán zöld vonallal jelzett) a megfigyelési idő alatt végig 25-30%-kal magasabb volt, mint a kínai kohorszé (ábrán piros vonallal jelzett).



38. ábra. Az anyatej teljes zsírkoncentrációjának változása európai és kínai kohorszokban.

Forrás: *(Baranyi-Pacza- és mtsai., 2024)*

Az alapmodell (fekete folytonos vonal **38. ábra**) olyan átlagos zsírkoncentrációkra illeszkedik, melyhez egy-egy időpontban több száz anya szolgáltatott mintát. Az eredeti mérések szórása 10-20 g/L (amint azt a Samuel és munkatársai *(Samuel és mtsai., 2022)* táblázatai mutatják), így a beillesztett pontok hibája kevesebb, mint 1 g/L. Ezen átlagok standard hibái kevesebbek mint 1 g/l, 100-200 anya esetében szinkronban mérve. Ezek az átlagos koncentrációk észrevehetően sima mintázatot követnek, még akkor is, ha az egyes egyénenkénti pályák zajosabbak, hasonlóan a ábrán látható fehérjék esetében. Az átlagok kis hibáját az okozza, hogy a kohorsz mérete kompenzálja a nyers adatok - egyedi anyák biológiai különbségei által okozott - nagy (kb. 15 g/L, ahogyan az Samuel és munkatársai *(Samuel és mtsai., 2022)* táblázataiból kiszámítható) standard eltéréseit. Az ANOVA teszt megerősítette, ami a grafikonon is jól ábrázolódik, hogy ez az egyedi anyák által okozott eltérés szignifikánsan nagyobb, mint a földrajzi elhelyezkedés (a kohorsz az EU-ból vagy Kínából származik-e) okozta eltérés.

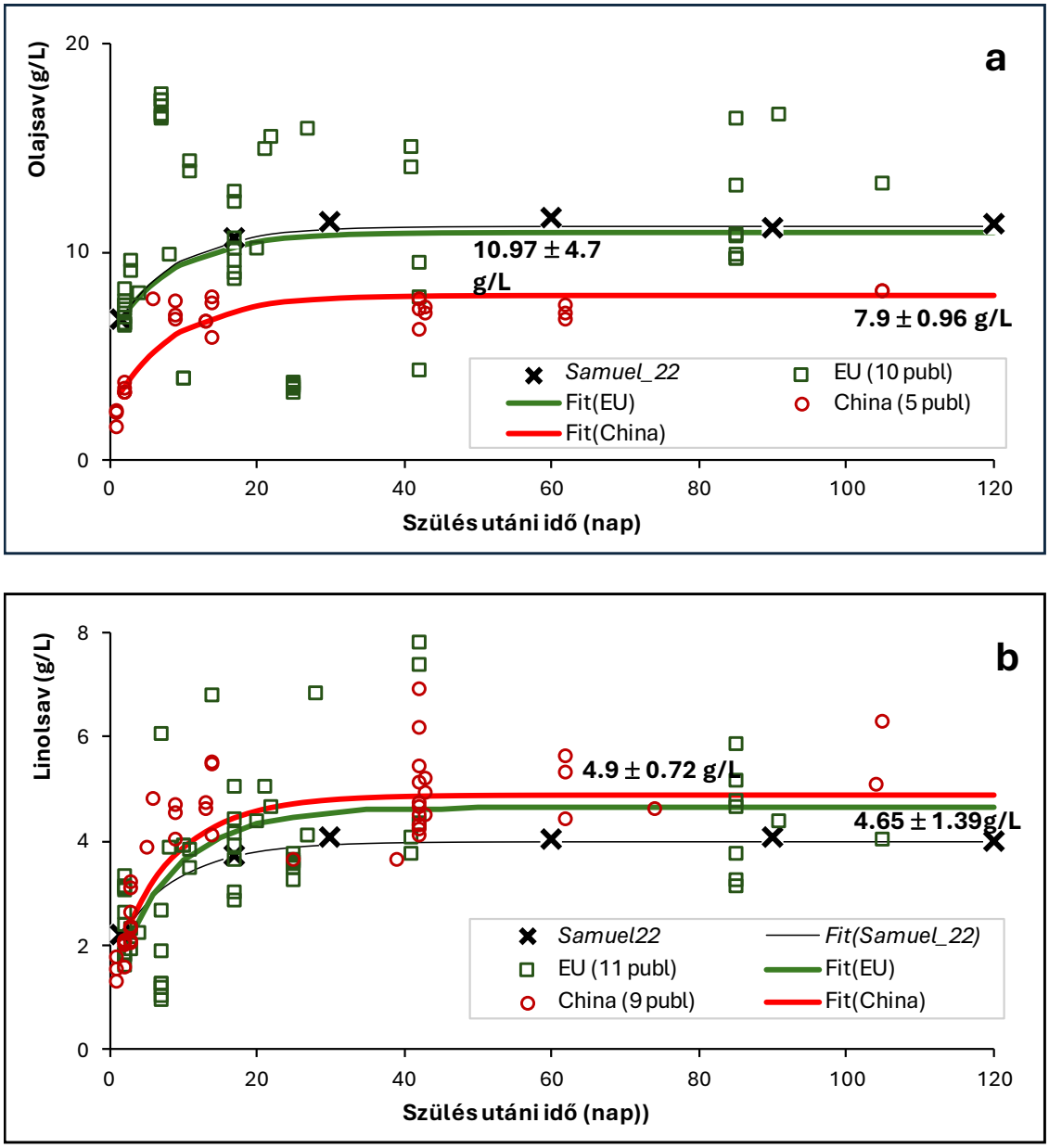
A **38. ábra** látható 9 független publikációból származó európai adatok (zöld négyzet) kisebb - 10 és 50 elemszám közötti - kohorszok átlagai. Az illesztett szaturációs görbe közel áll a Samuel és munkatársai *(Samuel és mtsai., 2022)* adataira illesztett referencia-modellünkhöz: az 1. napon mért zsírkoncentráció 22,2 g/L-ről a negyedik vizsgálati hónap végére 37,2 g/L-re emelkedett, míg ezek az értékek az referencia-modell esetében 19 g/L, illetve 40 g/L. A kínai kohorszokból származó adatok esetében ugyanezek a paraméterek 15 g/L és 30 g/L értéket mutatnak, amely a referencia -modelltől még távolabbra esik. Az európai kohorszok esetében az illeszkedés standard hibája 6,3 g/L, míg a kínai kohorszok esetében ez 4 g/L. Ezek a hibák lényegesen nagyobbak, mint Samuel és munkatársai *(Samuel és mtsai., 2022)* által közölt adatok standard hibái, de figyelembe véve, hogy az európai és a kínai adatpontokat jóval kisebb kohorszok generálták, és hogy a heterogenitás nemcsak a kohorszokból, hanem más tényezőkből is adódik (pl. az anya étrendje, a szülés módja, a csecsemő neme, a mérések mögött álló módszertan stb. az adatokat előállító publikációkban leírtak szerint), ez érthető. Ez azonban továbbra is lényegesen kisebb, mint a földrajzi helyzet okozta variabilitás.

Ezért, hasonlóan a fehérjékhez, az anyatej zsírtartalma esetében is az **egyes anyák közötti *sztochasztikus biológiai különbségből származó keresztmetszeti eltérés* nagyobb**, **mint a hosszanti vagy a földrajzi eredetű eltérés** (legalábbis az Európa és Kína tekintetében).

Az ábrán jól látható, hogy Wu és munkatársai *(Wu és mtsai., 2021)* megállapításával összhangban a kínai kohorszok zsírkoncentrációja a megfigyelési intervallum 120 napja alatt végig alacsonyabb, mint az európai kohorszoké.

Végül felmerül a kérdés, hogy ez az összzsírsav koncentrációban látható különbség az egyes zsírsavmolekulákra is érvényes-e. Erre a MilkyBase *(Pacza és mtsai., 2022)* adatbázis molekulaspecifikus adatainak felhasználásával kaphatunk választ..

A legnagyobb mennyiségben előforduló zsírsavmolekulák az olajsav (C18:1n-9) és a linolsav (C18:2n-6), ezek összege az anyatej teljes zsírsavtartalmának több mint felét adja. A **39.ábra** az olajsavra (**39.ábra** a) és a linolsavra (**39.ábra** b) vonatkozóan mutatja a MilkyBase *(Pacza és mtsai., 2022)* rendelkezésre álló adatait. Amint látható, a kínai kohorszokból származó koncentrációkban általában alacsonyabb az olajsav szintje, mint az európai kohorszokban, azonban a linolsav esetében fordított a helyzet, különösen az első két hétben. A Samuel és munkatársai *(Samuel és mtsai., 2022)* adatainak átlagait megerősítik az európai adatok átlagai, de ez utóbbiaknak sokkal nagyobb a standard hibájuk (6,3 g/L, illetve 1,9 g/L), mint amit a kínai kohorszok adatainak illesztése mutat (0,76 g/L, illetve 0,73 g/L).



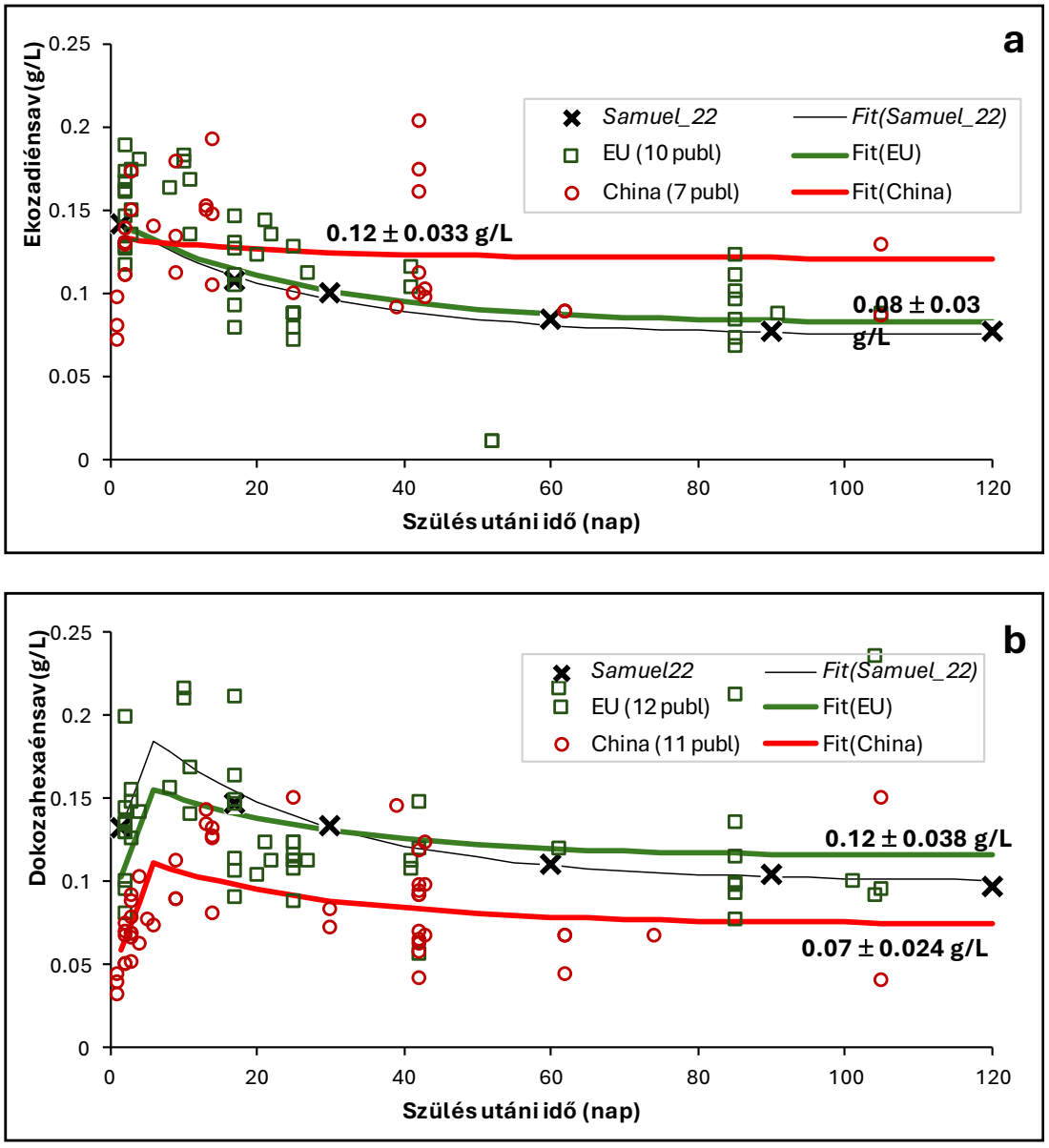
39.ábra. Az anyatej olajsav (C18:1 n-9) és a linolsav (C18:2 n-6) koncentrációja a szülést követő 120 napban.

Forrás: *(Baranyi-Pacza- és mtsai., 2024)*

A MilkyBase-ben *(Pacza és mtsai., 2022)* az eikozadiénsavra (C20:2n-6) és a dokozahexaénsavra (DHA; C22:6n-3) vonatkozóan is rendelkezésre állnak adatok. Az EU független adatainak illesztése mindkét esetben közel áll a Samuel és munkatársai *(Samuel és mtsai., 2022)* adatainak illesztéséhez. A **40. ábra**. (a.)-n a kezdeti lineáris fázist nem lehetett azonosítani, mivel hiányoznak a kolosztrumból származó adatok. A kínai kohorszok által szolgáltatott adatok az átlaguk körül szóródtak, így nem lehetett tendenciát megállapítani. A DHA esetében a kezdeti lineáris fázis mindkét földrajzi

régió esetében szignifikáns volt. A **40. ábra** rámutat, hogy míg az eikozadiénsav esetében a kínai kohorszokból származó koncentrációk általában magasabbak, mint az európai kohorszokból származó koncentrációk, míg a DHA esetében ez éppen fordítva van.

A **földrajzi elhelyezkedés okozta különbség** a teljes zsírsavkoncentrációk között tehát **nem arányosan oszlik meg a különböző zsírsavmolekulák között**.



40. ábra. Az anyatej eikozadiénsav (C20:2n-6) és a dokozahexaénsav (DHA; C22:6n-3) koncentrációja a szülést követő első 120 napban

Forrás: *(Baranyi-Pacza- és mtsai., 2024)*

5. KÖVETKEZTETÉSEK, JAVASLATOK

"No mathematical theories can be accepted by biologists without a most careful experimental verification" (Gauze, 1934)

Az adatbázis fejlesztése a publikációkból származó nagy mennyiségű adat rögzítésére összpontosított, függetlenül azok céljától, a vizsgált molekuláktól vagy kohorszoktól. Egy ilyen adatbázisnak viszonylag nagy volumenűnek kell lennie (figyelembe véve a tej biokémiai összetételének összetettségét), hogy átlépje azt a kritikus tömeget, amelytől kezdve az eredményeket jelentősnek tekinthetjük. Ezért, különösen az adatbázis-fejlesztés kezdetén, a szerzők által rendelkezésre bocsátott adatok mennyisége nagy szerepet játszott annak kiválasztásában, hogy mely dolgozatokat érdemes digitalizálni és rögzíteni. A könnyebb adatátvitel érdekében előnyben részesítettük azokat a publikációkat, melyek sok adatot tartalmazó táblázatokat közölnek. A fejlesztés fő kihívása a változatosság és a hitelesség volt, a "mit rögzítsünk" komoly döntés, ami akár elfogult is lehet. Módszerünk nem tartozott az adatbányászathoz, mivel az adatbázisba bevitt kiválasztott publikációk messze nem teljes körűek. Inkább egy prototípus, amely a felhasználóbarát adatgyűjtéshez egy olyan ontológiát biztosít, amely kompatibilis a matematikai modellezéssel. A közzétett adatok digitalizálása meglehetősen munkaigényes feladat, mivel például az ellenőrzés feladatát sem lehet teljesen automatizálni.

A szisztematikusan szervezett adatbázison a felhasználók

- automatizált keresést és statisztikákat futtathatnak, amelyek
- segíthetnek az adathiányok (azaz az új kutatási ötletek) azonosításában;
- a publikációkban található hibák felkutatásában; valamint
- a minták felismerésében, esetleg modellezésében és optimalizálásában az egészséges csecsemő és anya érdekében.

Összességében a MilkyBase adatbázis a jobb táplálkozási irányelvek kidolgozásához nyújthat segítséget. A jövőbeni anyatej-kutatások kiterjeszthetik az adatbázist és további anyatej összetételt befolyásoló tényezőket is vizsgálhatnak.

Az általunk létrehozott modell megmutatja, hogy az anyatej összetevők időbeli változásának mintázata egy-egy anya esetében -feltételezve, hogy az anyai feltételek nem változnak- matematikai eszközökkel előre jelezhető. Ez a mintázat nem lineáris , és olyan fázisokra osztható, melyek a hagyományos kolosztrum- átmeneti – érett tej felosztást igazolják. Egy-egy anya tejösszetételének ideális pályája az első héten gyors változáson megy keresztül, de ezt követően fokozatosan felveszi a szaturációs folyamatokra jellemző pályát, azaz lassú, de exponenciális ütemben konvergál az stacionárius állapothoz. A keresztmetszeti vizsgálatok nem tudják megerősíteni ezt a mintát, mivel mint láthattuk, a személyes (például genetikai) jellemzők okozta változékonyság nagyobb, mint az idő vagy más (például földrajzi) körülmények okozta változékonyság, ehhez nagy részletességű longitudinális adatra lenne szükség, lehetőleg komponensenként és anyánként. Továbbá - modellünk eredményei alapján - javasoljuk, hogy az anyatejkutatások során a kutatási tervek nem-ekvidisztáns mintavételi időpontokat tartalmazzanak, a szülést követő első két hétben gyakoribb mintavételezéssel. Ez a kutatói megközelítés hatékonyabban tudná jellemezni a kolosztrum időszak gyors változásait és pontosabb képet kaphatnánk az anyatej időbeli dinamikájáról.

6. ÚJ TUDOMÁNYOS EREDMÉNYEK

A MilkyBase adatbázis és újdonságai

1. A kutatás eredményeként – az anyatej kutatások területén először- **létrehoztunk egy adatbázist -a MilkyBase adatbázis-** ,az anyatej összetevőiről tudományosan publikált adatok alapján, mely **az anyatej összetételére vonatkozó** rekordokat tárolja és rendszerezi. Az adatbázis létrehozásakor a következő innovációkat vezettük be:

   1.1. **Megalkottuk az ontológiát**, amely **az anyatej összetételére ható tényezőkre (**pl. mérési körülmények, az anya/gyermek jellemzői, valamint környezeti és előzményi körülmények) **adott válaszként** tekinti az összetételre vonatkozó adatokat, amelyek a **magyarázó és válaszváltozókban** kerülnek beírásra. A rekordok „ok-okozati" hatást reprezentálnak, azaz különböző magyarázó körülmények - amelyek mellett a megfigyelések történtek - leképezését az anyatej összetételére, mint válaszváltozóra.

   1.2. Az adatbázis szerkezetét úgy alakítottuk ki, hogy alkalmas legyen az anyatej komponensek és összetétel időbeli változásának a követésére, azaz nem csak statikus, hanem **dinamikus (időfüggő) állapotok** is **rögzíthetők legyenek.** Ehhez a változók időbeli változását egy dedikált táblával reprezentáljuk, amelyek a pályák („időpont, mért érték") értékpárjait tartalmazza és az ezekre hivatkozó mutatók biztosítják, hogy az időfüggőség az adott mezők természetes attribútuma legyen.

   1.3. Bevezettük a **kiterjesztett numerikus változók**at, melyek lehetőséget adnak a bizonytalanságok feljegyzésére és mérésére, valamint az adatok **direkt és indirekt (származtatott) formá**ban való rögzíthetőségét, elősegítve ezzel a szélesebb körű adatgyűjtést és elemzést.

   1.4. Az adatbázis szerkezetét úgy definiáltuk, hogy az elemek csoportosítása **hierarchikus fa struktúrát** követ. A létrehozott struktúra lehetővé teszi az anyatej összetevőit tartalmazó gráf létrehozását és támogatja többszintű elemzéseket.

Az anyatej molekuláris összetételében lévő mintázatok felismerése

Az anyatejkutatások során először alkalmaztunk matematikai modellezést az anyatej összetevők jellemzésére, melynek során következő eredményeket kaptuk.

2. Az elsődleges modellezés során létrehoztunk egy **egyfázisú egyszerű szaturációs modellt**

$$y(t) = y_0 \cdot e^{-r \cdot t} + y_{End}(1 - e^{-r \cdot t}) \qquad (0 \le t, r \ge 0)$$

   A modell három paraméterrel rendelkezik, az y_0 kezdeti koncentráció és az y_{End} végső koncentráció, valamint a végső koncentrációhoz való exponenciális konvergenciának a sebessége (r).

3. A MilkyBase adatbázisban rögzített adatok alapján az anyatej **teljes fehérje koncentrációját** vizsgálva az egyfázisú egyszerű szaturációs modell segítségével következő eredményekre jutottunk:

   3.1. Megállapítottuk, hogy a létrehozott egyszerű szaturációs modell illeszkedése populációs szinten meglehetősen **erős** (mindhárom paramétert 20%-nál kisebb relatív hibával becsültük), valamint elég **rugalmas** ahhoz, hogy mind egyéni, mind populációs szinten illeszkedjen a **fehérjepályákhoz.**

   3.2. Megállapítottuk, hogy **nagyobb változatosság** mutatható ki a teljes fehérjekoncentrációban az **egyes anyák közötti biológiai különbségeknek köszönhetően** (úgynevezett keresztmetszeti eltérésnek- szórás kb.5 g/L), és jóval **kisebb** mértékű az **idő függvényében** történő (longitudinális eltérés- szórás kb. 1-2 g/L) eltérés.

4. Az anyatej összetevők időbeli pályáinak leírására megalkottunk egy **általános kétfázisú szaturációs modell**t (melynek speciális esete az egyfázisú egyszerű szaturációs modell), ahol a szaturációs fázist először egy kezdeti lineáris időszak előzi meg

$$y(t) = \begin{cases} y_0 + a \cdot t & (0 \le t < \lambda) \\ y_\lambda \cdot e^{-r \cdot (t-\lambda)} + y_{End}(1 - e^{-r \cdot (t-\lambda)}) & (\lambda \le t) \end{cases}$$

ahol $y_\lambda = y_0 + a \cdot \lambda$, $0 \le r$, $0 \le \lambda$.

A modellnek 5 paramétere van, a kezdeti koncentráció (y_0) és végső koncentráció (y_{End}), az anyatej komponens koncentrációjának változási gyorsasága a kezdeti szakaszban (a), a szaturációs ráta (r) és a kezdeti (kolosztrum) fázis időtartama (λ).

Az opcionális kezdeti – kolosztrumbeli - lineáris fázisnak két paramétere van: a vizsgált komponens koncentrációja a születés napján, és a koncentráció változás sebessége. Ezek a fázisok igazolják a hagyományos anyatej időbeli változás felosztást (kolosztrum – átmeneti – érett).

5. A kétfázisú modellt a MilkyBase adatbázisban rögzített **egyedi molekulák** adataira illesztve a következő eredményeket kaptuk:

   5.1. A kétfázisú modell jó illeszkedést mutatott a vizsgált zsírsav molekulák, oligoszacharidok valamint ásványi anyagok esetében is.

   5.2. Azt tapasztaltuk, hogy a C18:0, C18:1n9, C18:3n3 zsírsav molekulák trajektóriái a végső koncentrációs szintig gyorsan növekednek, a C20:2n6, C22:1n9 zsírsav molekulák, valamint a cink és szelénium mindvégig csökkenő tendenciát mutatnak. Továbbá a C22:6n3, C16:1n7 és C20:0 zsírsav molekulák, valamint a foszfor koncentrációja az anyatejben először lineáris növekszik, majd egy exponenciálisan csökkenő pályát követve konvergál a végső szinthez.

6. Másodlagos modellezés segítségével az anyatej összzsírsav tartalmának végső koncentráció szintjére gyakorolt földrajzi hatás vizsgálatakor a következő eredményeket kaptuk:

   6.1. Európa és Kína tekintetében az anyatej **teljes zsírtartalma esetén is** az egyes anyák közötti **sztochasztikus biológiai különbségből származó keresztmetszeti eltérés** (kb. 15 g/L**), nagyobb**, **mint** a **hosszanti** avagy a földrajzi eredetű eltérés (az európai kohorszok esetében 6,3 g/L, míg a kínai kohorszok esetében ez 4 g/L).

   6.2. Megállapítottuk, hogy a teljes zsírsavkoncentrációk közötti különbség, melyet a földrajzi elhelyezkedés okoz, **a nem arányosan oszlik** meg a különböző zsírsavmolekulák között.

7. GYAKORLATBAN ALKALMAZHATÓ EREDMÉNYEK

Elsődleges célunk egy olyan ontológia definiálása volt, amely annak feltárását segíti, hogy az anyatej összetétele hogyan függ különböző tényezőktől. Másodlagos cél, hogy egy ilyen ontológia használható legyen másfajta élelmiszerek kutatásában is, ahol a felhasználók saját adataikat is hasonló formátumban tárolhatják.

A létrehozott ontológia és az alkalmazott modellezési módszerek gyakorlatban is alkalmazható eredményei a következők:

1. A kutatók és a táplálkozási szakemberek a MilkyBase segítségével különböző tényezők alapján azonosíthatják a tej összetételének mintázatait. Ezek a minták segíthetnek annak megértését, hogy az anyai körülmények, a szülés körülményei vagy egyéb környezeti hatások hogyan befolyásolhatják az anyatej összetételét.
2. Az adatbázis segíthet feltárni a „kevésbé kutatott" területeket, irányt mutatva a további kutatási projekteknek.
3. A MilkyBase olyan platformként szolgál, ahol a felhasználók szabványosított formátumban bevihetik saját adataikat, elősegítve ezzel az élelmiszer-összetételre vonatkozó adatok **formátumainak a szabványosítását**, az adatbázisok kompatibilissé tételét és teljes potenciáljának kihasználását.
4. A MilkyBase mezőinek hierarchikus szervezése megkönnyíti az adatkészleteken belüli és azok közötti összehasonlítást, a statisztikai módszerek alkalmazását az adatok elemzésére, lehetőséget nyújtva az adathalmazon belül összefüggések és korrelációk feltárására. A könnyű kezelhetőség a kutatókat arra ösztönzi, hogy az általuk létrehozott adatokat ebben a formában rögzítsék, **elősegítve az ellenőrzést, együttműködést és adatmegosztást**.
5. .Az anyatej összetevők matematikai modellezésének bevezetése megteremti a lehetőségét a bennük rejlő mintázatok objektív alapon való felismerésére. Az előrejelzések, nagy előnyt jelenthetnek a kísérleti tervezésre és az adatok értelmezésére, valamint a kutatási és innovációs területek kiválasztására. A terület a gépi tanulási technikák integrálásával és az összetett és nagy adathalmazok kezelésével fejlődhet tovább. A jövő kihívásai közé tartozik a pontosabb és

hatékonyabb számítási modellek kifejlesztése, amelyek képesek kezelni a tudományos problémák növekvő összetettségét és méretét. A mesterséges intelligencia térhódításával lehetőséget nyújt arra, hogy ezek a programok könnyebben hasznosíthassák a méréseket, a felhalmozott közös tudást, segítve ezzel a döntéshozatalt is.

6. Továbbá - modellünk eredményei alapján - javasoljuk, hogy az anyatejkutatások során a kutatási tervek nem-ekvidisztáns mintavételi időpontokat tartalmazzanak, a szülést követő első két hétben gyakoribb mintavételezéssel. Ez a kutatói megközelítés hatékonyabban tudná jellemezni a kolosztrum időszak gyors változásait és pontosabb képet kaphatnánk az anyatej időbeli dinamikájáról.

8. ÖSSZEFOGLALÁS

Az anyatej a csecsemők optimális tápláléka, mely a növekedéshez, fejlődéshez és egészséghez nélkülözhetetlen tápanyagok és bioaktív összetevők összetett keveréke. Az anyatej kutatás az utóbbi évtizedekben rohamosan fejlődik, azonban a kiterjedt kutatások ellenére az anyatej biokémiai komplexitása még továbbra sem feltárt. A kutatás során létrehozott MilkyBase adatbázis egyik célja az volt, hogy rávilágítson arra, hogyan lehet ezt a tudáshiányt betölteni és megmutassa, hogy milyen előnyökkel járhatnak a nagy adatszolgáltatási módszerek az anyatej kutatás, és szélesebb körben a táplálkozástudomány számára.

Az adatbázis rekordjai a tudományos publikációkból származnak, melyeket gépi keresőmotoros és manuális módszerekkel gyűjtöttünk. Ez a kettős megközelítés biztosította a pontosságot és telje körűséget. A rögzített adatokat széleskörűen és körültekintően választottuk ki, digitalizáltuk és aktualizáljuk a megjelent tudományos publikációk alapján. Az adatszerkezet lehetőséget nyújt anyatejjel kapcsolatos részletes adatok rögzítésére a biokémiai összetételéről, valamint azok keletkezésének körülményeiről, mint például az anyai tényezőkről (étrendről, egészségi állapot, a szoptatási időszak), a csecsemő karakterisztikájáról, illetve a mérési körülményekkel kapcsolatosan.

A MilkyBase adatbázis egy VBA szervízmakrókkal kiegészített Excel munkafüzet, mely könnyű kezelhetősége révén a nem számítástechnikai háttérrel rendelkező kutatóknak is hozzáférhető. Cél, hogy az adatbázis forrást és sablont nyújtson a kutatók és laboratóriumok számára, hogy saját adataikat ebbe a szabványosított formátumba helyezzék. A platform használata megkönnyítheti a kutatók közötti együttműködést és adatmegosztást, ezzel egyfajta Wiki-filozófiát követő tudásmegosztást kezdeményezve.

Az adatbázis különféle elemzési módszereket támogat az anyatej összetételében rejlő mintázatok feltárására. A mezők hierarchikus szervezése megkönnyíti a statisztikai elemzéseket, a bizonytalanságok számszerűsítése (kiterjesztett numerikus változók) a mérési hibák és a változékonyság figyelembe vételét segíti elő, míg a dinamikus változóforma az anyatej összetétel változásainak időbeli dinamikáját segít feltárni.

A kutatók és a táplálkozástudományi szakemberek a MilkyBase segítségével különböző tényezők alapján azonosíthatják a tej összetételének mintázatait. Ezek a minták segíthetnek annak megértését, hogy az anyai körülmények, a szülés körülményei vagy egyéb környezeti hatások hogyan befolyásolhatják az anyatej összetételét.

A kutatás második fázisában a matematikai modellezés módszerét használtuk az anyatej komponenseinek időbeli változásában található minták feltárására. Egy két fázisból álló modellt dolgoztunk ki a mely egy rövidebb lineáris fázisból (a szülést követő kolosztrum időszak, melyet az összetevők gyors változása jellemez) és egy hosszabb autonóm fázisból áll (az összetevők koncentrációja egy stacionárius szinthez konvergál, a változások fokozatos bekövetkezésével), majd a modellt a MilkyBase-ben rögzített adatokra illesztettük. A különböző anyatej összetevők (fehérjék, zsírsavak oligoszacharidok és ásványi anyagok) koncentrációjának időbeli változásában határozott mintázatokat azonosítottunk.

Megállapítottuk, hogy az anyák közötti biológiai különbségek jelentősen befolyásolják az összetevők koncentrációját, például teljes fehérjekoncentrációban nagyobb változatosság mutatható az egyes anyák közötti biológiai különbségeknek köszönhetően, mint az idő függvényében történő eltérés, illetve az anyatej összzsírsav koncentrációja közötti *sztochasztikus biológiai különbségből származó keresztmetszeti eltérés is* nagyobb, mint a hosszanti vagy a földrajzi eredetű eltérés. Modellünk alapján javasoljuk, hogy a kísérleti tervek nem-lineáris mintavételi időpontokat kövessenek, hanem a szülés utáni első két hétben a rövidebb időközönként végezzenek méréseket, hogy a kolosztrum fázisban bekövetkező gyors változásokat pontosabban fel lehessen tárni.

9. SUMMARY

Human milk is the optimal nutrition for infants, a combination of nutrients and bioactive ingredients which are essential for growth, development, and health. The research on human milk has rapidly evolved over the last decades, but despite extensive research, the biochemical complexity of human milk remains unexplored. One of the purposes of the MilkyBase database was to highlight how this knowledge gap can be covered and to show the benefits that Big Data can potentially contribute to human milk research and, more extensively, to nutrition science.

The records of the database come from scientific publications which have been collected using automated search engine and manual methods. This dual approach ensured accuracy and completeness. The records were extensively and carefully selected, digitised, and updated based on published scientific publications. The data structure provides the opportunity to record detailed data on the biochemical composition of human milk and the conditions under which it was produced, such as maternal factors (diet, health status, breastfeeding period), infant characteristics and measurement methods.

The MilkyBase database is an Excel workbook with VBA service macros, which is user-friendly and accessible to researchers with a non- computational background. The ontology is supposed to provide a source and template for researchers and laboratories to put their own data into this standardised format. The use of the platform can encourage collaboration and data sharing between researchers, thus initiating a kind of Wiki-style knowledge sharing.

The database supports a range of analytical methods to reveal patterns in human milk composition. The hierarchical organisation of fields supports statistical analyses, the quantification of uncertainties ('extended numerical variables') helps to reflect measurement error and variability, while the dynamic variables serve to reveal the dynamics of changes of composition of human milk over time.

In addition, researchers and nutritionists can use MilkyBase to identify patterns in human milk composition based on several factors. These patterns can help understand how maternal conditions, birth circumstances or other environmental factors may shape the composition of human milk.

In the second phase of the research, mathematical modelling was used to identify patterns in the variation of breast milk components over time. A two-phase model was developed, consisting of a shorter linear phase (the postnatal colostrum period, characterised by rapid changes) and a longer autonomous phase (the concentration of components converges to a steady-state level with gradual changes), and the model was fitted to data recorded in MilkyBase. We identified a robust pattern of temporal variation in the concentration of different human milk components (proteins, fatty acids oligosaccharides and minerals).

We found that biological differences between mothers significantly affect the concentration of components, for example, there is greater variation in total protein concentration due to biological differences between mothers than variation over time, and the cross-sectional variation from stochastic biological differences in total fatty acid concentration in human milk is greater than longitudinal or geographic variation in total fatty acid concentration in human milk.

10. IRODALOM

1. *Agostoni, C. - Braegger, C. - Decsi, T. - Kolacek, S. - Koletzko, B. - Michaelsen, K. F. - Mihatsch, W. - Moreno, L. A. - Puntis, J. - Shamir, R. - Szajewska, H. - Turck, D. - van Goudoever, J. and Nutrition, E. C. o.*: 2009. Breast-feeding: A Commentary by the ESPGHAN Committee on Nutrition. *Journal of Pediatric Gastroenterology and Nutrition*. 49. (1). 112-125. 10.1097/MPG.0b013e31819f1e05

2. *Bailey, J. E.*: 1998. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnol Prog*. 14. (1). 8-20. 10.1021/bp9701269

3. *Ballard, O. and Morrow, A. L.*: 2013. Human Milk Composition. *Pediatric Clinics of North America*. 60. (1). 49-74. 10.1016/j.pcl.2012.10.002

4. *Banack, H. R. - Hayes-Larson, E. and Mayeda, E. R.*: 2021. Monte Carlo Simulation Approaches for Quantitative Bias Analysis: A Tutorial. *Epidemiologic Reviews*. 43. (1). 106-117. 10.1093/epirev/mxab012

5. *Barabási, A.-L. - Menichetti, G. and Loscalzo, J.*: 2020. The unmapped chemical complexity of our diet. *Nature Food*. 1. (1). 33-37. https://doi.org/10.1038/s43016-019-0005-1

6. *BarabasiLab*: 2024. Barabasi Lab. https://www.barabasilab.com/

7. *Baranyi, J. - McClure, P. - Sutherland, J. - Roberts, T. J. J. o. i. m. and biotechnology*: 1993. Modeling bacterial growth responses. *Journal of Industrial Microbiology*. 12. (3-5). 190-194. Doi 10.1007/Bf01584189

8. *Baranyi, J. and Tamplin, M. L.*: 2004. ComBase: A Common Database on Microbial Responses to Food Environments†. *Journal of Food Protection*. 67. (9). 1967-1971. https://doi.org/10.4315/0362-028X-67.9.1967

9. *Baranyi, J.*: 2005. Quantitative Microbial Ecology of Food: Evolution of mathematical modelling in food microbiology. *Acta Alimentaria - ACTA ALIMENT*. 34. (4). 335-337. 10.1556/AAlim.34.2005.4.1

10. *Baranyi, J. - Pacza, T. - Martins, M. L. - Thakkar, S. K. and Samuel, T. M.*: 2024. Modelling the temporal trajectories of human milk components. *BMC Pregnancy and Childbirth*. 24. (1). 739. 10.1186/s12884-024-06896-z

11. *Baranyi, J. - Rockaya, M. and Ellouze, M.*: 2024. From data to models and predictions in food microbiology. *Current Opinion in Food Science*. 57. 101177. https://doi.org/10.1016/j.cofs.2024.101177

12. *Barde, M. P. and Barde, P. J.*: 2012. What to use to express the variability of data: Standard deviation or standard error of mean? *Perspectives in Clinical Research*. 3. (3). https://journals.lww.com/picp/fulltext/2012/03030/what_to_use_to_express_the_variability_of_data_.7.aspx

13. *Boix-Amorós, A. - Collado, M. C. - Van'T Land, B. - Calvert, A. - Le Doare, K. - Garssen, J. - Hanna, H. - Khaleva, E. - Peroni, D. G. - Geddes, D. T. - Kozyrskyj, A. L. - Warner, J. O. and Munblit, D.*: 2019. Reviewing the evidence on breast milk composition and immunological outcomes. *Nutrition Reviews*. 77. (8). 541-556. https://doi.org/10.1093/nutrit/nuz019

14. *Buchanan, R.*: 1992. Predictive Microbiology: Mathematical Modeling of Microbial Growth in Foods. 250-260.

15. *Buchanan, R. - Whiting, R. C. and Damert, W. C.*: 1997. When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves* 1,* 2. *Food Microbiology*. 14. 313-326. 10.1006/fmic.1997.0125

16. *Buchanan, R. L.*: 1993. Predictive food microbiology. *Trends in Food Science & Technology*. 4. (1). 6-11. https://doi.org/10.1016/S0924-2244(05)80004-4

17. *Carr, L. E. - Virmani, M. D. - Rosa, F. - Munblit, D. - Matazel, K. S. - Elolimy, A. A. and Yeruva, L.*: 2021. Role of Human Milk Bioactives on Infants' Gut and Immune Health. *Front Immunol*. 12. 604080. https://doi.org/10.3389/fimmu.2021.604080

18. *Casavale, K. O. - Ahuja, J. K. C. - Wu, X. - Li, Y. - Quam, J. - Olson, R. - Pehrsson, P. - Allen, L. - Balentine, D. - Hanspal, M. - Hayward, D. - Hines, E. P. - McClung, J. P. - Perrine, C. G. - Belfort, M. B. - Dallas, D. - German, B. - Kim, J. - McGuire, M. - McGuire, M. - Morrow, A. L. - Neville, M. - Nommsen-Rivers, L. - Rasmussen, K. M. - Zempleni, J. and Lynch, C. J.*: 2019. NIH workshop on human milk composition: summary and visions. *Am J Clin Nutr*. 110. (3). 769-779. 10.1093/ajcn/nqz123

19. *Chavalarias, D. - Wallach, J. D. - Li, A. H. T. and Ioannidis, J. P. A.*: 2016. Evolution of ReportingPValues in the Biomedical Literature, 1990-2015. *JAMA*. 315. (11). 1141. https://doi.org/10.1001/jama.2016.1952

20. *Christian, P. - Smith, E. R. - Lee, S. E. - Vargas, A. J. - Bremer, A. A. and Raiten, D. J.*: 2021. The need to study human milk as a biological system. *The American Journal of Clinical Nutrition*. 113. (5). 1063-1072. https://doi.org/10.1093/ajcn/nqab075

21. *Church, S. M.*: 2006. The history of food composition databases. *Nutrition Bulletin*. 31. (1). 15-20. 10.1111/j.1467-3010.2006.00538.x

22. *Cox, D.*: 1990. Role of Models in Statistical Analysis. *Statistical Science*. 5. 10.1214/ss/1177012165

23. *Cukier, K. N. - Mayer-Schönberger, V. and Pitici, M.*: 2014. The Rise of Big Data: How It's Changing the Way We Think about the World.

24. *De Weerth, C. - Aatsinki, A.-K. - Azad, M. B. - Bartol, F. F. - Bode, L. - Collado, M. C. - Dettmer, A. M. - Field, C. J. - Guilfoyle, M. - Hinde, K. - Korosi, A. - Lustermans, H. - Mohd Shukri, N. H. - Moore, S. E. - Pundir, S. - Rodriguez, J. M. - Slupsky, C. M. - Turner, S. - Van Goudoever, J. B. - Ziomkiewicz, A. and Beijers, R.*: 2022. Human milk: From complex tailored nutrition to bioactive impact on child cognition and behavior. *Critical Reviews in Food Science and Nutrition*. 63. (26). 1-38. 10.1080/10408398.2022.2053058

25. *Dym, C. L.*: 2004. Principles of Mathematical Modeling. *Academic Press*. Burlington

26. *Eidelman, A. I. - Schanler, R. J. - Johnston, M. - Landers, S. - Noble, L. - Szucs, K. and Viehmann, L.*: 2012. Breastfeeding and the Use of Human Milk. *Pediatrics*. 129. (3). e827-e841. https://doi.org/10.1542/peds.2011-3552

27. *Elsevier*: 2022.Scopus https://www.scopus.com/

28. *Ene-Obong, H. - Schönfeldt, H. C. - Campaore, E. - Kimani, A. - Mwaisaka, R. - Vincent, A. - El Ati, J. - Kouebou, P. - Presser, K. - Finglas, P. and Charrondiere, U. R.*: 2019. Importance and use of reliable food composition data generation by nutrition/dietetic professionals towards solving Africa's nutrition problem: constraints and the role of FAO/INFOODS/AFROFOODS and other stakeholders in future initiatives. *Proceedings of the Nutrition Society*. 78. (4). 496-505. 10.1017/S0029665118002926

29. *Fischer, L. M. - da Costa, K. A. - Galanko, J. - Sha, W. - Stephenson, B. - Vick, J. and Zeisel, S. H.*: 2010. Choline intake and genetic polymorphisms influence choline metabolite concentrations in human breast milk and plasma123. *The American Journal of Clinical Nutrition*. 92. (2). 336-346. https://doi.org/10.3945/ajcn.2010.29459

30. *Forum, W. E.*: 2012. Big Data, Big Impact: New Possibilities for International Development.

31. *Fusch, G. - Rochow, N. - Choi, A. - Fusch, S. - Poeschl, S. - Ubah, A. O. - Lee, S. Y. - Raja, P. and Fusch, C.*: 2015. Rapid measurement of macronutrients in breast milk: How reliable are infrared milk analyzers? *Clin Nutr*. 34. (3). 465-476. 10.1016/j.clnu.2014.05.005

32. *Galante, L. - Milan, A. M. - Reynolds, C. M. - Cameron-Smith, D. - Vickers, M. H. and Pundir, S.*: 2018. Sex-Specific Human Milk Composition: The Role of Infant Sex in Determining Early Life Nutrition. *Nutrients.* 10. (9). 10.3390/nu10091194

33. *Galit, S.*: 2010. To Explain or to Predict? *Statistical Science*. 25. (3). 289-310. 10.1214/10-STS330

34. *Gauze, G. F.*: 1934. The struggle for existence, by G. F. Gause.

35. *Gertosio, C. - Meazza, C. - Pagani, S. and Bozzola, M.*: 2016. Breastfeeding and its gamut of benefits. *Minerva Pediatr*. 68. (3). 201-212. https://www.ncbi.nlm.nih.gov/pubmed/26023793

36. *Gidrewicz, D. A. and Fenton, T. R.*: 2014. A systematic review and meta-analysis of the nutrient content of preterm and term breast milk. *BMC Pediatr*. 14. 216. 10.1186/1471-2431-14-216

37. *Golan, Y. - Kambe, T. and Assaraf, Y. G.*: 2017. The role of the zinc transporter SLC30A2/ZnT2 in transient neonatal zinc deficiency. *Metallomics*. 9. (10). 1352-1366. 10.1039/c7mt00162b

38. *Gries, D. and Schneider, F. B.*: 1993. Boolean Expressions. *A Logical Approach to Discrete Math.* 10.1007/978-1-4757-3837-7_325-40. 10.1007/978-1-4757-3837-7_3

39. *Hartmann, S.*: 2005. The World as a Process: Simulations in the Natural and Social Sciences.

40. *Heller, S. R. - McNaught, A. - Pletnev, I. - Stein, S. and Tchekhovskoi, D.*: 2015. InChI, the IUPAC International Chemical Identifier. *Journal of Cheminformatics*. 7. (1). 23. 10.1186/s13321-015-0068-4

41. *Hooton, F. - Menichetti, G. and Barabási, A.-L.*: 2020. Exploring food contents in scientific literature with FoodMine. *Scientific Reports*. 10. (1). 16191. 10.1038/s41598-020-73105-0

42. *Horta, B. L.*: 2019. Breastfeeding: Investing in the Future. *Breastfeeding Medicine*. 14. (S1). S-11-S-12. 10.1089/bfm.2019.0032

43. *John, A. - Sun, R. - Maillart, L. - Schaefer, A. - Hamilton Spence, E. and Perrin, M. T.*: 2019. Macronutrient variability in human milk from donors to a milk bank: Implications for feeding preterm infants. *PLOS ONE*. 14. (1). e0210610. 10.1371/journal.pone.0210610

44. *Kim, M. H. - Shim, K. S. - Yi, D. Y. - Lim, I. S. - Chae, S. A. - Yun, S. W. - Lee, N. M. - Kim, S. Y. and Kim, S.*: 2019. Macronutrient Analysis of Human Milk according to Storage and Processing in Korean Mother. *Pediatr Gastroenterol Hepatol Nutr*. 22. (3). 262-269. 10.5223/pghn.2019.22.3.262

45. *Kim, S. Y. and Yi, D. Y.*: 2020. Components of human breast milk: from macronutrient to microbiome and microRNA. *Clin Exp Pediatr*. 63. (8). 301-309. 10.3345/cep.2020.00059

46. *Kramer, M. S.*: 2010. "Breast is best": The evidence. *Early Human Development*. 86. (11). 729-732. https://doi.org/10.1016/j.earlhumdev.2010.08.005

47. *Kunz, C. - Rudloff, S. - Schad, W. and Braun, D.*: 1999. Lactose-derived oligosaccharides in the milk of elephants: comparison with human milk. *British Journal of Nutrition*. 82. (5). 391-399. 10.1017/S0007114599001798

48. *Laney, D.*: 2001. 3D Data Management Controlling Data Volume Velocity And Variety. *Application Delivery Strategies.* 1-4.

49. *Léké, A. - Grognet, S. - Deforceville, M. - Goudjil, S. - Chazal, C. - Kongolo, G. - Dzon, B. E. and Biendo, M.*: 2019. Macronutrient composition in human milk from mothers of preterm and term neonates is highly variable during the lactation period. *Clinical Nutrition Experimental*. 26. 59-72. https://doi.org/10.1016/j.yclnex.2019.03.004

50. *Levenberg, K. J. Q. o. A. M.*: 1944. A METHOD FOR THE SOLUTION OF CERTAIN NON – LINEAR PROBLEMS IN LEAST SQUARES. 2. 164-168.

51. *Liu, Y. - Liu, X. and Wang, L.*: 2019. The investigation of fatty acid composition of breast milk and its relationship with dietary fatty acid intake in 5 regions of China. *Medicine*. 98. (24). e15855. 10.1097/MD.0000000000015855

52. *Lönnerdal, B.*: 2016. Bioactive Proteins in Human Milk: Health, Nutrition, and Implications for Infant Formulas. *J Pediatr*. 173 Suppl. S4-9. 10.1016/j.jpeds.2016.02.070

53. *Lyons, K. E. - Ryan, C. A. - Dempsey, E. M. - Ross, R. P. and Stanton, C.*: 2020. Breast Milk, a Source of Beneficial Microbes and Associated Benefits for Infant Health. 12. (4). 1039. https://www.mdpi.com/2072-6643/12/4/1039

54. *Mahdinia, E. - Liu, S. - Demirci, A. and Puri, V. M.*: 2020. Microbial Growth Models. *Food Safety Engineering*. 10.1007/978-3-030-42660-6_14357-398. 10.1007/978-3-030-42660-6_14

55. *Manyika, J. - Chui, M. - Brown, B. - Bughin, J. - Dobbs, R. - Roxburgh, C. and Byers, A. H.*: 2011. Big data: The next frontier for innovation, competition, and productivity. https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/big%20data%20the%20next%20frontier%20for%20innovation/mgi_big_data_full_report.pdf

56. *Marconi, S. - Durazzo, A. - Camilli, E. - Lisciani, S. - Gabrielli, P. - Aguzzi, A. - Gambelli, L. - Lucarini, M. and Marletta, L.*: 2018. Food Composition Databases: Considerations about Complex Food Matrices. *Foods*. 7. (1). 10.3390/foods7010002

57. *Marquardt, D. W.*: 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. 11. (2). 431-441. 10.1137/0111030

58. *Martysiak-Żurowska, D. - Szlagatys-Sidorkiewicz, A. and Zagierski, M.*: 2013. Concentrations of alpha- and gamma-tocopherols in human breast milk during the first months of lactation and in infant formulas. *Maternal & Child Nutrition*. 9. (4). 473-482. https://doi.org/10.1111/j.1740-8709.2012.00401.x

59. *Marvin, H. J. - Janssen, E. M. - Bouzembrak, Y. - Hendriksen, P. J. and Staats, M.*: 2017. Big data in food safety: An overview. *Crit Rev Food Sci Nutr*. 57. (1549-7852 (Electronic)). 2286-2295. 10.1080/10408398.2016.1257481

60. *May, R. M.*: 2001. Stability and Complexity in Model Ecosystems. *Princeton University Press*.

61. *Mayer-Schönberger, V. and Cukier, K.*: 2013. Big data: A revolution that will transform how we live, work, and think. *Houghton Mifflin Harcourt*. Boston, Massachusetts

62. *McDonald, K. and Sun, D.-W.*: 1999. Predictive food microbiology for the meat industry: a review. *Int J Food Microbiol*. 52. (1). 1-27. https://doi.org/10.1016/S0168-1605(99)00126-9

63. *McMeekin, T. A. - Olley, J. - Ratkowsky, D. A. and Ross, T.*: 2002. Predictive microbiology: towards the interface and beyond. *Int J Food Microbiol*. 73. (2-3). 395-407. 10.1016/s0168-1605(01)00663-8

64. *MeSH*: 2020. Medical Subject Headings. https://www.nlm.nih.gov/mesh/meshhome.html

65. *Moltó-Puigmartí, C. - Plat, J. - Mensink, R. P. - Müller, A. - Jansen, E. - Zeegers, M. P. and Thijs, C.*: 2010. FADS1 FADS2 gene variants modify the association between fish intake and the docosahexaenoic acid proportions in human milk1234. *The American Journal of Clinical Nutrition*. 91. (5). 1368-1376. https://doi.org/10.3945/ajcn.2009.28789

66. *Morgenstern, J. D. - Rosella, L. C. - Costa, A. P. - de Souza, R. J. and Anderson, L. N.*: 2021. Perspective: Big Data and Machine Learning Could Help Advance Nutritional Epidemiology. *Adv Nutr*. 12. (3). 621-631. 10.1093/advances/nmaa183

67. *Murray, J. D.*: 2003. Mathematical Biology I. An Introduction.

68. *Nagele, P.*: 2003. Misuse of standard error of the mean (sem) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *British Journal of Anaesthesia*. 90. (4). 514-516. 10.1093/bja/aeg087

69. *Newburg, D. S.*: 2001. Bioactive Components of Human Milk. *Bioactive Components of Human Milk.* 10.1007/978-1-4615-1371-1_13-10. 10.1007/978-1-4615-1371-1_1

70. *NIST*: 2022. Unit Conversion, National Institue of Standards and Technology. https://www.nist.gov/pml/owm/metric-si/unit-conversion

71. *Nyquist, S. K. - Gao, P. - Haining, T. K. J. - Retchin, M. R. - Golan, Y. - Drake, R. S. - Kolb, K. - Mead, B. E. - Ahituv, N. - Martinez, M. E. - Shalek, A. K. - Berger, B. and Goods, B. A.*: 2022. Cellular and transcriptional diversity over the course of human lactation. 119. (15). e2121720119. doi:10.1073/pnas.2121720119

72. *O'Malley, M. A. and Parke, E. C.*: 2018. Microbes, mathematics, and models. *Studies in History and Philosophy of Science Part A*. 72. 1-10. https://doi.org/10.1016/j.shpsa.2018.07.001

73. *Pacza, T. - Martins, M. L. - Rockaya, M. - Müller, K. - Chatterjee, A. - Barabási, A.-L. and Baranyi, J.*: 2022. MilkyBase, a database of human milk composition as a function of maternal-, infant- and measurement conditions. *Scientific Data*. 9. (1). 557. 10.1038/s41597-022-01663-1

74. *Pacza, T. - Martins, M. L. - Rockaya, M. - Müller, K. - Chatterjee, A. - Barabási, A.-L. and Baranyi, J.*: 2022. MilkyBase, a database of human milk composition as a function of maternal-, infant- and measurement conditions. https://figshare.com/s/c44b92932fc1a5785cd3

75. *Patro-Gołąb, B. - Zalewski, B. M. - Kołodziej, M. - Kouwenhoven, S. - Poston, L. - Godfrey, K. M. - Koletzko, B. - van Goudoever, J. B. and Szajewska, H.*: 2016. Nutritional interventions or exposures in infants and children aged up to 3 years and their effects on subsequent risk of overweight, obesity and body fat: a systematic review of systematic reviews. *Obes Rev*. 17. (12). 1245-1257. https://doi.org/10.1111/obr.12476

76. *Perrella, S. - Gridneva, Z. - Lai, C. T. - Stinson, L. - George, A. - Bilston-John, S. and Geddes, D.*: 2021. Human milk composition promotes optimal infant growth, development and health. *Seminars in Perinatology*. 45. (2). 151380. https://doi.org/10.1016/j.semperi.2020.151380

77. *Picciano, M. F.*: 2001. Nutrient Composition of Human Milk. *Pediatric Clinics of North America*. 48. (1). 53-67. https://doi.org/10.1016/S0031-3955(05)70285-6

78. *Pirt, S. J.*: 1975. Principles of microbe and cell cultivation. *Blackwell Scientific Publications.*

79. *PTFI*: 2021. Periodic Table of Food Initiative https://foodperiodictable.org/

80. *PubMed*: 2022. PubMed. Bethesda (MD): National Library of Medicine (US). [1946] https://www.ncbi.nlm.nih.gov/pubmed/

81. *Qureshi, S.*: 2020. Why Data Matters for Development? Exploring Data Justice, Micro-Entrepreneurship, Mobile Money and Financial Inclusion. *Information Technology for Development*. 26. (2). 201-213. 10.1080/02681102.2020.1736820

82. *Ratkowsky, D. A. - Lowry, R. K. - McMeekin, T. A. - Stokes, A. N. and Chandler, R. E.*: 1983. Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *Journal of Bacteriology*. 154. (3). 1222-1226. 10.1128/jb.154.3.1222-1226.1983

83. *Reinsel, D. - Gantz, J. and Rydning, J.*: 2018. Data Age 2025 - The Digitization of the World From Edge to Core. *IDC White Paper.* #US44413318. https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

84. *Rodríguez-Cruz, M. - Alba, C. - Aparicio, M. - Checa, M. - Fernández, L. and Rodríguez, J.*: 2020. Effect of Sample Collection (Manual Expression vs. Pumping) and Skimming on the Microbial Profile of Human Milk Using Culture Techniques and Metataxonomic Analysis. *Microorganisms*. 8. (9). 1278. 10.3390/microorganisms8091278

85. *Rollins, N. C. - Bhandari, N. - Hajeebhoy, N. - Horton, S. - Lutter, C. K. - Martines, J. C. - Piwoz, E. G. - Richter, L. M. and Victora, C. G.*: 2016. Why invest, and what it will take to improve breastfeeding practices? *Lancet*. 387. (10017). 491-504. https://doi.org/10.1016/s0140-6736(15)01044-2

86. *Rossum, C. - Büchner, F. and Hoekstra, J.*: 2005. Quantification of health effects of breastfeeding - Review of the literature and model simulation. *Annals of Nutrition and Metabolism*. 51.

87. *Salciccioli, J. D. - Crutain, Y. - Komorowski, M. and Marshall, D. C.*: 2016. Sensitivity Analysis and Model Validation. *Secondary Analysis of Electronic Health Records.* 10.1007/978-3-319-43742-2_17263-271. 10.1007/978-3-319-43742-2_17

88. *Saltelli, A.*: 2019. A short comment on statistical versus mathematical modelling. *Nature Communications*. 10. (1). 3870. 10.1038/s41467-019-11865-8

89. *Samuel, T. M. - Binia, A. - de Castro, C. A. - Thakkar, S. K. - Billeaud, C. - Agosti, M. - Al-Jashi, I. - Costeira, M. J. - Marchini, G. - Martínez-Costa, C. - Picaud, J.-C. - Stiris, T. - Stoicescu, S.-M. - Vanpeé, M. - Domellöf, M. - Austin, S. and Sprenger, N.*: 2019. Impact of maternal characteristics on human milk oligosaccharide composition over the first 4 months of lactation in a cohort of healthy European mothers. *Scientific Reports*. 9. (1). 11767. 10.1038/s41598-019-48337-4

90. *Samuel, T. M. - Zhou, Q. - Giuffrida, F. - Munblit, D. - Verhasselt, V. and Thakkar, S. K.*: 2020. Nutritional and Non-nutritional Composition of Human Milk Is Modulated by Maternal, Infant, and Methodological Factors. *Frontiers in Nutrition*. 7. 576133. https://doi.org/10.3389/fnut.2020.576133

91. *Samuel, T. M. - Thielecke, F. - Lavalle, L. - Chen, C. - Fogel, P. - Giuffrida, F. - Dubascoux, S. - Martinez-Costa, C. - Haaland, K. - Marchini, G. - Agosti, M. - Rakza, T. - Costeira, M. J. - Picaud, J. C. - Billeaud, C. and Thakkar, S. K.*: 2022. Mode of Neonatal Delivery Influences the Nutrient Composition of Human Milk: Results From a Multicenter European Cohort of Lactating Women. *Front Nutr*. 9. 834394. 10.3389/fnut.2022.834394

92. *Sánchez, C. - Franco, L. - Regal, P. - Lamas, A. - Cepeda, A. and Fente, C.*: 2021. Breast Milk: A Source of Functional Compounds with Potential Application in Nutrition and Therapy. *Nutrients*. 13. (3). 1026. 10.3390/nu13031026

93. *SciELO*: 2022.SciELO – Scientific Electronic Library Online https://scielo.org/

94. *Shenhav, L. and Azad, M. B.*: 2022. Using Community Ecology Theory and Computational Microbiome Methods To Study Human Milk as a Biological System. *mSystems*. 7. (1). e01132-01121. 10.1128/msystems.01132-21

95. *Silva, C. - Valle, B. - Matos, U. - Amaral, Y. - Moreira, M. and Vieira, A.*: 2021. Influence of different breast expression techniques on human colostrum macronutrient concentrations. *Journal of Perinatology*. 41. (5). 1-4. 10.1038/s41372-021-00989-9

96. *Strogatz, S. H.*: 2018. Nonlinear Dynamics and Chaos with Student Solution manual : With Applications to Physics, Biology, Chemistry and Engineering (2nd ed.). *Taylor & Francis Group*.

97. *Vaux, D. L.*: 2012. Know when your numbers are significant. *Nature*. 492. (7428). 180-181. https://doi.org/10.1038/492180a

98. *Victora, C. G. - Smith, P. G. - Vaughan, J. P. - Nobre, L. C. - Lombardi, C. - Teixeira, A. M. - Fuchs, S. M. - Moreira, L. B. - Gigante, L. P. and Barros, F. C.*: 1987. Evidence for protection by breast-feeding against infant deaths from infectious diseases in Brazil. *Lancet*. 2. (8554). 319-322. 10.1016/s0140-6736(87)90902-0

99. *Victora, C. G. - Bahl, R. - Barros, A. J. - França, G. V. - Horton, S. - Krasevec, J. - Murch, S. - Sankar, M. J. - Walker, N. and Rollins, N. C.*: 2016. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *Lancet*. 387. (10017). 475-490. https://doi.org/10.1016/s0140-6736(15)01024-7

100. *Weisberg, M.*: 2013. Simulation and Similarity: Using Models to Understand the World. 10.1093/acprof:oso/9780199933662.001.0001 10.1093/acprof:oso/9780199933662.001.0001

101. *WHO*: 2003. Global Strategy for Infant and Young Child Feeding. *Fifthy-fourth world health assembly*. https://apps.who.int/iris/bitstream/handle/10665/42590/9241562218.pdf?sequence=1. (1). 8-8.

102. *WOS*: 2022.Web of Science https://www.webofscience.com/wos

103. *Wu, C. - Buyya, R. and Ramamohanarao, K.*: 2016. Chapter 1 - Big Data Analytics = Machine Learning + Cloud Computing. *Big Data.* https://doi.org/10.1016/B978-0-12-805394-2.00001-53-38. https://doi.org/10.1016/B978-0-12-805394-2.00001-5

104. *Wu, W. - Balter, A. - Vodsky, V. - Odetallh, Y. - Ben-Dror, G. - Zhang, Y. and Zhao, A.*: 2021. Chinese Breast Milk Fat Composition and Its Associated Dietary Factors: A Pilot Study on Lactating Mothers in Beijing. *Front Nutr*. 8. 606950. 10.3389/fnut.2021.606950

105. *Yeung, A. W.*: 2023. Food Composition Databases (FCDBs): A Bibliometric Analysis. *Nutrients.* 15. (16). 10.3390/nu15163548

106. *Yi, D. and Kim, S.*: 2021. Human Breast Milk Composition and Function in Human Health: From Nutritional Components to Microbiome and MicroRNAs. *Nutrients*. 13. (9). 3094. 10.3390/nu13093094

107. *Zivkovic, A. M. - German, J. B. - Lebrilla, C. B. and Mills, D. A.*: 2011. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. *Proc Natl Acad Sci U S A*. 108 Suppl 1. (Suppl 1). 4653-4658. 10.1073/pnas.1000083107

108. *Zwietering, M. H. - Jongenburger, I. - Rombouts, F. M. and van 't Riet, K.*: 1990. Modeling of the bacterial growth curve. *Appl Environ Microbiol*. 56. (6). 1875-1881. 10.1128/aem.56.6.1875-1881.1990

11. ÁBRAJEGYZÉK

11.1. TÁBLÁZATOK

12. PUBLIKÁCIÓK AZ ÉRTEKEZÉS TÉMAKÖRÉBEN

DEBRECENI EGYETEM
EGYETEMI ÉS NEMZETI KÖNYVTÁR
H-4002 Debrecen, Egyetem tér 1, Pf.: 400
Tel.: 52/410-443, e-mail: publikaciok@lib.unideb.hu

Nyilvántartási szám: DEENK/567/2024.PL
Tárgy: PhD Publikációs Lista

Jelölt: Pacza Tünde
Doktori Iskola: Táplálkozás- és Élelmiszertudományi Doktori Iskola. Élelmiszertudományi doktori program
MTMT azonosító: 10084787

A PhD értekezés alapjául szolgáló közlemények

Idegen nyelvű tudományos közlemények külföldi folyóiratban (4)

1. Baranyi, J., Csorba, S., Farkas, Z., **Pacza, T.**, Jóźwiak, Á.: Internal dynamics of patent reference networks using the Bray-Curtis dissimilarity measure.
   J Big Data. 11 (1), 1-10, 2024. EISSN: 2196-1115.
   DOI: http://dx.doi.org/10.1186/s40537-024-00883-z
   IF: 8.6 (2023)

2. Baranyi, J.*, **Pacza, T.***, Martins, M. L., Thakkar, S. K., Samuel, T. M.: Modelling the temporal trajectories of human milk components.
   BMC Pregnancy Childbirth. 24 (1), 1-13, 2024. EISSN: 1471-2393.
   DOI: http://dx.doi.org/10.1186/s12884-024-06896-z
   *Megosztott első szerzős közlemény.
   IF: 2.8 (2023)

3. Martins, M. L., **Pacza, T.**, Müller, K. E., Baranyi, J.: A computational approach to nutrition science reveals the dynamics of the protein content of human milk.
   Innovative Food Science & Emerging Technologies. 82, 1-5, 2022. ISSN: 1466-8564.
   DOI: http://dx.doi.org/10.1016/j.ifset.2022.103167
   IF: 6.6

4. **Pacza, T.**, Martins, M. L., Rockaya, M., Müller, K. E., Chatterjee, A., Barabási, A. L., Baranyi, J.: MilkyBase, a database of human milk composition as a function of maternal-, infant- and measurement conditions.
   Sci Data. 9 (1), 1-7, 2022. EISSN: 2052-4463.
   DOI: http://dx.doi.org/10.1038/s41597-022-01663-1
   IF: 9.8

DEBRECENI EGYETEM Egyetemi és Nemzeti Könyvtár

DEBRECENI EGYETEM
EGYETEMI ÉS NEMZETI KÖNYVTÁR
H-4002 Debrecen, Egyetem tér 1, Pf.: 400
Tel.: 52/410-443, e-mail: publikaciok@lib.unideb.hu

Idegen nyelvű absztrakt kiadványok (1)

5. **Pacza, T.**, Martins, M. L., Müller, K. E., Baranyi, J.: MilkyBase- A Database for Molecular-Level Mapping of the Composition of the Human Milk.
In: International Milk Genomic Consortium : IMGC HYBRID Symposium 2023, IMGC, Cork, 1, 2023.

További közlemények

Idegen nyelvű tudományos közlemények hazai folyóiratban (1)

6. Mposula, Z., **Pacza, T.**, Szepesi, J., Máthé, E.: Lifestyle and socio-economic inequalities in diabetes prevalence in Madadeni Township, South Africa.
Acta Med. Sociol. 14 (37), 5-21, 2023. ISSN: 2062-0284.
DOI: http://dx.doi.org/10.19055/ams.2023.12/15/1

A közlő folyóiratok összesített impakt faktora: 27,8
A közlő folyóiratok összesített impakt faktora (az értekezés alapjául szolgáló közleményekre): 27,8

A DEENK a Jelölt által az iDEa Tudóstérbe feltöltött adatok bibliográfiai és tudománymetriai ellenőrzését a tudományos adatbázisok és a Journal Citation Reports Impact Factor lista alapján elvégezte.

Debrecen, 2024.11.14.

DEBRECENI EGYETEM * Egyetemi és Nemzeti Könyvtár *

13. KÖSZÖNETNYILVÁNÍTÁS

Ezúton szeretnék köszönetet mondani témavezetőmnek Dr. Baranyi Józsefnek a lehetőségért és támogatásért, valamint a PhD időszakom és a doktori munkám elkészítése alatt nyújtott folyamatos szakmai segítségért és útmutatásért.

Köszönöm kutatótársamnak és barátomnak Mayara Lopes Martinsnak, a közös munkát, az együtt töltött PhD-s éveket.

Köszönöm Dr. Máthé Endrének a DE MÉK Táplálkozástudományi Intézet vezetőjének, hogy lehetővé tette és segítette doktori munkám elvégzését az intézet keretein belül, valamint Szepesi Juditnak, hogy támogató segítsége végig kísérte a PhD-s éveimet. Valamint köszönöm az intézet többi dolgozójának és doktorandusz hallgató társaimnak is a képzés alatt nyújtott támogatást.

Végezetül (de nem utolsó sorban) köszönöm férjemnek, Dr. Vámosi Györgynek és gyerekeimnek (Marci, Ábel, Boldi, Boglár, Bendi és Mimi), hogy végig bíztattak, segítettek és támogattak. Nélkülük nem jöhetett volna létre ez a PhD disszertáció.

14. NYILATKOZATOK

NYILATKOZAT

Ezen értekezést a Debreceni Egyetem **Táplálkozás- és Élelmiszertudományi Doktori Iskola (Élelmiszertudományi programja)** keretében készítettem, a Debreceni Egyetem doktori (Ph.D.) fokozatának elnyerése céljából.

Debrecen, 20…….

……..

a jelölt aláírása

NYILATKOZAT

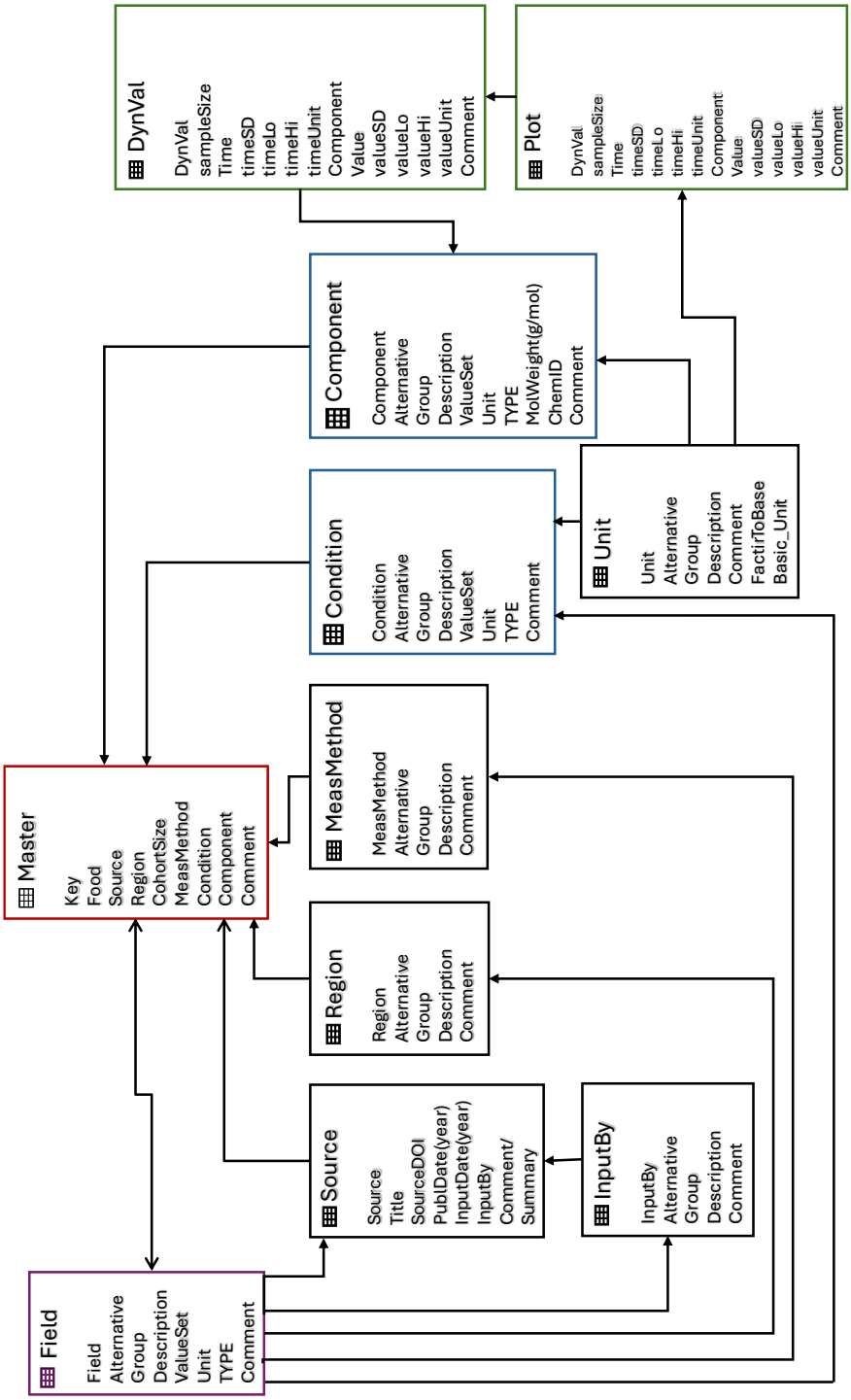
Tanúsítom, hogy **Vámosiné Pacza Tünde** doktorjelölt **2020 - 2024** között a fent megnevezett Doktori Iskola keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult, az értekezés a jelölt önálló munkája. Az értekezés elfogadását javaslom.

Debrecen, 20…….

……..

a témavezető aláírása

15. MELLÉKLETEK



Kiegészítő információ 1. A MilkyBase adatbázis kapcsolódó tábláinak szerkezete

Kiegészítő információ 2: Az elemzésekhez felhasznált publikációk forrásadatai.
* Forrásazonosító : “Source ID” a MilkyBase adatbázisban
(https://figshare.com/articles/dataset/MilkyBase_database/20540454?file=48906571)

Ábra	Forrás azonosító*	Cím	DOI	Publ. Dátum
7.ábra.	Martysiak-Zurowska_12	Concentrations of alpha- and gamma-tocopherols in human breast milk during the first months of lactation and in infant formulas	doi.org/10.1111/j.1740-8709.2012.00401.x	2012
29.ábra.	Kunz_99	Lactose-derived oligosaccharides in the milk of elephants comparison with human milk	doi.org/10.1017/S0007114599001798	1999
29.ábra.	Coppa_99	Oligosaccharides in human milk during different phases of lactation	doi.org/10.1111/j.1651-2227.1999.tb01307.x	1999
29.ábra.	Roldan_20	Human Milk Oligosaccharides and Their Association With Late-Onset Neonatal Sepsis in Peruvian Very-Low-Birth-Weight Infants	doi.org/10.1093/ajcn/nqaa102	2020
31.ábra.	Liu_19	The investigation of fatty acid composition of breast milk and its relationship with dietary fatty acid intake in 5 regions of China	doi.org/10.1097/MD.0000000000015855	2019
32.ábra.	Kunz_99	Lactose-derived oligosaccharides in the milk of elephants’ comparison with human milk	doi.org/10.1017/S0007114599001798	1999
32.ábra.	Coppa_99	Oligosaccharides in human milk during different phases of lactation	doi.org/10.1111/j.1651-2227.1999.tb01307.x	1999
32.ábra.	Roldan_20	Human Milk Oligosaccharides and Their Association With Late-Onset Neonatal Sepsis in Peruvian Very-Low-Birth-Weight Infants	doi.org/10.1093/ajcn/nqaa102	2020
33.ábra., 34.ábra.	John_19	Macronutrient variability in human milk from donors to a milk bank. Implications for feeding preterm infants	doi.org/10.1371/journal.pone.0210610	2019

Kiegészítő információ 2: Az elemzésekhez felhasznált publikációk forrásadatai.
* Forrásazonosító : “Source ID” a MilkyBase adatbázisban
(https://figshare.com/articles/dataset/MilkyBase_database/20540454?file=48906571)

Ábra	Forrás azonosító*	Cím	DOI	Publ. Dátum
35.ábra.	Sanchez-Pozo_86	Changes in the Protein Fractions of Human Milk During Lactation	doi.org/10.1159/000177172	1986
35.ábra.	Maly_18	Preterm human milk macronutrient concentration is independent of gestational age at birth	doi.org/10.1136/archdischild-2016-312572	2018
35.ábra.	Montagne_99	Immunological and nutritional composition of human milk in relation to prematurity and mothers' parity during the first 2 weeks of lactation	doi.org/10.1097/00005176-199907000-00018	1999
35.ábra.	Saarela_05	Macronutrient and energy contents of human milk fractions during the first six months of lactation	doi.org/10.1111/j.1651-2227.2005.tb02070.x	2005
35.ábra., 38.ábra.	Bauer_11	Longitudinal Analysis of Macronutrients and Minerals in Human Milk Produced by Mothers of Preterm Infants	doi.org/10.1016/j.clnu.2010.08.003	2011
35.ábra., 38.ábra.	Michaelsen_90	Variation in Macronutrients in Human Bank Milk: Influencing Factors and Implications for Human Milk Banking	doi.org/10.1097/00005176-199008000-00013	1990
35.ábra., 38.ábra.	Sann_81	Comparison of the Composition of Breast Milk From Mothers of Term and Preterm Infants	doi.org/10.1111/j.1651-2227.1981.tb07182.x	1981
35.ábra., 39.ábra., 40.ábra.	vanBeusekom_93a	Milk of patients with tightly controlled insulin-dependent diabetes mellitus has normal macronutrient and fatty acids composition	doi.org/10.1093/ajcn/57.6.938	1993

Kiegészítő információ 2: Az elemzésekhez felhasznált publikációk forrásadatai.
* Forrásazonosító : “Source ID” a MilkyBase adatbázisban
(https://figshare.com/articles/dataset/MilkyBase_database/20540454?file=48906571)

Ábra	Forrás azonosító*	Cím	DOI	Publ. Dátum
36.ábra., 38.ábra., 39b.ábra., 40b.ábra.	Liu_19	The investigation of fatty acid composition of breast milk and its relationship with dietary fatty acid intake in 5 regions of China	doi.org/10.1097/MD.0000000000015855	2019
36.ábra., 38.ábra., 39.ábra., 40.ábra.	Samuel_22	Mode of Neonatal Delivery Influences the Nutrient Composition of Human Milk: Results From a Multicenter European Cohort of Lactating Women	doi.org/10.3389/fnut.2022.834394	2022
37.ábra.	Samuel_19	Impact of maternal characteristics on human milk oligosaccharide composition over the first 4 months of lactation in a cohort of healthy European mothers	doi.org/10.1038/s41598-019-48337-4	2019
38.ábra.	Antonakou_10	Breast milk tocopherol content during the first six months in exclusively breastfeeding Greek women	doi.org/10.1007/s00394-010-0129-4	2010
38.ábra.	Czosnykowska_18	Breast milk macronutrient components in prolonged lactation	doi.org/10.3390/nu10121893	2018

Kiegészítő információ 2: Az elemzésekhez felhasznált publikációk forrásadatai.
* Forrásazonosító : “Source ID” a MilkyBase adatbázisban
(https://figshare.com/articles/dataset/MilkyBase_database/20540454?file=48906571)

Ábra	Forrás azonosító*	Cím	DOI	Publ. Dátum
38.ábra.	LondonDatabase	The Composition of Mature Human Milk.	assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/743819/The_Composition_of_Mature_Human_Milk__1977_.pdf	1977
38.ábra.	Maas_98	Development of macronutrient composition of very preterm human milk	doi.org/10.1017/S0007114598001743	1998
38.ábra.	Paulaviciene_20	Circadian Changes in the Composition of Human Milk Macronutrients Depending on Pregnancy Duration: A Cross-Sectional Study	doi.org/10.1186/s13006-020-00291-y	2020
38.ábra.,	Lipkie_15	Longitudinal Survey of Carotenoids in Human Milk from Urban Cohorts in China, Mexico, and the USA	doi.org/10.1371/journal.pone.0127729	2015
38.ábra., 39.ábra., 40.ábra.	Fidler_01	Fat Content and Fatty Acid Composition of Fresh, Pasteurized, or Sterilized Human Milk	doi.org/10.1007/978-1-4615-1371-1_60	2001
38.ábra., 39.ábra., 40.ábra.	Wu_20	Lactational Changes of Fatty Acids and Fat-Soluble Antioxidants in Human Milk From Healthy Chinese Mothers	doi.org/10.1017/S0007114520000239	2020

Kiegészítő információ 2: Az elemzésekhez felhasznált publikációk forrásadatai.
* Forrásazonosító : “Source ID” a MilkyBase adatbázisban
(https://figshare.com/articles/dataset/MilkyBase_database/20540454?file=48906571)

Ábra	Forrás azonosító*	Cím	DOI	Publ. Dátum
38.ábra., 39b.ábra., 40.ábra.	Yip_20	Quantification of breast milk trans fatty acids and trans fat intake by Hong Kong lactating women	doi.org/10.1038/s41430-020-0604-9	2020
39.ábra., 40.ábra.	Giuffrida_22	Human milk fatty acid composition and its association with maternal blood and adipose tissue fatty acid content in a cohort of women from Europe	doi.org/10.1007/s00394-021-02788-6	2022
39.ábra., 40.ábra.	Giuffrida_16	Temporal Changes of Human Breast Milk Lipids of Chinese Mothers	doi.org/10.3390/nu8110715	2016
39.ábra., 40.ábra.	Jiang_16	Changes in Fatty Acid Composition of Human Milk Over Lactation Stages and Relationship With Dietary Intake in Chinese Women	doi.org/10.1039/c6fo00304d	2016
39.ábra., 40.ábra.	SalaVila_04	The Source of Long-Chain PUFA in Formula Supplements Does Not Affect the Fatty Acid Composition of Plasma Lipids in Full-Term Infants	doi.org/10.1093/jn/134.4.868	2004
39.ábra., 40.ábra.	SalaVila_06	Influence of dietary source of docosahexaenoic and arachidonic acids on their incorporation into membrane phospholipids of red blood cells in term infants	doi.org/10.1016/j.plefa.2005.10.003	2006
39.ábra., 40.ábra.	Thakkar_19	Temporal Progression of Fatty Acids in Preterm and Term Human Milk of Mothers from Switzerland	doi.org/10.3390/nu11010112	2019

Kiegészítő információ 2: Az elemzésekhez felhasznált publikációk forrásadatai.
* Forrásazonosító : “Source ID” a MilkyBase adatbázisban
(https://figshare.com/articles/dataset/MilkyBase_database/20540454?file=48906571)

Ábra	Forrás azonosító*	Cím	DOI	Publ. Dátum
39.ábra., 40.ábra.	Wesolowska _19	Lipid profile lipase bioactivity and lipophilic antioxidant content in high pressure processed donor human milk	doi.org/10.3390/nu11091 972	2019
39.ábra., 40.ábra.	Wu_19	Fatty acid positional distribution (sn -2 fatty acids) and phospholipid composition in Chinese breast milk from colostrum to mature stage	doi.org/10.1017/S000711 4518002994	2019
39.ábra., 40.ábra	Xiang_99	Composition of long chain polyunsaturated fatty acids in human milk and growth of young infants in rural areas of northern China	doi.org/10.1080/0803525 9950170268	1999
39.ábra., 40b.ábra.	Sanders_92	The Influence of a Vegetarian Diet on the Fatty Acid Composition of Human Milk and the Essential Fatty Acid Status of the Infant	doi.org/10.1016/s0022-3476(05)81239-9	1992
39.ábra., 40.ábra.	Xiang_00	Long-chain polyunsaturated fatty acids in human milk and brain growth during early infancy	doi.org/10.1080/0803525 00750028735	2000
39b.ábra.	Ehrenkranz_ 84	Total Lipid Content and Fatty Acid Composition of Preterm Human Milk	doi.org/10.1097/0000517 6-198411000-00021	1984
39.ábra.	Wardell_81	Effect of Pasteurization and of Freezing and Thawing Human Milk on Its Triglyceride Content	doi.org/10.1111/j.1651-2227.1981.tb05724.x	1981
39b.ábra., 40.ábra.	Ni_21	Total and Sn-2 Fatty Acid Profile in Human Colostrum and Mature Breast Milk of Women Living in Inland and Coastal Areas of China	doi.org/10.1159/0005103 79	2021

Kiegészítő információ 2: Az elemzésekhez felhasznált publikációk forrásadatai.
* Forrásazonosító : “Source ID” a MilkyBase adatbázisban
(https://figshare.com/articles/dataset/MilkyBase_database/20540454?file=48906571)

Ábra	Forrás azonosító*	Cím	DOI	Publ. Dátum
39b.ábra., 40b.ábra.	Peng_21	Xiang Study: an association of breastmilk composition with maternal body mass index and infant growth during the first 3 month of life.	doi.org/10.4162/nrp.2021.15.3.367	2021
40a.ábra.	Jagodic_20	Dietary Habits of Slovenian Inland and Coastal Primiparous Women and Fatty Acid Composition of Their Human Milk Samples	doi.org/10.1016/j.fct.2020.111299	2020
40.ábra.	Precht_99	C18 1 C18 2 and C18 3 trans and cis fatty acid isomers including conjugated cis δ9, trans δ11 linoleic acid (CLA) as well as total fat composition of German human milk ...	doi.org/10.1002/(SICI)1521-3803(19990801)43:4<233::AID-FOOD233>3.0.CO;2-B	1999
40b.ábra.	Olafsdottir_06	Polyunsaturated fatty acids in the diet and breast milk of lactating Icelandic women with traditional fish and cod liver oil consumption	doi.org/10.1159/000091685	2006
40b.ábra.	Sanders_78	Studies of vegans: the fatty acid composition of plasma choline phosphoglycerides, erythrocytes, adipose tissue, and breast milk, and some indicators of susceptibility to ischemic heart disease in vegans and omnivore controls	doi.org/10.1093/ajcn/31.5.805	1978
40b.ábra.	Wang_20	Effect of lactation stages and dietary intake on the fatty acid composition of human milk (A study in northeast China)	doi.org/10.1016/j.idairyj.2019.104580	2020

Kiegészítő információ 2: Az elemzésekhez felhasznált publikációk forrásadatai.
* Forrásazonosító : “Source ID” a MilkyBase adatbázisban
(https://figshare.com/articles/dataset/MilkyBase_database/20540454?file=48906571)

Ábra	Forrás azonosító*	Cím	DOI	Publ. Dátum
40b.ábra.	Zhao_18	Differences in the Triacylglycerol and Fatty Acid Compositions of Human Colostrum and Mature Milk.	doi.org/10.1021/acs.jafc.8b00868	2018